Reading Mastery and Learning Disabled Students:

A Comment on the What Works Clearinghouse Review

NATIONAL INSTITUTE FOR DIRECT INSTRUCTION

Jean Stockard and Timothy W. Wood

July, 2012

Reading Mastery and Learning Disabled Students

A Comment on the What Works Clearinghouse Review

A recent report by the What Works Clearinghouse (2012) examined two studies of the use of *Reading Mastery* with learning disabled students and concluded that it had "no discernible effects on reading comprehension and potentially negative effects on alphabetics, reading fluency, and writing." This conclusion is in stark contrast to dozens of studies of *Reading Mastery* and other elements of the Direct Instruction (DI) corpus of material. This research has consistently found strong positive effects of the programs on academic achievement for students of all ability levels.

The Office of Research and Evaluation at the National Institute for Direct Instruction examined the WWC's report and the research articles used to develop its findings and found very serious problems with its conclusions. The WWC analysis was based on only two articles. One (Cooke, Gibbs, Campbell, & Schalvis, 2004) compared two very similar Direct Instruction programs, *Reading Mastery* and *Horizons*, and found that students in both programs made gains over the academic year that were significantly greater than those made by students in national and state level populations. Gains in both programs were similar, leading the WWC to conclude that *Reading Mastery* was no better than its comparison program. They ignored the fact that students performed significantly better than the national or state norms or that the comparison program was similar on all but a very few characteristics, essentially a modified version of *Reading Mastery*. The second article reviewed (Herrera, Logan, Cooker, Morris, & Lyman, 1997) involved two groups of students, both of whom appear to have received *Reading Mastery* as part of the schools' "usual and customary school day curriculum." One group of students also received 45 minutes of supplemental phonemic related instruction, from their regular classroom teachers, that involved motor activities to accompany practice of skills. Not surprisingly, the group receiving the additional instruction had significantly larger gains than those who did not have additional learning time. Despite these differences in exposure and the fact that both groups appear to have had *Reading Mastery* as their usual reading curriculum, the WWC used these results to suggest that *Reading Mastery* could have potentially negative effects. The first section of this report provides additional details on each of these studies.

The WWC uses what can be termed an "exclusive approach," putting a great deal of emphasis on "internal validity" and focusing only on studies that incorporate a select set of characteristics, involving the nature of the design, subject assignment, testing procedures, setting, and time of publication. In short, the WWC appears to be searching for a perfectly designed experiment, implicitly suggesting that a perfectly designed experiment will give the

best results. Not surprisingly, very few studies pass its screening procedures. In contrast, meta-analyses and literature reviews often take a much broader view of the literature, looking at the totality of information that is available. While a WWC report may find only one or two studies that meet its criteria, the meta-analyses and other reviews of Direct Instruction typically report on dozens of studies. The second section of this document reports the results of several meta-analyses and literature reviews of DI programs, focusing only on those that report results separately for special education students.

Over the last four and one-half decades many studies have appeared regarding the efficacy of Direct Instruction. The National Institute for Direct Instruction is compiling and abstracting this vast bibliography. The final section of this document provides citations to and abstracts of many individual studies that pertain to the content of the WWC report. The full NIFDI bibliography is available on-line at www.nifdi.org/research.

The WWC Report

A valid test of the efficacy of any intervention requires that the treatment received by the control group and the treatment group be distinct. The WWC clearly includes this criterion in its list of reasons that a study may fail to meet WWC standards, noting that a study "may fail to meet WWC evidence standards" if "the measures of effect cannot be attributed solely to the intervention—the intervention was combined with another intervention" (What Works Clearinghouse, 2011, p. 15). The WWC conclusions regarding *Reading Mastery (RM)* and learning disabled students were based on analyses of two studies, both of which should have been excluded from analysis based on this criterion. This section summarizes the designs of these studies and explains why they do not provide a valid test of the efficacy of *RM*.

A Comparison of Two Direct Instruction Programs

The first study reviewed by the WWC (Cooke, et al., 2004) is titled "A Comparison of *Reading Mastery Fast Cycle* and *Horizons Fast Track A—B* on the Reading Achievement of Students with Mild Disabilities." It examined gains in reading achievement over the school year of learning disabled students studying with one or the other of these programs and compared these gains with national norms and with each other.

Reading Mastery and *Horizons* are both part of the DI corpus of curricular material and were developed and authored by S. Engelmann and colleagues. The first three pages of the Cooke, et al. paper describe the close alignment of the programs, stating that "*Horizons* shares many important design features with Reading Mastery" (p. 140). Seventeen very specific points of design similarity are cited (p. 140, citing Engelmann, 2000), and only five points of difference are described (pp. 140-141). The major difference between the programs is the prompts used to help students learn the relationships between letters and sounds, primarily involve differences with orthography at the early stages. In other words, *Horizons* may be seen as a modified version of *Reading Mastery.* Thus the study reported by

Cooke and associates did not compare *Reading Mastery* with another type of reading program. It compared *Reading Mastery* with a slightly altered version of the same program.

The authors report results using norm-based standard scores, comparing students' standing to other students in their grades at the testing time. Such a norm-based comparison allows the calculation of effects relative to the total population, for, as explained by the authors, "if students progressed at an average rate across the course of the study, they would show no change in these scores" (p. 146). They found that, relative to other students in the nation and state, students in both groups made statistically greater gains over time. These results appeared on all measures used. Effect sizes ranged from .09 to .36 for sub-tests of the WJ-R and .71 to .78 for subtests of a state assessment. (Effect sizes of .25 or higher have traditionally been seen as educationally important (Tallmadge, 1977).) The gains were similar for students in both *Horizons* and *Reading Mastery*.

The WWC report focused on the lack of difference between *Horizons* and *Reading Mastery*, rather than a) students in both programs had significantly stronger gains than students in national or state level comparison groups and b) the programs were essentially equivalent in design. An appropriate reading of the article would be that it provided confirmation of the efficacy of the Direct Instruction approach in teaching reading. Instead, the WWC chose to report the study as showing "no discernible effects" of *Reading Mastery*.

A Comparison of Reading Mastery and Additional Intervention

The second article reviewed by the WWC (Herrera, et al., 1997) compared growth over time in achievement of students who received only *Reading Mastery* with those who received an extra 45 minutes daily instruction involving "sensorimotor, procedural activities patterned to phonemes, graphemes and phonographs" (p. 71). *Reading Mastery* was described as "the usual and customary prescribed curriculum for the school day" (p. 78). Six classrooms of students receiving only *Reading Mastery* were treated by the WWC as the intervention group for their review.

The intervention group in Herrera, et al.'s analysis was comprised of students in five other randomly selected classes within the same district. The teachers in these classrooms were trained, in a 60-hour program, to use a sensorimotor procedure to teach reading "for forty-five minutes per school day as a supplement to the usual and customary school day curriculum." An appendix to the article describes sample activities in the approach, which is termed the "*Stabilized Learning System*." It describes how "visual representations of words are simultaneously created and expressed via movement" (p. 85). Students were given explicit instruction in various phonemic elements accompanied by associated movements, designed to help "ensure stabilization of the movement and represented processes" (p. 85). Movements involved hand motions or activities on floor grids, where students would "perform all of the above [phonemic] activities in simultaneous-oral fashion (saying the

word, then the letter names while 'moving' them or 'stepping' them, then saying the word at the end) in simultaneous-sounding fashion."

The authors described this approach as "implicit," because of the centrality of sensorimotor functions rather than emphasis on the rule based elements of learning to read. While described as implicit, the program appears to be very different than the "implicit" learning approach used by the so-called "whole language" movement, for it used explicit reinforcement of learning related to "phonemes, graphemes, and phonographs" (p. 71). In addition, all of the teachers assigned to the program used *Reading Mastery* in their classrooms as part of the district's "usual and customary prescribed curriculum" (p. 78), and it would be expected that the skills and procedures used there would extend to their work in the sensorimotor component. Teachers were well trained in the intervention, and were reported to have "received sixty hours of instruction involving these [sensorimotor] procedures" (p. 78).

For their analysis the WWC treated this intervention group as the control condition and the group that did not receive *Stabilized Learning (SL)*, only having *Reading Mastery*, as the intervention or treatment condition, implicitly assuming that the two treatments were independent of each other. However, a careful reading of the article casts doubt on this assumption. As noted above, *Reading Mastery* was "the usual and customary prescribed curriculum for the school day" (p. 78). The authors reported that the sensorimotor intervention "was implemented for forty-five minutes per school day as a supplement to the usual and customary school day curriculum" (p. 78). While the *SL* program "was integrated into the school day with no extra time allotment required" we have not been able to find any place within the article that explicitly states that the students using the *SL* system were not exposed to *RM*. If *SL* were indeed, as stated, a "supplement to the usual and customary prescribed curriculum," they would have also received *Reading Mastery*.

Thus, the analysis appears to have compared gains in reading skills of students exposed to *Reading Mastery* to those exposed to *Reading Mastery* plus an additional 45 minutes of instruction in phonics-related skills. In short, the design does not appear to provide a test of the efficacy of *Reading Mastery* but, instead, a test of the extent to which 45 minutes of additional teaching in phonemic related matters can supplement the gains made by *Reading Mastery*. As would be expected, the students receiving this additional instruction had stronger gains over time. An internet search using Google Scholar found no other scholarly studies of this program. If, in fact, the *Stabilized Learning System* did provide superior results to *Reading Mastery* there should be other tests and reports of this finding.

Conclusion

When examined carefully, neither of the studies cited by the WWC as demonstrating the ineffectiveness of *Reading Mastery* actually does so. In one study (Cooke, et al., 2004), students studying with *Reading Mastery* had significantly stronger gains than those in

national and state populations. The WWC ignored these results and focused instead on the lack of difference between *Reading Mastery* students and those in another Direct Instruction program, claiming that this demonstrated a lack of effectiveness. The other study appeared to use nonequivalent interventions, with both groups exposed to *Reading Mastery* as part of the usual curriculum and the comparison group having significant extra instructional time designed to reinforce the phonemic understandings gained in the schools' usual and customary reading curriculum, *Reading Mastery*.

A large body of literature has documented the efficacy of Direct Instruction curricula, including many studies that looked at special education populations, including learning disabled students. The next section describes summaries of this work.

Meta-Analyses and Literature Reviews

The literature related to Direct Instruction is very large, with publications appearing over the last fifty years. A number of authors have reviewed the literature and summarized the major findings. These summaries build upon the epistemological tradition found in the classic literature on experimental design (Campbell and Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002) and a very large methodological literature within the social sciences that stresses the importance of repeated studies, or replications, of analyses (See Stockard, 2011 for a summary). Writers in this epistemological tradition emphatically state that there can be no "perfect" experiment and that it is essential to examine effects of interventions over a broad range of settings and conditions to amass knowledge.

The authors of the literature reviews and meta-analyses described below have systematically reviewed dozens of studies of Direct Instruction conducted in many different locales and with a wide variety of subjects. Four of the reports were systematic literature reviews; four were quantitative meta-analyses. All of the reviews concluded that Direct Instruction curricula are highly effective. Their methods and conclusions are summarized below. Only reports that included specific mention of results pertaining to special education students are included. When results regarding *Reading Mastery* and/or learning disabilities or special education students were reported separately, they are noted. The systematic literature reviews are listed first, followed by the meta-analyses. In each section reports are listed alphabetically.

Systematic Literature Reviews

Systematic literature reviews provide narrative summaries of the literature, detailing the findings and noting key similarities and differences in the results.

1) Kinder, D., Kubina, R., & Marchand-Martella, N. (2005), Special education and Direct Instruction: An effective combination. *Journal of Direct Instruction*, 5 (1), 1-36. Also

distributed by SRA/McGraw-Hill as *Special Education and Direct Instruction: An Effective Combination.*

This article summarizes 37 studies that used Direct Instruction materials with students with high-incidence disabilities, ranging in age from pre-school to high school, and 8 studies that used DI materials with students with low-incidence disabilities. Of the 45 studies, over 90 percent found positive effects for the Direct Instruction programs. Seven of the studies involved school-aged populations with high incidence disabilities using *Reading Mastery* or *DISTAR* (the precursor to *Reading Mastery*) as compared to another program. Learning disabilities are typically considered a high incidence condition and, thus, these studies are most relevant to considering the conclusions of the recent WWC report. (They were Branwhite, 1983; Kuder, 1990; Marston, Deno, Kim, Diment, & Rogers, 1995; O'Connor, Jenkins Cole, & Mills, 1993; O'Connor & Jenkins, 1995; Richardson, DiBenedetto, Christ, Press, & Winsberg, 1978; and Stein & Goldman, 1980). Only one of these studies (Marston, et al., 1995) found that the comparison group outperformed the students in the DI program. One of the studies (Glang, Singer, Cooley, & Tish, 1992) looked at the impact of *Reading Mastery* for students with low incidence disabilities using a single subject design (two students with traumatic brain injury). Results with this study, as well as with the other 7 studies of students with low incidence conditions, indicated positive effects.

2) Przychodzin, A.M., Marchand-Martella, N., Martella, R.C., & Azim, D. (2004). Direct Instruction Mathematics Programs: An Overview and Research Summary. *Journal of Direct Instruction, 4 (1)*, 53-84.

This article reviews 12 studies of DI math programs published since 1990. All but one of the studies showed positive results. Four of the studies looked at students with disabilities (Glang, Singer, Cooley, & Tish, 1991; McKenzie, Marchand-Martella, Martella, & Moore, 2004; Sommers, 1991; and Young, Baker, and Martin,1990), and all showed results in favor of Direct Instruction.

3) Przychodzin-Havis, A.M., Marchand-Martella, N., Martella, R.C., Miller, D.A., Warner, B.L., & Chapman, S. (2005). An Analysis of Corrective Reading Research. *Journal of Direct Instruction, 5 (1)*, 37-65.

This article reviewed 28 published studies of the Direct Instruction *Corrective Reading* program. Twenty-three of the studies used teachers to deliver instruction (4 in general education classrooms, 12 in special education, and 7 in alternative education settings). Five studies looked at implementations by paraprofessionals or peer instructions in general education or special education settings. Twenty six of the 28 studies (92.8%) found positive results. Only one study found greater gains with another intervention. Similar results

appeared with different types of assessments (e.g. standardized tests or curriculum-based measures), in different settings, with different types of instructors, and with different research designs.

4) Schieffer, C., Marchand-Martella, N.E., Martella, R.C., Simonsen, F.L., & Waldron-Soler K.M. (2002). An analysis of the *Reading Mastery* Program: Effective components and research review. *Journal of Direct Instruction, 2(2)*, 87-119.

This article reviewed twenty-one studies of *Reading Mastery* that compared its use to that of another program. Four of these studies looked at general education populations, eight focused on general education remedial readers, and nine looked at special education students. Results in fourteen of these studies (67%) favored RM, other programs were favored in three (14%), and there were no differences in the remainder. Interestingly, they included the Herrera, et al. (1997), study noted above as one of the three studies that favored other programs.

Meta-Analyses

Meta-analyses summarize study results using quantitative methods and, usually, more restrictive criteria for selecting studies than the literature reviews described above. Meta-analyses typically use an "effect size" metric to summarize the results, averaging the effects over numerous studies. The effect size is simply the difference between two average scores divided by the common standard deviation. In other words, it describes a difference between two groups in standard deviation units (a standard metric). A number of years ago Tallmadge suggested that an effect of .25 standard deviations was educationally important (Tallmadge, 1977). All of the meta-analyses of Direct Instruction programs have found average effect sizes far surpassing this level.

1) Adams, G.L. & Engelmann, S. (1995). *Research on Direct Instruction: 25 Years Beyond Distar.* Seattle, WA: Educational Achievement Systems.

Adams (who conducted the meta-analysis) limited his review to studies that met the following criteria, which are similar to those used in other analyses:

- The study included a comparison group.
- Pretest scores were available and showed no significant differences between the groups.
- Descriptive statistics were provided for both groups (means, standard deviations, and sample sizes).

- The study lasted for more than one session.
- The DI intervention was not combined with another, incompatible program.
- More than one subject was in each group no single-subject designs).
- The study involved an entire program, not separate components of DI.
- The study used Direct Instruction programs developed by Engelmann and associates (not the so-called "di").

Thirty-seven studies were found that met these criteria, and these studies included 173 comparisons of mean scores. Of these comparisons, 87.3% favored the Direct Instruction condition, and 12% favored the comparison program (50% would be expected by chance). Sixty-four percent of the comparisons were statistically significant in favor of DI, while 1.2% were statistically significant in favor of the comparison program (5% would be expected by chance). Of the 173 comparisons, 101 involved students in special education. The average effect size across these 101 studies was .76. When the effect sizes were aggregated to the level of each study, the average effect size across the 21 studies including special education students was .90.

2) Coughlin, C. (2011). *Research on the Effectiveness of Direct Instruction Programs: An Updated Meta-Analysis*, Poster presented at the Annual Meetings of the Association for Behavior Analysis International, May, 2011, NIFDI Technical Report 2011-4.

Couglin limited this meta-analysis to randomized control studies. Twenty studies with 95 separate comparisons were identified. Nine of the studies involved students with disabilities. Over a third of the studies examined *Reading Mastery* or its precursor *DISTAR*. The average effect size across all the studies was .66. Effect sizes were similar for studies of general education students (.69) and those with special education students (.71). They were slightly smaller, on average, for reading and other programs (.56 and .54, respectively) than for language (.81) and mathematics (1.03). The effect sizes were regressed on methodological characteristics of the studies including sample size, whether or not fidelity was monitored, the length of the intervention, the type of test used to measure achievement, and the year the study was conducted. The only statistically significant relationship was with sample size, with smaller effects found for larger samples.

3) Hattie, John A.C. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London and New York: Routledge.

Hattie used a meta-analysis technique to review the results of over 800 meta-analyses of research studies related to student achievement. He summarized the results of four meta-analyses that included DI, incorporating 304 studies, 597 effects and over 42,000 students. He found that the average effect size associated with DI was .59 and noted that the positive results were "similar for regular (*d*=.99) and special education and lower ability students

(*d*=0.86), ... [and] similar for the more low-level word-attack (*d*=.64) and also for high-level comprehension (*d*=.54)" (pp. 206-207). No other curricular program showed such consistently strong effects with students of different ability levels, of different ages, and with different subject matters.

4) White, W.A.T. (1988). A meta-analysis of the effects of Direct Instruction in special education, *Education and Treatment of Children, 11 (4),* 364-374.

Twenty-five studies of the effectiveness of Direct Instruction programs with special education students were examined using quantitative meta-analysis techniques. The analysis was limited to studies with a comparison group of students assigned to a non-DI treatment prior to the start of the intervention. Fifty-three percent of the outcomes significantly favored DI (5% would be expected by chance). The average effect size across all comparisons was .84 standard deviations. Similar effects appeared with different conditions, ages, skill areas, and type of outcome measure.

Scholarly Studies of Direct Instruction

This section includes citations to and abstracts of a large number of individual research studies regarding the impact of the Direct Instruction curricula on students' reading achievement. The first section focuses only on studies that examined the impact of *Reading Mastery*, or its precursor *DISTAR*, on the reading achievement of students with disabilities, thus paralleling most closely the focus of the WWC report. The second section provides summaries of a number of articles that focus on students with disabilities using other elements of the Direct Instruction body of curricula. The third section examines published data from school districts around the country that have used *Reading Mastery* with special education students, reporting the magnitude of change using effect sizes.

Reading Mastery and Disabled Students

This section only includes studies that report results of the use of Reading Mastery and outcomes with special education students.

Frankhauser, M. A., Tso, M. E., & Martella, R. C. (2001). A comparison of curriculum-specified reading checkout timings and daily 1-minute timings on student performance in Reading Mastery. *Journal of Direct Instruction, 1(2),* 85-96.

This study gathered data on words read per minute and number of errors on curriculum-specified reading checkouts; in addition, the number of readings needed to reach criteria on these checkouts with and without the addition of a daily timing was recorded. Four second- and third-grade students with reading disabilities who received reading instruction using *Reading Mastery II* or *Fast Cycle* (Engelmann & Bruner, 1995) participated in this study.

Comparisons were made of each student's progress across phases. Results showed no overall change in mean words read per minute, errors per minute, or number of timings to meet curriculum-specified criteria at reading checkouts for all students. These results suggest that the systematic practice and curriculum-specified reading checkouts within the Reading Mastery lessons provide the structure needed for students in need of reading remediation to make consistent progress in reading.

Gersten, R. M., & Maggs, A. (1982). Teaching the general case to moderately retarded children: Evaluation of a five year project. *Analysis and Intervention in Developmental Disabilities, 2,* 329-343.

This study examined the effects of a five-year program for pre-adolescent students with low-moderate intelligence levels that utilized the Direct Instruction programs, DISTAR Language and DISTAR Reading. Students were administered the Stanford-Binet as a pretest and posttest measure and student gains were compared to the standardization sample of the Stanford-Binet. Additionally, the Peabody Picture Vocabulary Test was administered as a posttest to examine concurrent validity. Results indicate that students who received the DISTAR program made statistically significant gains when compared to the norm sample.

Kuder, S. J. (1990). Effectiveness of the DISTAR reading program for children with learning disabilities. *Journal of Learning Disabilities*, *23(1),* 69-71

The reading achievement of students with learning disabilities (N=24) who received DISTAR instruction was compared to that of similar students using basal reader materials. No statistically significant differences were found between groups in overall reading scores, but the DISTAR group outperformed the basal reading group on the Word Attack subtest.

O'Connor, R. E., Jenkins, J. R., Cole, K. N., & Mills, P. (1993). Two approaches to reading instruction with children with disabilities: Does program design make a difference? *Exceptional Children, 59(4),* 312-323.

This study analyzed 2 approaches to reading instruction for children with disabilities in relation to reading achievement. The study compared Direct Instruction Reading Mastery I and II, and Addison Wesley's Meet the Superkids and The Superkids' Club. Results indicated no significant difference between the instructional programs at the end of treatment or one year later. The only notable difference was for children who made advanced progress, where the Direct Instruction group registered larger reading gains on the California Achievement Test (CAT) sound recognition subtest and the Peabody Individual Achievement Test (PIAT) reading recognition and spelling subtests.

Riepl, J. H., Marchand-Martella, N. E., & Martella, R. C. (2008). The effects of Reading Mastery Plus on the beginning reading skills of students with intellectual and developmental disabilities. *Journal of Direct Instruction, 8(1),* 29-39.

This study explored the effects of "Reading Mastery Plus" in grades K-2 at a Title 1 school with six students with intellectual and developmental disabilities (IDD). Two students in each of grades kindergarten, 1, and 2 participated. "Diagnostic Indicators of Basic Early Literacy Skills (DIBELS)" spring fluency probes were used to assess changes in students' beginning reading skills and risk status. Results indicated students made large overall gains on fluency measures. Additionally, there were several improvements in the "DIBELS" risk status categories.

Somerville, D. E., & Leach, D. J. (1988). Direct or indirect instruction?: An evaluation of three types of intervention programme for assisting students with specific reading disabilities. *Educational Research, 30(1),* 46-53.

Three intervention programs (psychomotor, self-esteem enhancement, and Direct Instruction) were conducted with 40 children who had reading difficulties. The direct instruction program resulted in the greatest gains. Post-intervention questionnaires completed by subjects, parents, and teachers indicated that perceived success differed significantly from measured success.

Stein, C., & Goldman, J. (1980). Beginning reading instruction for children with minimal brain dysfunction. *Journal of Learning Disabilities*, *13(4),* 52-55.

This study examined the differential effect of two reading programs, DISTAR and Palo Alto Reading Program, on early elementary aged children diagnosed with minimal brain dysfunction. Results indicated that the DISTAR program was more effective in producing larger mean gains in reading achievement, as measured by the Peabody Individual Achievement Test (PIAT).

Stockard, J. (2008). *Reading Achievement in a Direct Instruction School and a "Three Tier" Curriculum School, NIFDI Technical Report 2008-5*. Eugene, Oregon: National Institute for Direct Instruction

This report examines data from two schools within the same Oregon school district. One school adopted the *Reading Mastery* Direct Instruction program as the core reading curriculum for all primary children, while the other used a "three tiered" model, occasionally employing DI for students that teachers felt would benefit from the instruction. There were

almost equal numbers of students from each school and there were no significant differences between the schools in the students' eligibility for free or reduced lunch, their racial-ethnic characteristics, or their special education designation. A slightly modified version of DIBELS was used to assess achievement gains over the time span of the study. Data were available for 2 cohorts of students who were in the schools from kindergarten through third grade. Students in the Direct Instruction School had statistically significant higher gains in Oral Reading Fluency (ORF) from first through third grade than students in the Control School. These differences were especially marked for students in special education. By the end of third grade, the effect size for ORF for the total group of students was .42, while the effect size for special education students was .73. By the end of the study the reading scores of special education students in the DI school were essentially equivalent to those of the general education students in the non-DI school. The odds of a child in the DI school having levels of oral reading fluency that would indicate a possibility of academic failure were less than half that of a child in the control school. Similar results occurred with the measures of onset recognition fluency, phonemic segmentation fluency, and nonsense word fluency, consistently favoring the Direct Instruction students.

Zayac, R. (2008). Direct Instruction reading: Effects of the Reading Mastery Plus – Level K curriculum on preschool children with developmental delays. Dissertation from Auburn University.

This study investigated the effects of the Reading Mastery Plus - Level K program on preschool children with developmental delays. Findings indicated that preschool-aged children both with and without developmental delays were able to acquire beginning reading skills. While the research design inhibited the identification of any functional relationships between the Reading Mastery Plus - Level K program and the participants' reading gains, the data showed that young children with developmental delays can acquire skills that are necessary to begin reading. The author suggests that this is an important finding, especially considering the fact that the number of children with autism spectrum disorder is increasing.

Studies of other DI Curricula and Students with Disabilities

While *Reading Mastery* is typically used with students in grades K-3, *Corrective Reading* is the program prescribed for older students. Thus, many of the studies of Direct Instruction curricula with older special education students involve *Corrective Reading*. Some also involve exposure to other programs. This section includes citations to and abstracts of a number of studies involving special education students who were exposed to Direct Instruction curricula other than *Reading Mastery*. Because all of the Direct Instruction programs are built using the same techniques and criteria for effective instruction, results

from all of these studies are important to consider. As with the set of works reviewed above, results were overwhelmingly in favor of the Direct Instruction curricula.

Anderson, D. M. (2002). Using Reasoning and Writing to teach writing skills to students with learning disabilities and behavioral disorders. *Journal of Direct Instruction, 2(1),* 49-55.

This study examined the effectiveness of the Direct Instruction writing program, *Reasoning and Writing*, with 10 fourth and fifth grade students with learning disabilities and/or behavioral disorders. Instruction with Reasoning and Writing occurred for six weeks, and pretest and posttest scores on the Test of Written Language-2 (TOWL-2) were used to indicate program effects. As a group, results indicated that students made both statistically and educationally significant gains overall on the TOWL-2.

Benner, G. J. (2007). The relative impact of remedial reading instruction on the basic reading skills of students with emotional disturbance and learning disabilities. *Journal of Direct Instruction, 7(1),* 1-15.

This study investigated the effects of a Direct Instruction reading program, *Corrective Reading*, on the reading achievement of elementary and middle school students with emotional disturbance (ED) and learning disabilities (LD). Comparison students receiving non-DI instruction were selected based on demographics. Results indicated that students in the *Corrective Reading* program made significantly greater gains than comparison students; additionally, students with ED were more responsive to Corrective Reading than students with LD.

Benner, G. J., Kinder, D., Beaudoin, K. M., Stein, M., & Hirschmann, K. (2005). The effects of the “Corrective Reading Decoding” program on the basic reading skills and social adjustment of students with high-incidence disabilities. *Journal of Direct Instruction, 5(1),* 67-80.

The purpose of this study was to examine the effects of the "Corrective Reading Decoding B1" program on the basic reading skills, social adjustment, and treatment responsiveness of elementary and middle school students with high-incidence disabilities (N = 51). Students were provided an average of 3 40-45 minute lessons per week over the course of nearly 4 months. Statistically and educationally significant improvements were found between students who received "Corrective Reading Decoding Level B1" (n = 28) and those in the comparison condition (n = 23) on measures of basic reading skills and social adjustment. A large percentage of students who experienced below average basic reading skills (i.e., nonresponders) at pretest performed in the average range at posttest (i.e., responders).

Booth, A., Hewitt, D., Jenkins, W., & Maggs, A. (1979). Making retarded children literate: A five year study. *The Australian Journal of Mental Retardation, 5(7),* 257-260.

The Distar model of reading and language instruction was shown to be effective in developing the reading and language skills of 12 retarded Australian children beyond the usual expectations for such children, and in fostering mastery of most of the basic literacy skills. A battery of measures (e.g., Peabody Picture Vocabulary Test) were administered to 12 subjects. At the beginning of the study they averaged 10 years of age and had measured IQs ranging from 35 to 55. Instruction occurred with the Distar language program, and involved all levels of the Distar language and reading programs. Results indicated that students demonstrated mastery of most of the basic literacy skills.

Bracey, S., Maggs, A., & Morath, P. (1975). Teaching arithmetic skills to moderately mentally retarded children using direct verbal instruction: Counting and symbol identification. *Australian Journal of Mental Retardation, 3,* 200-204.

This study analyzed the impact of Direct Instruction's arithmetic program and behavioral techniques to teach arithmetic skills to a group of 6 institutionalized moderately mentally retarded children (IQs approximately 35-50). The aim of the study was to determine whether the subjects made significant gains in 4 types of arithmetical skills: object counting, making lines from numerals, the meaning of plus, and increment additions. Results revealed significant improvement ($p < .05$) in each of the skill areas.

Campbell, M. (1981). A study of Corrective Reading as an effective and appropriate program for reading-disabled, learning-handicapped secondary students. Report presented to Faculty of School of Education. San Diego, CA: San Diego State University.

The study showed that the *Corrective Reading Program* was successful with upper elementary children identified as having mild learning disabilities and junior high school underachievers.

Carnine, D., Engelmann, S., Hofmeister, A., & Kelly, B. (1987). Effects of instructional design variables on vocabulary acquisition of LD students: A study of computer-assisted instruction. *Journal of Learning Disabilities, 20,* 206-213.

This study compared two computer-assisted instructional vocabulary programs used with 25 learning disabled high school students. Results indicated that the Direct Instruction program, which utilized smaller teaching sets and cumulative review exercises, was more

effective in helping students achieve mastery than the program using a large teaching set and no cumulative reviews.

Carnine, D. (1989). Teaching complex content to learning disabled students: The role of technology. *Exceptional Children, 55,* 524-533.

The article documents the use of Direct Instruction procedures with technological adjuncts to teach learning disabled students. Examples are given of effective technology use to teach earth science/chemistry, fractions, health, reasoning skills, and vocabulary, and findings from several studies are briefly reviewed. The article documents how such a comprehensive intervention reduces performance differences between students with learning disabilities and their peers, while using technology to minimize, or even reduce, the demands placed on the teacher.

Clark, D. S. (2001). Components of effective reading instruction for reading disabled students: An evaluation of a program combining code- and strategy-instruction. Unpublished doctoral dissertation, University of Toronto.

This study randomly assigned 44 disabled readers, seven to ten years of age to an experimental treatment intervention, *Phono-Graphix*, or to the Direct Instruction program, *Corrective Reading*. Both programs placed emphasis on decontextualized word identification training with instructional focus at the level of the phoneme. Each group received a mean of 10.5 hours of formal, one-to-two instruction and an equivalent number of hours of assigned homework over a 7.5-week period over the summer. No reliable pretest-posttest gains were demonstrated for either program on any of the standardized measures of word identification. A large treatment effect was found for the Phono-Graphix group, however, on an experimental measure of word identification. Further analysis revealed that this treatment effect was attributable to large, reliable improvement in reading orthographically irregular words.

Cole, K. N., Mills, P. E., & Dale, P. S. (1989). Comparison of the effects of academic and cognitive curricula for young handicapped children one and two years post program. *Topics in Early Childhood Special Education, 9(3),* 110-127.

Follow-up testing of students who had participated when ages 3-7 in two early intervention programs, Direct Instruction and Mediated Learning, was conducted at 1 year and 2 years post-intervention. Children from both programs generally maintained or increased cognitive and academic skills following intervention. Trends suggesting differential effects for subscales of the Peabody Individual Achievement Test and the Stanford-Binet (fourth

edition) were noted in the first-year follow-up. The differential effects decreased by the second year post-intervention, although the overall benefits of early intervention appeared to continue.

Dale, P. S. & Cole, K. N. (1988). Comparison of academic and cognitive programs for young handicapped children. *Exceptional Children, 54(5),* 439-447.

Two highly contrasting models of preschool education for mildly handicapped children were compared. Direct Instruction led to greater gains on the Test of Early Language Development and the Basic Language Concepts test. Mediated Learning led to greater gains on the McCarthy Verbal and Memory scales and Mean Length of Utterance measure.

Darch, C. & Kameenui, E. J. (1987). Teaching LD students critical reading skills: A systematic replication. *Learning Disability Quarterly, 10,* 82-91.

Contrasted 2 approaches to teaching learning disabled (LD) students 3 critical reading skills, using 25 LD students in Grades 4–6. Subjects were randomly assigned to treatment groups using Direct Instruction or discussion with a workbook. The former featured specific rules and strategies to detect instances of faulty arguments, whereas in the latter group subject involvement was encouraged through discussions on how to use critical reading skills. Both groups received 40-minute lessons for 12 consecutive school days. Three dependent measures designed to assess students' knowledge of the three critical reading skills were administered. Subjects in the direct instruction group were found to significantly outperform their counterparts in the discussion/workbook group on each measure.

Darch, C. & Simpson, R. G. (1990). Effectiveness of visual imagery versus rule-based strategies in teaching spelling to learning disabled students. *Research in Rural Education, 7(1),* 61-70.

Among 28 upper elementary learning-disabled students in a summer remedial program, those that were taught spelling with explicit rule-based, Direct Instruction, strategies out-performed students presented with a visual imagery mnemonic on unit tests, a posttest, and a standardized spelling test.

Darch, C., Eaves, R. C., Crowe, D. A., Simmons, K., & Conniff, A. (2006). Teaching spelling to students with learning disabilities: A comparison of rule-based strategies versus traditional instruction. *Journal of Direct Instruction, 6(1),* 1-16.

This study compared the effects of the Direct Instruction program, Spelling Mastery, and traditional spelling instruction on spelling achievement of special education elementary students receiving services for a learning disability. Unit tests, a standardized spelling test, a sentence-writing test, a transfer test, and a maintenance test were administered to all students. Instruction occurred daily, for 30 minutes, for four consecutive weeks. Four different word types (i.e., regular, morphological, spelling rule, and irregular) were introduced as instruction progressed. After receiving instruction in one of the instructional groups, the students were compared on scores from unit tests, a standardized test, a sentence-writing test, a transfer test, and a maintenance test. Results indicate that students in the Spelling Mastery group significantly outperformed comparison students on words of high predictability (regular, morphological, and spelling-rule words).

Darch, C., & Gersten, R. (1985). Effects of teacher presentation rate and praise on LD students' oral reading performance. *British Journal of Educational Psychology, 55,* 295-303.

Four learning disabled students were taught beginning reading with a Direct Instruction program. The teacher systematically varied the rate of instructional presentation (rapid pace ***vs.*** slow pace) and frequency of praising (praise ***vs.*** no praise). A modified reversal design was used to evaluate both the isolated and interactive effects of the individual instructional components. Experimentation lasted 25 school days. The results indicated that increases in teacher presentation rate and increased use of praise led to improvement in levels of on-task behavior and correct student responding. Each of the variables in isolation had an effect. Praise appeared to be more powerful than presentation rate. The combination of the two variables had the most powerful effect. These findings extend the research of earlier studies to learning disabled students and confirm other studies regarding the presentation style used in Direct Instruction programs.

Darch, C. (1990). Comprehension instruction for high school learning disabled students. *Research in Rural Education 5,* 43-49.

This article presents an instructional approach for helping learning disabled high school students comprehend content area instruction in regular classrooms that combines the use of Direct Instruction and advanced organizers. It also identifies and discusses general methods teachers can use to teach learning disabled students successfully in regular classrooms.

Flores, M. M., & Ganz, J. B. (2007). Effectiveness of Direct Instruction for teaching statement inference, use of facts, and analogies to students with developmental

disabilities and reading delays. *Focus on Autism &Other Developmental Disabilities, 22(4),* 244-251.

This study investigated the effects of a Direct Instruction (DI) reading comprehension program implemented with students who had developmental disabilities, including autism spectrum disorders (ASD) and reading delays. There is little research in the area of reading comprehension for students with ASD and no research as to the effectiveness of reading comprehension DI. A multiple-probe-across-behaviors design was employed. A functional relationship between Direct Instruction and reading comprehension skills and behaviors was demonstrated across all behavioral conditions and across students. Results and their implications are discussed.

Flores, M., Shippen, M., Alberto, P. & Crowe, L. (2004). Teaching letter-sound correspondence to students with moderate intellectual disabilities. *Journal of Direct Instruction, 4(2),* 173-188

This study examined the efficacy of systematic and explicit instruction in phonic decoding for 6 elementary students with moderate intellectual disabilities. A multiple probe across behaviors with embedded changing conditions design was employed in order to analyze the effect of Direct Instruction on teaching the following skills with regard to consonant-vowel-consonant (CVC) words: letter-sound identification, continuous sound blending, sounding out, and decoding (sounding out then telescoping). Results indicate that the majority of students demonstrated an understanding of letter-sound correspondence and decoding basics following intervention.

Franklin, M. E., Little, E., & Teska, J.A. (1987). Effective teaching strategies used with the mildly handicapped in the mainstream. *Focus on Exceptional Children, 20,* 7-11.

The article explores evidence of the benefits and limitations of four instructional strategies (Direct Instruction, classroom management, cooperative grouping, and metacognition) on the successful integration of mildly handicapped children in the mainstream.

Ganz, J. & Flores, M. (2009). The effectiveness of Direct Instruction for teaching language to children with autism spectrum disorder. Indentifying materials. *Journal of Autism and Developmental Disorders, 39,* 75-83.

This study examined effects of the Direct Instruction program, *Language for Leaning*, on improving language ability of elementary students with autism spectrum disorders. A single-subject changing criterion design was employed. Probes were given 2-3 times a week to measure students' ability in identifying items made from different materials. Percentage of

non-overlapping data was calculated to supplement visual analysis when determining the effectiveness of the intervention. Results indicate that a functional relation was demonstrated between the implementation of *Language for Learning* and an increase in correctly naming presented items; additionally, the percentage of non-overlapping data was at least 90% for all participants, suggesting that the intervention was effective at increasing the number of correct responses over time.

Gersten, R. (1985). Direct Instruction with special education students: A review of evaluation research. *Journal of Special Education, 19*, 41-50.

Studies evaluating the effectiveness of direct instruction curricula and teaching procedures are reviewed and, in some instances, critiqued. The six studies indicate that direct instruction tends to produce higher academic gains for handicapped children than traditional approaches. They also suggest that some of the more subtle principles of direct instruction—such as insistence on complete (rather than partial) mastery of each step in the learning process—are important. The author contends that future research should more carefully measure program implementation and begin to look more precisely at specific instructional variables. Suggestions for appropriate research designs are presented.

Hempenstall, K. (1997). The effects on the phonological processing skills of disabled readers of participating Direct Instruction reading programs. Australian Digital Theses Program, RMIT University Library.

This thesis examines the effects of two Direct Instruction reading programs on the phonological processes of students with teacher-identified reading problems in nine northern and western Melbourne primary schools. In the first study students (mean age 9.7 years years) were assigned to the treatment condition or to wait-list comparison groups. Based on the results of a program placement test of rate and accuracy, students were assigned to one of two entry points into the *Corrective Reading* program (A, B1). The students in the intervention group received 60-65 lessons (in groups of five to ten students) from teachers at their schools, or, for some students, at a resource centre for surrounding schools. An additional study, with younger (mean age 8.8 years) less advanced readers involved a similar design and teaching approach. The program, *Teach Your Child to Read in 100 Easy Lessons*, was presented to thirteen students in two settings. When compared with a similar cohort of wait-list students, the students in each program made statistically significant and educationally important gains in such phonologically-based processes as word attack, phonemic awareness, and spelling; and, statistically significant gains, of at least moderate effect size, in phonological recoding in lexical access and phonological recoding in working memory.

Hicks, S. C. (2011). Effects of direct instruction on the use of and response to prepositions by students with an intellectual disability. Unpublished doctoral dissertation, University of North Carolina at Charlotte.

Hicks, S. C., Stevenson, K. M., Wood, C. L., Cooke, N. L., & Mims, P. (2011). Effects of direct instruction on the acquisition of prepositions by students with intellectual disabilities. Journal of Applied Behavior Analysis, 44, 675-679

Students with an intellectual disability often struggle with significant language delays or impairments. Although this population of students can acquire language skills, they often require methods of explicit instruction of language skills to do so. Direct Instruction (DI), a system of explicit and systematic instruction, could be one of these methods. The purpose of this study was to investigate the effects of DI on the use of and response to prepositions by three elementary school students with an intellectual disability. A multiple baseline design across prepositions was used in this study with replication across students. The researcher used DI to model examples and nonexamples (i.e., "This is above," "This is not above.") of three prepositions (e.g., above, behind, beneath) to the students. In addition to the instructional sessions, students participated in three generalization activities. Results of this study showed a functional relationship between Direct Instruction and students' use of and response to prepositions. Students demonstrated the ability to use and respond to prepositions consistently after receiving DI on each of the three target prepositions. Furthermore, all three students maintained the skill up to 56 days from instruction on each of the prepositions. These findings are important to this population of students because of the need for explicit and systematic instruction of language skills; it has been demonstrated that DI is an effective instructional tool in teaching these skills in an efficient and effective way.

Horner, R. H., & Albin, R. W. (1988). Research on general-case procedures for learners with severe disabilities. *Education and Treatment of Children, 11,* 175-188

This paper presents an overview of the Direct Instruction principles and guidelines that define general-case programing and then reviews recent research findings related to general-case instruction with learners with severe disabilities.

Infantino, J. & Hempenstall, K. (2006). Effects of a decoding program on a child with Autism Spectrum Disorder. *Australasian Journal of Special Education, 30*(2), 126-144.

This case study examined the effects of a parent-presented Direct Instruction decoding program on the reading and language skills of a child with high functioning Autism Spectrum

Disorder. Following the 23 hour intervention, reading comprehension, listening comprehension and fluency skills improved to grade level, whilst statistically significant improvements were also noted in receptive language skills. There were no significant changes in phonological and decoding skills for which various possibilities are explored, including the student's prior double deficit in phonological skills. The findings are consistent with research suggesting that increased intervention intensity, along with greater emphasis on phonological skills may be necessary to advance the decoding skills of children displaying a double deficit. The reading and language outcomes were sufficiently promising to warrant further studies employing methodologically sound group designs with this population.

Kelly, B., Carnine, D., Gersten, R., & Grossen, B. (1986). Effectiveness of videodisc instruction in teaching fractions to learning-disabled and remedial high school students. *Journal of Special Education Technology, 8,* 5-9.

This study compared the effectiveness of an interactive videodisc curriculum incorporating principles of instructional design with a traditional basal curriculum to teach basic fraction skills to 28 high school students, including 17 mildly handicapped students. The videodisc curriculum's superiority at posttest and at two-week maintenance were attributed to better instructional design.

Kelly, B., Gersten, R., & Carnine, D. (1990). Student error patterns as a function of curriculum design: Teaching fractions to remedial high school students and high school students with learning disabilities. *Journal of Learning Disabilities, 23,* 23-29.

This study compared the relative effectiveness of a Direct Instruction curriculum, which incorporated three empirically derived principles of curriculum design, with a basal approach in teaching basic fractions concepts to students with learning disabilities and other low performing students in high school remedial math classes. Results indicated that, although both programs were reasonably successful in teaching the material, the Direct Instruction was significantly more effective. Mean scores on a curriculum-referenced test were 96.5% for that group and 82.3% for the basal group. Secondary analyses of item clusters revealed that areas of weakness in the performance of the basal group could be directly linked to hypothesized flaws in its curriculum design.

Kitz, W. & Thorpe, H. (1995). A comparison of the effectiveness of videodisc and traditional algebra instruction for college-age students with learning disabilities. *Remedial & Special Education, 16(5),* 295-307.

This study compared the effectiveness of two algebra programs, the Direct Instruction-based Mastering Equations, Roots, and Exponents delivered via videodisc and a traditional text-based program. 26 college-age students with learning disabilities were randomly assigned to receive instruction with either the videodisc or traditional program. All participants were administered assessments from both programs. Results indicate that students who were in the videodisc group significantly outperformed students receiving traditional instruction not only on the Mastering Equations, Roots, and Exponents post-test, but also on the Project Success Algebra Placement Test (tied to the text-based program), despite the control group's higher pre-test scores on this test and daily practice with the material assessed. Additionally, students in the videodisc group received significantly higher grades in their first algebra class in the semester following the completion of instruction.

Kroesbergen, E. & Van Luit, J. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24*(2), 97-114.

This article presents the results of a meta-analysis of 58 studies of mathematics interventions for elementary students with special needs. Interventions in three different domains were selected: preparatory mathematics, basic skills, and problem-solving strategies. The majority of the included studies described interventions in the domain of basic skills. In general, these interventions were also the most effective. Furthermore, a few specific characteristics were found to influence the outcomes of the studies. In addition to the duration of the intervention, the particular method of intervention proved important: Direct instruction and self-instruction were found to be more effective than mediated instruction. Interventions involving the use of computer-assisted instruction and peer tutoring showed smaller effects than interventions not including these supports.

Leiss, R. H., & Proger, B. B. (1974). Language training for trainable mentally retarded: Annual project report – second year – ESEA Title III. ERIC 097 789, 83p.

During the 1973-74 school year, 230 trainable mentally retarded (TMR) children (ages 7 to 14 years) were exposed to one of two language training conditions: Distar or Peabody. A population of 116 students who continued from the first year of the project and 114 new entries were assigned in as random a fashion as possible to either Distar or Peabody. Subjects were divided into low IQ (21-43) and high IQ (44-53) groups. Sex was built into the design, as was pretest-posttest and new entries versus continuees. Thus, a five-factor, 2 x 2 x 2 x 2 x 2 repeated-measures design was subjected to analysis of variance for each of three basic criteria: Peabody Picture Vocabulary Test, Illinois Test of Psycholinguistic Abilities, and Mecham Verbal Language Development Scale. Seven children were selected randomly from each of the 16 between-factor cells to yield a total of 112 children. Longitudinal analyses were also conducted on just the continuing students with pre- and posttest data from the

three basic measures from both years of the project to yield a treatments-by-IQ-by-Sex-by-Measures (2 x 2 x 2 x 4) design. While no significant differences emerged for the high-IQ children, the low-IQ children were aided more by Distar than by Peabody. In the 5-way designs, gain in the total sample was not marked. However, when one considers only the continuing students, significant gains in language functioning did occur.

Leiss, R. H., & Proger, B. B. (1977). Language training for trainable mentally retarded children: ITPA, Peabody, and DISTAR techniques. ERIC 140 527, 39p.

To determine the most effective language training activities for trainable mentally retarded (TMR) children, the variables of degree of previous language training, IQ, Peabody language treatment program versus Distar language treatment program, pretest versus posttest, and sex were examined with 122 TMR subjects (7 to 14 years old). Results of the Peabody Picture Vocabulary Test, the Illinois Test of Psycholinguistic Abilities, and the Mechann Verbal Language Development Scale indicated that subjects were significantly better aided by the Distar program than by the Peabody program, and that only the children previously exposed to language training showed any significant growth in language. A second study examined in depth the nature of previous language experience on Distar performance.

Maggs, A., & Morath, P. (1976). Effects of direct verbal instruction on intellectual development of institutionalized moderately retarded children: A 2-year study. *The Journal of Special Education, 10(4),* 357-364.

This study compared the effects of two programs on the language gains of students with moderate to severe mental retardation across two years of implementation: DISTAR Language and the Peabody Language Kit. Results indicate that students receiving instruction with DISTAR Language achieved significantly greater gains on all assessment measures than the control group. Additionally, DISTAR students progressed at approximately the same rate as typically developing children, gaining 22.5 mental age months across 24 months, whereas control students only gained 7.5 mental age months in the same period of time.

Maggs, A., & Morath, P. The effects of direct verbal instruction on intellectual development of institutionalized moderately retarded children: A two-year study. Journal of Special Education. 1976, 10, 357-364

This study examined the intellectual development of 28 moderately retarded children (IQ's approximately 30-55), using Distar Language 1 Kit, together with positive reinforcement, modeling, and shaping techniques, in a two year intervention project. Results revealed

significant differences in favor of the experimental treatment conditions on all dependent variables (i.e., the Basic Concept Inventory, Reynell Verbal Comprehension, Stanford-Binet (L-M) Intelligence, Piaget's Class Inclusion, Piaget's Seriation, and Bruner's Matrix) related to the subjects' performance. Subjects seemed to be able to maintain approximately a "normal" rate of intellectual development over the 2-yr period.

Marston, D., Deno, S. L., Kim, D., Diment, K., & Rogers, D. (1995). Comparison of reading intervention approaches for students with mild disabilities. *Exceptional Children, 62(1),* 20-37.

This study examined effects of six different reading instructional strategies for students with mild disabilities: peer tutoring, reciprocal teaching, computer-aided instruction, effective teaching principles, Direct Instruction using Corrective Reading, and direct instruction principles applied to a basal reading series. Additionally, a comparison group was included. Student achievement and instructional ecology were assessed with curriculum-based measures and two observation systems. Results indicate that, after 10 weeks of implementation, students receiving computer-aided instruction and instruction that applied direct instruction principles to a basal reading series made significantly greater gains in reading achievement than the comparison group.

Owens, S. H., Fredrick, L. D., & Shippen, M. E. (2004). Training a paraprofessional to implement Spelling Mastery and examining its effectiveness for students with learning disabilities. *Journal of Direct Instruction*, *4(2),* 153-172.

This study examined: 1) the effectiveness and efficiency of training a paraprofessional to implement Spelling Mastery, 2) the effectiveness of Spelling Mastery with students with learning disabilities, and 3) the maintenance and generalizability of spelling skills. Results indicate that students were able to learn, maintain, and generalize spelling skills following instruction with Spelling Mastery; additionally, paraprofessionals were able to effectively implement Spelling Mastery.

Polloway, E. A., Epstein, M. H., Polloway, C. H., Patton, J. R., & Ball, D. W. (1986). Corrective Reading program: An analysis of effectiveness with learning disabled and mentally retarded students. *Remedial and Special Education 7(4),* 41-47.

The Corrective Reading Program (CRP) was used with a group of learning disabled (LD) and educable mentally retarded (EMR) adolescents with data collected on the achievement of these students in the domains of reading recognition and comprehension. When compared to reading progress made in prior years, both groups showed significantly greater

improvement. LD students experienced larger achievement gains than EMR students in both recognition and comprehension with differences in the former domain being statistically significant.

Proger, B. B & Leiss, R. H. (1976). Language training for TMR children: Third-year results and comparison with first two years: The Peabody, DISTAR, and ITPA programs. Norristown, PA: Montgomery County Intermediate Unit 23 Special Education Center.

Effects of the different components of the Distar language program were tested on 48 trainable mentally retarded (TMR) children. Three designs were used comparing IQ, sex, continuation status, selected standardized test results, and replicates. Results indicated that the degree of previous formal language training (continuation status) did not have any effect on current functioning. The Distar program produced steady, acceptable progress of the TMR children throughout the school year. In addition, the Distar program proved more effective when compared with the Peabody Language Development program received by 40 TMR children.

Scruggs, T. E., Mastropieri, M., & Levin, J. R. (1985). Vocabulary acquisition by mentally retarded students under direct and mnemonic instruction. *American Journal of Mental Deficiency, 89,* 546-551.

Twenty EMR junior-high-school students learned the definitions of unfamiliar English vocabulary words under either direct instruction or mnemonic instruction in a crossover design. When in the mnemonic-instruction condition, students remembered 50% more vocabulary definitions than when they were in the direct-instruction condition. Only three of the 20 students failed to exhibit superior performance when in the mnemonic instruction condition. Supplementary analysis of the response data revealed distinctly different error patterns in the two instructional conditions.

Shippen, M. E., Houchins, D. E., Calhoon, M. B., Furlow, C. F., & Sartor, D. L. (2006). The effects of comprehensive school reform models in reading for urban middle school students with disabilities. *Remedial and Special Education, 27(6),* 322-328.

The authors compared the effects of two CSR models ("Success for All" and "Direct Instruction") in reading for urban middle school students with disabilities who were performing 2 or more years below grade level in reading. The results indicated that students with disabilities showed little or no reading skill gain from either comprehensive school reform model and remained markedly behind.

Somerville, D. E., & Leach, D. J. (1988). Direct or indirect instruction?: An evaluation of three types of intervention programme for assisting students with specific reading disabilities. *Educational Research, 30(1),* 46-53.

Three intervention programs (psychomotor, self-esteem enhancement, and Direct Instruction) were conducted with 40 children who had reading difficulties. The Direct Instruction program resulted in the greatest gains. Post-intervention questionnaires completed by subjects, parents, and teachers indicated that perceived success differed significantly from measured success.

Stephens, M. A. (1993). Developing and implementing a curriculum and instructional program to improve reading achievement of middle-grade students with learning disabilities in a rural school district. ERIC 359 492.

A program was developed and implemented to improve the reading achievement of 56 learning-disabled (LD) students in grades five through eight in a rural school district in South Carolina. Only 12% of the students were meeting the state standard in reading. Baseline test data indicated that reading achievement was from two to six grade levels below grade placement. Problems identified were absence of written basic skills curriculum aligned with the tests used to measure achievement; lack of consistent, effective instruction in the basic skills areas; and the need for more parental support and involvement in the education of students with special needs. Solution strategies for improving reading achievement of LD students included the development of functional basic skill curriculum guides, training for teachers in the use of direct instruction techniques, field testing the curriculum guide, inclusion of the basic skills objectives in the individualized education programs of the students, implementation of a direct-instruction model in the teaching of reading, use of the Corrective Reading Program, and involvement of parents. As a result of intervention strategies, 25% of the targeted students met the state standard on basic skill tests. Over 55% of the targeted students gained two or more grade levels in reading. The discrepancy between the students' estimated ability and reading achievement was reduced by 41%. Parent involvement increased from 10% to 75%.

Viel, K.A. (2008). *The Effects of Direct Instruction in Writing on English Speakers and English Language Learners with Disabilities*, Unpublished PhD Dissertation, Georgia State University.

Viel-Ruma, K., Houchins, D., Jolivette, K., Fredrick, L., & Gama, R. (2010). Direct Instruction in written expression: The effects on English speakers and English Language Learners with disabilities. *Learning Disabilities Research & Practice, 25*(2), 97-108.

Students with disabilities often struggle with writing tasks. In order to improve the written expression performance of high school students with deficits in written expression, a Direct Instruction writing program was implemented. The participants were six high school students in programs for individuals with learning disabilities. Three of the six students were served in programs for students who are English Language Learners. Using a multiple-probe across-participants design, the effect of the writing program was examined. The intervention was implemented over a 5-week period with maintenance checks conducted 2 and 4 weeks after the termination of instruction. Results were variable, but there appeared to be a positive trend in student writing performance as measured by correct word sequence, length of text, and the TOWL-3. Implications for practice and future directions are also provided.

Walker, B., Shippen, M. E., Alberto, P., Houchins, D. E., & Cihak, D. F. (2006). Using the Expressive Writing program to improve the writing skills of high school students with learning disabilities. *Journal of Direct Instruction, 6(1),* 35-47.

The complex nature of written expression presents difficulty for many students, particularly those with learning disabilities (LD). The purpose of this study was to investigate the effects of the DI writing program, *Expressive Writing*, for high school students with learning disabilities using a single-subject design methodology. Results indicated that the *Expressive Writing* program improved the writing skills of the students in this study. Students also were able to generalize and maintain the writing skills learned during intervention.

Young, M., Baker, J., & Martin, M. (1990). Teaching basic number skills to students with a moderate intellectual disability. *Education and Training in Mental Retardation, 25,* 83-93.

Performance of 5 students (ages 8 and 10) with moderate intellectual disability was compared as they received the DISTAR Arithmetic instructional intervention and a Discrimination Learning Theory (DLT) intervention not requiring a fast-paced verbal student response. Academic engagement and skill mastery were greater during the DLT intervention.

Analyzing Case-Study Data from Schools

SRA, the publisher of Reading Mastery and other Direct Instruction programs, routinely posts information from schools regarding changes in standardized test scores after they began to use *Reading Mastery* and other Direct Instruction programs. This section describes results given by SRA from four different districts regarding changes in reading test scores of special education students after using *Reading Mastery*, sometimes in conjunction with *Corrective Reading* for students in older grades. All reports are available from the SRA research report website (http://mheresearch.com/product_info.php?segID=4&subID=21&proID=38&catID=7).

In a recent article, Stockard (2011) explained how the comparisons in these reports conform to the tenets of a cohort control group design as described by the classical literature on research design (Campbell and Stanley,1963; Cook & Campbell, 1979; Shadish, et al, 2002). The cohort control group design is an internally valid design especially appropriate for settings such as schools and provides good external validity. The article also showed how the data presented in the SRA reports could be used to calculate effect sizes, similar to those used in the meta-analyses described in an earlier section. Recall that Tallmadge (1977) suggested that effect sizes of .25 or larger be considered educationally important. All of the effect sizes calculated from the data in these reports and reported below far surpassed this level. (See Stockard (2011) for similar analyses for a large number of other districts with data regarding general education and total student populations.)

SRA/McGraw-Hill. (2006). *Exceptional education and regular education students excel with Direct Instruction*.

The Iredell Statesville School District in North Carolina has approximately 21,000 students in grades pre-K – 12. Almost three-fourths are Caucasian, 18% African American. Thirty-five percent qualify for free or reduced lunch. In fall 2003, teachers in the special education department adopted SRA/McGraw-Hill's Direct Instruction programs *Reading Mastery* and *Corrective Reading* for intervention in Grades K–12 because they felt students with disabilities were falling further and further behind their regular education peers. The students responded so well that the programs were implemented with at-risk students in regular education the following school year (2004–2005). They too responded well, and reading scores increased for both groups. The percentage of special education students attaining adequate yearly progress changed from 43 percent in 2002-03, before implementation of RM and CR, to 66% in the 2005-06 academic year. This represents an effect size of .46.

SRA/McGraw-Hill. (2006). *Reading proficiency more than doubles among Putnam County special education students*.

The Putnam County School District in Cookeville, Tennessee serves about 10,000 students. The student population is 91% Caucasian, and 47% qualify for free or reduced lunch. Reading proficiency among special education students in Grades 3-8 jumped from 37% in 2003, before implementing R*eading Mastery* and *Corrective Reading* to 85% in the 2005-2006 school year, corresponding to an effect size of .98. School officials reported that socioeconomic status had no influence on reading proficiency, reporting that 100% of special education students in one of the poorest schools in the district read proficiently in 2005. A school district researcher found statistical significance in spelling and

comprehension abilities among students exposed to *Corrective Reading*. In addition, greater mean gains were achieved by those same students in all decoding, spelling, and comprehension tests.

SRA/McGraw-Hill. (2008). *Direct Instruction reduces special education referrals in Louisiana school district by half*.

The Rapides Parish school district in Louisiana serves approximately 24,000 students, about half of whom are Caucausian and half are African American. Sixty percent classify for free or reduced-price lunch. At the start of the 2006-07 the district introduced *Reading Mastery* and *Corrective Reading* in three targeted schools due to concerns with high numbers of students being referred for special education. Once this early intervention began the number of referrals for special education evaluations dropped by 50% and the number of students who qualified for special education after referral dropped from 100% to 42%, an effect size of 1.28.

SRA/McGraw-Hill. (2008). *Special education students at California elementary school achieve AYP with Direct Instruction.*

Virginia Primrose School in Fontana, California, serves slightly more than 800 students in grades Pre-K to 5. The school's population is 80% Hispanic, all students qualify for free or reduced-price lunch, 87% are transient, and 55% are English learners. The school's principal introduced *Reading Mastery* in grades K-3 and *Corrective Reading* in grades 4-5 after finding that the majority of fifth graders read at only the second grade level. After introducing these curricula scores on state assessments rapidly improved. The percentage of special education students achieving AYP moved from 21.2% before implementation to 37.2% by the end of the second year of implementation, an effect size of .35.

References

Branwhite, A. B. (1983). Boosting reading skills by Direct Instruction. *British Journal of Educational Psychology, 53*, 291-298.

Campbell, D.T. & Stanley, J.C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Cook, T.D. & Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

Cooke, N. L., Gibbs, S. L., Campbell, M. L., & Shalvis, S. L. (2004). A comparison of Reading Mastery Fast Cycle and Horizons Fast Track A-B on the reading achievement of students with mild disabilities. *Journal of Direct Instruction, 4*(2), 139-151.

Engelmann, S. (2000). About reading – A comparison between Reading Mastery and Horizons. *Effective School Practices, 18* (3), 15-26.

Glang, A., Singer, G., Cooley, E., & Tish, N. (1991). Direct Instruction: Applications with students with brain injury. *ADI News, 11*(1), 23-28.

Glang, A., Singer, G., Cooley, E., & Tish, N. (1992). Tailoring Direct Instruction techniques for use with elementary students with brain injury. *Journal of Head Trauma Rehabilitation, 7*(4), 93-108.

Herrera, J.A., Logan, C.H., Cooker, P.G., Morris, D.P., & Lyman, D.E. (1997). Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children. Learning disability or learning difference? *Reading Improvement, 34*(2), 71-89.

Kuder, S. J. (1990). Effectiveness of the DISTAR reading program for children with learning disabilities. *Journal of Learning Disabilities, 23*(1), 69-71.

Marston, D., Deno, S. L., Kim, D., Diment, K., & Rogers, D. (1995). Comparison of reading intervention approaches for students with mild disabilities. *Exceptional Children, 62*(1), 20-37.

McKenzie, M. A., Marchand-Martella, N. E., Martella, R. C., & Moore, M.E. (2004). Teaching basic math skills to preschoolers using Connecting Math Concepts Level K. *Journal of Direct Instruction, 4*(1), 85-94.

O'Connor, R. E., Jenkins, J. R., Cole, K. N., & Mills, P. (1993). Two approaches to reading instruction with children with disabilities: Does program design make a difference? *Exceptional Children, 59* (4), 312-323.

O'Connor, R. E., & Jenkins, J. R. (1995). Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. *The Journal of Special Education, 29*(3), 255-275.

Richardson, E., DiBenedetto, B., Christ, A., Press, M., & Winsberg, B. G. (1978). An assessment of two methods for remediating reading deficiencies. *Reading Improvement, 15*(2), 82-95.

Shadish, W.R.. Cook, T.D. & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Sommers, J. (1991). Direct Instruction programs produce significant gains with at-risk middle school students. *Direct Instruction News, 11*(1), 7-14.

Stein, C., & Goldman, J. (1980). Beginning reading instruction for children with minimal brain dysfunction. *Journal of Learning Disabilities, 13*(4), 52-55.

Stockard, J. (2011) "Merging the Accountability and Scientific Research Requirements of the No Child Left Behind Act: Using Cohort Control Groups" Jean Stockard, *Quality and Quantity: International Journal of Methodology*, electronic version available on-line, December, 2011.

Tallmadge, G. K. (1977). *The Joint Dissemination Review Panel Ideabook.* Washington DC: National Institute of Education and U.S. Office of Education.

What Works Clearinghouse (2012). *WWC Intervention Report, Reading Mastery and Students with Learning Disabilities*. Washington, D.C.: Institute of Education Sciences

What Works Clearinghouse (2011). *Procedures and Standards Handbook, Version 2.1.* Washington, D.C.: Institute of Education Sciences.

Young, M., Baker, J., & Martin, M. (1990). Teaching basic number skills to students with a moderate intellectual disability. *Education and Training in Mental Retardation, 25* (1), 83-93.

From: jstockard@nifdi.org
Sent: Tuesday, August 21, 2012 8:54 PM
To: info@whatworks.ed.gov
Subject: IES WWC Website: Contact Us: Publications, Intervention Reports,
Reference ID Number: 1757800276

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: I am writing to request a Quality Review of the recent report on Reading Mastery for students with learning disabilities. I believe that the
report misinterpreted both of the studies that were reviewed and should be
removed from the website immediately.

The WWC analysis was based on two articles. One (Cooke, Gibbs, Campbell, &
Schalvis, 2004) compared two very similar Direct Instruction programs, Reading
Mastery and Horizons, and found that students in both programs made gains over
the academic year that were significantly greater than those made by students
in national and state level populations. Gains in both programs were similar,
leading the reviewers to conclude that Reading Mastery was no better than its
comparison program. They ignored the fact that students in Reading Mastery
performed significantly better than the national or state norms and that the
comparison program was similar on all but a very few characteristics, essentially a modified version of Reading Mastery. To use a medical analogy,
the WWC decision would be equivalent to a researcher claiming that aspirin was
ineffective because Tylenol performed equally well.

The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman, 1997)
involved two groups of students, both of whom appear to have received Reading
Mastery as part of the schools' "usual and customary school day curriculum."
One group of students also received 45 minutes of supplemental phonemic related instruction, from their regular classroom teachers, that involved motor activities to accompany practice of skills. Not surprisingly, the group

receiving the additional instruction had significantly larger gains than those who did not have additional learning time. Despite these differences in exposure and the fact that both groups appear to have had Reading Mastery as their usual reading curriculum, the WWC used these results to suggest that Reading Mastery could have potentially negative effects. A much more accurate conclusion would be that students who receive additional phonics related instruction have greater skill growth.

I have prepared a more extensive analysis of these concerns, which I would be happy to send you. Because the problems with the analysis are so serious I believe that the report should be immediately taken off your website. In addition, given the very serious misinterpretations of the literature described above I will continue to examine WWC reports and procedures.

Please acknowledge receipt of my concerns and let me know the time line for your review.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction

From: Jean Stockard <jstockard@nifdi.org>
Sent: Thursday, August 23, 2012 3:41 PM
To: WhatWorks
Cc: jstockard@nifdi.org
Subject: Re: What Works Clearinghouse (WWC 3801)
Attachments: RM and LD Response 7.25.12.pdf

Thank you for acknowledging receipt of my e-mail. I have attached a more lengthy analysis of both the review and related literature.
Please ask if more information would help in your review.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> Thank you for contacting the What Works Clearinghouse (WWC).
>
> We have received your email about the Reading Mastery Intervention Report.
> The WWC Quality Review Team is reviewing your email and will prepare a
> written response. The WWC Quality Review Team responds to concerns raised
> about WWC reviews published on our website. These quality reviews are
> undertaken when concerned parties present evidence that a WWC review may
> be inaccurate. When a quality review is conducted, a researcher who was
> not involved in the initial review undertakes an independent assessment of
> the study in question.
>
> If a quality review concludes that the original review contained errors, a
> revised version of the review will be published on the web site. However,
> it is WWC policy that the existing report will remain on the web site, and
> be revised only if the WWC team determines it does contain errors. These
> quality reviews are one of tools used to ensure that the standards
> established by the Institute of Educational Sciences (IES) are upheld on
> every review conducted by the What Works Clearinghouse.
>
> At this time, we are unable to provide a timeline for a QRT response.
> Please submit further comments that you would like given to the QRT. You
> may send them in response to this email.
>
> Thank you,
>

> What Works Clearinghouse
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more
> information, please visit http://ies.ed.gov/ncee/wwc/.
>
>
>
>
>
> -----Original Message-----
>
> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>
> Sent: Tuesday, August 21, 2012 8:54 PM
>
> To: info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>
>
> Subject: IES WWC Website: Contact Us: Publications, Intervention Reports,
> Reference ID Number: 1757800276
>
>
>
> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>, this email was
> automatically sent through the Contact link on the WWC website.
>
>
>
> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>
>
>
> Message: I am writing to request a Quality Review of the recent report on
> Reading Mastery for students with learning disabilities. I believe that
> the report misinterpreted both of the studies that were reviewed and
> should be removed from the website immediately.
>
>
>
> The WWC analysis was based on two articles. One (Cooke, Gibbs, Campbell, &
> Schalvis, 2004) compared two very similar Direct Instruction programs,
> Reading Mastery and Horizons, and found that students in both programs
> made gains over the academic year that were significantly greater than
> those made by students in national and state level populations. Gains in
> both programs were similar, leading the reviewers to conclude that Reading

> Mastery was no better than its comparison program. They ignored the fact
> that students in Reading Mastery performed significantly better than the
> national or state norms and that the comparison program was similar on all
> but a very few characteristics, essentially a modified version of Reading
> Mastery. To use a medical analogy, the WWC decision would be equivalent to
> a researcher claiming that aspirin was ineffective because Tylenol
> performed equally well.
>
>
>
> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman, 1997)
> involved two groups of students, both of whom appear to have received
> Reading Mastery as part of the schools' "usual and customary school day
> curriculum." One group of students also received 45 minutes of
> supplemental phonemic related instruction, from their regular classroom
> teachers, that involved motor activities to accompany practice of skills.
> Not surprisingly, the group receiving the additional instruction had
> significantly larger gains than those who did not have additional learning
> time. Despite these differences in exposure and the fact that both groups
> appear to have had Reading Mastery as their usual reading curriculum, the
> WWC used these results to suggest that Reading Mastery could have
> potentially negative effects. A much more accurate conclusion would be
> that students who receive additional phonics related instruction have
> greater skill growth.
>
>
>
> I have prepared a more extensive analysis of these concerns, which I would
> be happy to send you. Because the problems with the analysis are so
> serious I believe that the report should be immediately taken off your
> website. In addition, given the very serious misinterpretations of the
> literature described above I will continue to examine WWC reports and
> procedures.
>
>
>
> Please acknowledge receipt of my concerns and let me know the time line
> for your review.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research

>
> National Institute for Direct Instruction
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

From: Jean Stockard <jstockard@nifdi.org>
Sent: Tuesday, September 04, 2012 8:12 PM
To: WhatWorks
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Follow Up Flag: Follow up
Flag Status: Flagged

Dear WWC Staff:
I am writing with regards to the article by Herrera and associates (1997) that was included in your recent review of Reading Mastery for Learning Disabled students. I wondered if you had any additional information on the Herrera et al. study that was obtained from the authors in the course of developing your report. If so, could I possibly have access to that information? I have been trying to contact them, but have not been successful.

Thank you very much for your quick responses to queries.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> We received your email and attachment. We provided it to the QRT.
>
> Thank you,
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Thursday, August 23, 2012 2:41 PM
> To: WhatWorks
> Cc: jstockard@nifdi.org
> Subject: Re: What Works Clearinghouse (WWC 3801)
>
> Thank you for acknowledging receipt of my e-mail. I have attached a more
> lengthy analysis of both the review and related literature.
> Please ask if more information would help in your review.
>
> Sincerely,

>
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>> Dear Dr. Stockard,
>>
>> Thank you for contacting the What Works Clearinghouse (WWC).
>>
>> We have received your email about the Reading Mastery Intervention
>> Report.
>> The WWC Quality Review Team is reviewing your email and will prepare
>> a written response. The WWC Quality Review Team responds to concerns
>> raised about WWC reviews published on our website. These quality
>> reviews are undertaken when concerned parties present evidence that a
>> WWC review may be inaccurate. When a quality review is conducted, a
>> researcher who was not involved in the initial review undertakes an
>> independent assessment of the study in question.
>>
>> If a quality review concludes that the original review contained
>> errors, a revised version of the review will be published on the web
>> site. However, it is WWC policy that the existing report will remain
>> on the web site, and be revised only if the WWC team determines it
>> does contain errors. These quality reviews are one of tools used to
>> ensure that the standards established by the Institute of Educational
>> Sciences (IES) are upheld on every review conducted by the What Works
>> Clearinghouse.
>>
>> At this time, we are unable to provide a timeline for a QRT response.
>> Please submit further comments that you would like given to the QRT.
>> You may send them in response to this email.
>>
>> Thank you,
>>
>> What Works Clearinghouse
>>
>> The What Works Clearinghouse was established by the U.S. Department of
>> Education's Institute of Education Sciences to provide educators,
>> policymakers, researchers, and the public with a central and trusted
>> source of scientific evidence of what works in education. For more
>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>
>>
>>
>>

>>
>> -----Original Message-----
>>
>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>> Sent: Tuesday, August 21, 2012 8:54 PM
>>
>> To: info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>
>>
>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>> Reports, Reference ID Number: 1757800276
>>
>>
>>
>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>, this email was
>> automatically sent through the Contact link on the WWC website.
>>
>>
>>
>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>
>>
>>
>> Message: I am writing to request a Quality Review of the recent report
>> on Reading Mastery for students with learning disabilities. I believe
>> that the report misinterpreted both of the studies that were reviewed
>> and should be removed from the website immediately.
>>
>>
>>
>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>> Campbell, & Schalvis, 2004) compared two very similar Direct
>> Instruction programs, Reading Mastery and Horizons, and found that
>> students in both programs made gains over the academic year that were
>> significantly greater than those made by students in national and
>> state level populations. Gains in both programs were similar, leading
>> the reviewers to conclude that Reading Mastery was no better than its
>> comparison program. They ignored the fact that students in Reading
>> Mastery performed significantly better than the national or state
>> norms and that the comparison program was similar on all but a very
>> few characteristics, essentially a modified version of Reading
>> Mastery. To use a medical analogy, the WWC decision would be
>> equivalent to a researcher claiming that aspirin was ineffective because
>> Tylenol performed equally well.
>>
>>
>>

>> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman,
>> 1997) involved two groups of students, both of whom appear to have
>> received Reading Mastery as part of the schools' "usual and customary
>> school day curriculum." One group of students also received 45 minutes
>> of supplemental phonemic related instruction, from their regular
>> classroom teachers, that involved motor activities to accompany practice
>> of skills.
>> Not surprisingly, the group receiving the additional instruction had
>> significantly larger gains than those who did not have additional
>> learning time. Despite these differences in exposure and the fact that
>> both groups appear to have had Reading Mastery as their usual reading
>> curriculum, the WWC used these results to suggest that Reading Mastery
>> could have potentially negative effects. A much more accurate
>> conclusion would be that students who receive additional phonics
>> related instruction have greater skill growth.
>>
>>
>>
>> I have prepared a more extensive analysis of these concerns, which I
>> would be happy to send you. Because the problems with the analysis are
>> so serious I believe that the report should be immediately taken off
>> your website. In addition, given the very serious misinterpretations
>> of the literature described above I will continue to examine WWC
>> reports and procedures.
>>
>>
>>
>> Please acknowledge receipt of my concerns and let me know the time
>> line for your review.
>>
>>
>>
>> Sincerely,
>>
>>
>>
>> Jean Stockard, Ph.D.
>>
>> Director of Research
>>
>> National Institute for Direct Instruction
>>
>>
>>
>
>
> Jean Stockard, Ph.D.
> Director of Research

> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

From: Jean Stockard <jstockard@nifdi.org>
Sent: Tuesday, October 02, 2012 10:59 AM
To: WhatWorks
Cc: jstockard@nifdi.org
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear WWC Staff,

Some recent events have prompted me to follow up on my request for a quality review of WWC Report on Reading Mastery for Learning Disabled Students (reference number is 1757800276), which was submitted on August 21, 2012.

First, we received confirmation from an author of one of the articles cited in the review (Dr. Philip Cooker, a co-author of the Herrera, et al. study), that our interpretation of the work, as explained in my earlier correspondence, was correct. A copy of this correspondence is attached below.

Second, we asked highly regarded independent researchers to examine our concerns. They have confirmed our interpretation of the problems with the WWC report and note additional problems with WWC procedures and policies.

Third, our office has been told that, because of the WWC posting, schools are beginning to abandon the use of Reading Mastery with learning disabled students. In addition to the fiscal damages that will accrue to the authors and publisher from these decisions, such actions would deprive many students of access to a curriculum that has repeatedly been shown to be highly effective.

Finally, we were requested to respond to Senator Lisa Murkowski, whose office cited the WWC report in correspondence with schools. We have sent her our explanations of the serious errors in your review.

Given the very serious ramifications of your report, I again ask that you immediately remove it from your website and that you provide a clear timeline for the completion of your review including the steps taken to date. If you are not able to do so, we will need to pursue additional avenues to address this situation. Please let me know, as soon as possible, your intentions in this matter.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

phone: 541-505-5710 or toll-free 877-485-1973

Below is the correspondence from Dr. Cooker

From: "PHILIP G COOKER" <cooker@olemiss.edu>
Subject: RE: Query on article published in 1997 -
Date: Thu, September 6, 2012 9:12 am
To: "jstockard@nifdi.org" <jstockard@nifdi.org>

Hello,

sorry for the late response. I did not conduct the research, but edited and coordinated the efforts. Of the authors, I only knew Dr. Logan. It has been quite a
while since the publication and I have little recollection of details. It sounds as
if your interpretation is correct. Since reading is not my area of specialization, I
do not have knowledge of other research involving the system that was used.

Philip G. Cooker, Ph.D.
Mississippi Board of Psychology
CE Coordinator
407 Timber Lane
Oxford, MS 38655
(662) 801-1664

From: Jean Stockard [jstockard@nifdi.org]
Sent: Tuesday, September 04, 2012 6:51 PM
To: (b)(6) PHILIP G COOKER
Subject: Query on article published in 1997 -

Dear Professors Herrera and Cooker,

This is a note that I sent a few weeks ago. I am afraid it may have gotten lost in your inbox.

I am writing for a quick clarification/check on two items in your article, "Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children," which was published in 1997 in Reading Improvement.

First, my understanding is that all students in the study had Reading Mastery as "the usual and customary prescribed curriculum for the school day" (p. 78), and that the intervention group was given the Stabilized

Learning System “as a supplement to the usual and customary school day curriculum” (p. 78). In other words, both the control group and the intervention group had Reading Mastery for their usual curriculum, and the intervention group received additional instruction via the Stabilized Learning System. Is this interpretation correct?

Second, I’ve not been able to find other writings on the Stabilized Learning System. Do you know if there are other studies of the program?

Thank you very much for your help with my queries. If I can provide any clarification, please just ask.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> Thank you for your email. As was documented in the table notes in Appendix
> C.1 of the report, we attempted to contact the author (multiple times),
> but did not receive a response.
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Tuesday, September 04, 2012 7:12 PM
> To: WhatWorks
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear WWC Staff:
> I am writing with regards to the article by Herrera and associates (1997)
> that was included in your recent review of Reading Mastery for Learning
> Disabled students. I wondered if you had any additional information on the
> Herrera et al. study that was obtained from the authors in the course of
> developing your report. If so, could I possibly have access to that
> information? I have been trying to contact them, but have not been
> successful.
>
> Thank you very much for your quick responses to queries.

>
> Sincerely,
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>> Dear Dr. Stockard,
>>
>> We received your email and attachment. We provided it to the QRT.
>>
>> Thank you,
>>
>> What Works Clearinghouse
>>
>> -----Original Message-----
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>> Sent: Thursday, August 23, 2012 2:41 PM
>> To: WhatWorks
>> Cc: jstockard@nifdi.org
>> Subject: Re: What Works Clearinghouse (WWC 3801)
>>
>> Thank you for acknowledging receipt of my e-mail. I have attached a
>> more lengthy analysis of both the review and related literature.
>> Please ask if more information would help in your review.
>>
>> Sincerely,
>>
>>
>> Jean Stockard, Ph.D.
>> Director of Research
>> National Institute for Direct Instruction Eugene, Oregon
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>> Dear Dr. Stockard,
>>>
>>> Thank you for contacting the What Works Clearinghouse (WWC).
>>>
>>> We have received your email about the Reading Mastery Intervention
>>> Report.
>>> The WWC Quality Review Team is reviewing your email and will prepare
>>> a written response. The WWC Quality Review Team responds to concerns
>>> raised about WWC reviews published on our website. These quality
>>> reviews are undertaken when concerned parties present evidence that a
>>> WWC review may be inaccurate. When a quality review is conducted, a
>>> researcher who was not involved in the initial review undertakes an

>>> independent assessment of the study in question.
>>>
>>> If a quality review concludes that the original review contained
>>> errors, a revised version of the review will be published on the web
>>> site. However, it is WWC policy that the existing report will remain
>>> on the web site, and be revised only if the WWC team determines it
>>> does contain errors. These quality reviews are one of tools used to
>>> ensure that the standards established by the Institute of Educational
>>> Sciences (IES) are upheld on every review conducted by the What Works
>>> Clearinghouse.
>>>
>>> At this time, we are unable to provide a timeline for a QRT response.
>>> Please submit further comments that you would like given to the QRT.
>>> You may send them in response to this email.
>>>
>>> Thank you,
>>>
>>> What Works Clearinghouse
>>>
>>> The What Works Clearinghouse was established by the U.S. Department
>>> of Education's Institute of Education Sciences to provide educators,
>>> policymakers, researchers, and the public with a central and trusted
>>> source of scientific evidence of what works in education. For more
>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>>
>>>
>>>
>>>
>>>
>>> -----Original Message-----
>>>
>>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>>
>>> Sent: Tuesday, August 21, 2012 8:54 PM
>>>
>>> To: info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>
>>>
>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>>> Reports, Reference ID Number: 1757800276
>>>
>>>
>>>
>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>, this email was
>>> automatically sent through the Contact link on the WWC website.
>>>
>>>
>>>

>>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>>
>>>
>>>
>>> Message: I am writing to request a Quality Review of the recent
>>> report on Reading Mastery for students with learning disabilities. I
>>> believe that the report misinterpreted both of the studies that were
>>> reviewed and should be removed from the website immediately.
>>>
>>>
>>>
>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>>> Instruction programs, Reading Mastery and Horizons, and found that
>>> students in both programs made gains over the academic year that were
>>> significantly greater than those made by students in national and
>>> state level populations. Gains in both programs were similar, leading
>>> the reviewers to conclude that Reading Mastery was no better than its
>>> comparison program. They ignored the fact that students in Reading
>>> Mastery performed significantly better than the national or state
>>> norms and that the comparison program was similar on all but a very
>>> few characteristics, essentially a modified version of Reading
>>> Mastery. To use a medical analogy, the WWC decision would be
>>> equivalent to a researcher claiming that aspirin was ineffective
>>> because Tylenol performed equally well.
>>>
>>>
>>>
>>> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman,
>>> 1997) involved two groups of students, both of whom appear to have
>>> received Reading Mastery as part of the schools' "usual and customary
>>> school day curriculum." One group of students also received 45
>>> minutes of supplemental phonemic related instruction, from their
>>> regular classroom teachers, that involved motor activities to
>>> accompany practice of skills.
>>> Not surprisingly, the group receiving the additional instruction had
>>> significantly larger gains than those who did not have additional
>>> learning time. Despite these differences in exposure and the fact
>>> that both groups appear to have had Reading Mastery as their usual
>>> reading curriculum, the WWC used these results to suggest that
>>> Reading Mastery could have potentially negative effects. A much more
>>> accurate conclusion would be that students who receive additional
>>> phonics related instruction have greater skill growth.
>>>
>>>
>>>
>>> I have prepared a more extensive analysis of these concerns, which I
>>> would be happy to send you. Because the problems with the analysis

>>> are so serious I believe that the report should be immediately taken
>>> off your website. In addition, given the very serious
>>> misinterpretations of the literature described above I will continue
>>> to examine WWC reports and procedures.
>>>
>>>
>>>
>>> Please acknowledge receipt of my concerns and let me know the time
>>> line for your review.
>>>
>>>
>>>
>>> Sincerely,
>>>
>>>
>>>
>>> Jean Stockard, Ph.D.
>>>
>>> Director of Research
>>>
>>> National Institute for Direct Instruction
>>>
>>>
>>>
>>
>>
>> Jean Stockard, Ph.D.
>> Director of Research
>> National Institute for Direct Instruction Eugene, Oregon
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>

From: Doug Carnine <dcarnine@uoregon.edu>
Sent: Wednesday, October 10, 2012 2:43 PM
To: Joy.Lesnick@ed.gov
Cc: WhatWorks; jstockard@nifdi.org
Subject: Reading Mastery report by WWC
Attachments: Dr. Stockard email to WWC. 10-2-12.doc; IES-WWC email 10-9-12.doc

The following email is also attached: IES-WWC email

10-10-12

Dear Dr. Lesnick:

It is painful and depressing to have to come out of retirement, briefly, to write you about the inaccuracies of the recent report posted on the WWC website regarding the Direct Instruction program, Reading Mastery, and its effect on learning disabled students. Dr. Jean Stockard, research director of the National Institute for Direct Instruction, submitted a request for a quality review of the Reading Mastery report on August 21. Here is a summary of the problems Dr. Stockard found with the report:

The WWC analysis was based on only two articles, a fraction of the available studies. One of the articles (Cooke, Gibbs, Campbell, & Schalvis, 2004)
compared two very similar Direct Instruction programs, Reading Mastery and Horizons that were written by the same senior author (Siegfried Engelmann). The study found that students in both programs made gains over the academic year that were significantly greater than would be expected given national norms (of the Woodcock Johnson-Revised) or the performance of other students in the state (on the state K-2 literacy test) (see page 146 of the article).
Gains in both programs were similar, leading the WWC to conclude that Reading Mastery was no better than its comparison program. They ignored the fact that students in
Reading Mastery performed significantly better than the national or state norms and that the comparison program was similar on all but a very few
characteristics, essentially a modified version of Reading Mastery. To use a medical analogy, the WWC decision would be

equivalent to a researcher who compared ibuprofen and aspirin finding similar positive effects for both and then claiming that aspirin was ineffective. (I realize this is a challenge in research with two treatments rather than a treatment and a control.)

The second article (Herrera, Logan, Cooke, Morris, & Lyman, 1997) involved two groups of students, both of whom appear to have received Reading Mastery as part of the schools' ³usual and customary school day curriculum.² The article was not clearly written. After reading it several times, Dr. Stockard interpreted the article to indicate that one group of students also received 45 minutes of supplemental phonemic related instruction, from their regular classroom teachers. Dr. Stockard's interpretation was tentatively confirmed in an e-mail sent by the second author of the study. (Correspondence of Dr. Stockard with the WWC indicated that they had unsuccessfully tried to contact only the senior author). Not surprisingly, the group receiving the additional instruction had significantly larger gains than those who did not have additional learning time. Despite these differences in exposure and the fact that both groups had Reading Mastery as their usual reading curriculum, the authors of the WWC report used these results to conclude that Reading Mastery could have potentially negative effects. The accurate conclusion would be that students who receive additional reading-related instruction had greater growth in reading related skills (as measured by a district developed curriculum based measure of vocabulary and reading, see pp. 78-79). To again use a medical analogy, this would be equivalent to comparing the impact of taking one aspirin or one aspirin and one Tylenol and, upon finding that one Tylenol and one aspirin resulted in better pain relief, concluding that one aspirin had a ³potentially negative² effect.

There have been earlier concerns about the quality of the reviews of Direct Instruction programs conducted by WWC. This quote is from a September 8, 2008 communication from Mathematica

in response to an inquiry by Dr. Stockard about the exclusion of a study on the effects of a Reading Mastery implementation sponsored by the Rodeo Institute for Teacher Effectiveness (RITE). Coleen Carlson and David Francis conducted the study. [3]A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program[2] (Dynarski, p. 5). The Carlson and Francis study involved thousands of students, hundreds of teachers, and dozens of schools and used sophisticated multivariate techniques to analyze the data. The study was excluded because the teachers received training, which might suggest that any evaluation of a literacy program in which professional development is being conducted at the school should be excluded because of the potential confound. Alternatively the study should have been reviewed in the category of school reform model, as is the case for SFA. When I contrast the excluded Carlson and Francis study with the Herrera, Logan, Cooke, Morris, & Lyman, (1997) study in which students received Reading Mastery and an additional 45 minutes of reading instruction, I can only say I am perplexed. Does not 45 minutes of extra reading instruction constitute a confound? According to the 2011 WWC guidelines this study should not have met standards for inclusion: [3]A study may fail to meet WWC standards if the measures of effect cannot be attributed solely to the intervention-the intervention was combined with another intervention[2] (pp.15).

I have attached Dr. Stockard's most recent appeal to WWC, where you can read the reasons for the request that the review be immediately removed from the WWC website until the review can be completed. I am going to suggest that Dr. Jean Stockard send you the earlier submitted appeals that have resulted in inadequate responses.

As you may know, several meta-analyses in several peer-reviewed journals have consistently found strong support for the efficacy of Direct Instruction programs, including Reading Mastery.

Could the current review be biased because of an inadequate literature search? The 2011 WWC guidelines state their intervention reports [3]summarize all studies published during a specific time period that examine the effectiveness of an intervention[2] (pp.1). However several relevant studies were excluded from the report. Even if some of the studies did not meet the WWC standards they should have been referenced in the report for not meeting standards.

If you have questions about the specifics of this letter or other documents that have been submitted, please submit them to Dr. Jean Stockard, whose contact information follows.

Sincerely,

Doug Carnine, Professor Emeritus
University of Oregon
Phone: 541 485 3781

Cc: Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
Phone: 541-505-5710 or toll-free 877-485-1973
jstockard@nifdi.org

From: Jean Stockard <jstockard@nifdi.org>
Sent: Friday, November 02, 2012 4:09 PM
To: WhatWorks
Cc: jstockard@nifdi.org
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear What Works Clearinghouse Help Desk,

I am writing in regards to my request for a quality review of the WWC Report on Reading Mastery for Learning Disabled Students (reference number is 1757800276), which was submitted on August 21, 2012. I sent additional information regarding the request a few weeks later, and Dr. Keith Smolkowski, an independent researcher, sent an extensive memo regarding the review last week.

The WWC Procedures and Standards Handbook (p. 5) notes that if a Quality Review results in a validation of the WWC's original assessment the person requesting the review is to receive a response. If the original assessment is not validated, reports are made to others within the WWC and IES. In this case, there is, however, no stated provision for notifying the person who asked for the review.

More than two months have elapsed since my original request for a Quality Review. I must assume, given the length of time since my request was submitted and the relatively simple job of reviewing the two articles involved, that my concerns with the WWC report on Reading Mastery and Learning Disabled students have been validated but that, per the stated procedures, I have not been informed of this decision.

As in my earlier correspondences, which are appended to this e-mail, I ask that you immediately remove the erroneous report from the web and clearly indicate that previous postings provided incorrect information. In addition to producing possible fiscal damages for the authors and publisher of the Reading Mastery program, continued dissemination gives very erroneous information to the public.

If I am mistaken in my assumption regarding the outcome of the review, please send me the details of your decision.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> The quality review is in process and we will expedite the inclusion of the
> information on the Herrera, et al. study that was provided yesterday.
>
>
>
> What Works Clearinghouse
>
>
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Tuesday, October 02, 2012 9:59 AM
> To: WhatWorks
> Cc: jstockard@nifdi.org
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
>
>
> Dear WWC Staff,
>
>
>
> Some recent events have prompted me to follow up on my request for a
> quality review of WWC Report on Reading Mastery for Learning Disabled
> Students (reference number is 1757800276), which was submitted on August
> 21, 2012.
>
>
>
> First, we received confirmation from an author of one of the articles
> cited in the review (Dr. Philip Cooker, a co-author of the Herrera, et
> al.
>
> study), that our interpretation of the work, as explained in my earlier
> correspondence, was correct. A copy of this correspondence is attached
> below.
>
>
>
> Second, we asked highly regarded independent researchers to examine our
> concerns. They have confirmed our interpretation of the problems with the
> WWC report and note additional problems with WWC procedures and policies.
>

>
>
> Third, our office has been told that, because of the WWC posting, schools
> are beginning to abandon the use of Reading Mastery with learning disabled
> students. In addition to the fiscal damages that will accrue to the
> authors and publisher from these decisions, such actions would deprive
> many students of access to a curriculum that has repeatedly been shown to
> be highly effective.
>
>
>
> Finally, we were requested to respond to Senator Lisa Murkowski, whose
> office cited the WWC report in correspondence with schools. We have sent
> her our explanations of the serious errors in your review.
>
>
>
> Given the very serious ramifications of your report, I again ask that you
> immediately remove it from your website and that you provide a clear
> timeline for the completion of your review including the steps taken to
> date. If you are not able to do so, we will need to pursue additional
> avenues to address this situation. Please let me know, as soon as
> possible, your intentions in this matter.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction Eugene, Oregon
>
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>
> Below is the correspondence from Dr. Cooker
>
>
>
> From: "PHILIP G COOKER"
> <cooker@olemiss.edu<mailto:cooker@olemiss.edu>>
>
> Subject: RE: Query on article published in 1997 -

>
> Date: Thu, September 6, 2012 9:12 am
>
> To: "jstockard@nifdi.org<mailto:jstockard@nifdi.org>"
> <jstockard@nifdi.org<mailto:jstockard@nifdi.org>>
>
>
>
> _____
>
>
>
> Hello,
>
>
>
> sorry for the late response. I did not conduct the research, but edited
> and coordinated the efforts. Of the authors, I only knew Dr. Logan. It has
> been quite a while since the publication and I have little recollection of
> details. It sounds as if your interpretation is correct. Since reading is
> not my area of specialization, I do not have knowledge of other research
> involving the system that was used.
>
>
>
> Philip G. Cooker, Ph.D.
>
> Mississippi Board of Psychology
>
> CE Coordinator
>
> 407 Timber Lane
>
> Oxford, MS 38655
>
> (662) 801-1664
>
>
>
> _____
>
> From: Jean Stockard [jstockard@nifdi.org]
>
> Sent: Tuesday, September 04, 2012 6:51 PM
>
> To: (b)(6) PHILIP G
> COOKER
>

> Subject: Query on article published in 1997 -
>
>
>
> Dear Professors Herrera and Cooker,
>
>
>
> This is a note that I sent a few weeks ago. I am afraid it may have gotten
> lost in your inbox.
>
>
>
> I am writing for a quick clarification/check on two items in your article,
> “Phonological awareness and phonetic-graphic conversion: A study of the
> effects of two intervention paradigms with learning disabled children,”
>
> which was published in 1997 in Reading Improvement.
>
>
>
> First, my understanding is that all students in the study had Reading
> Mastery as “the usual and customary prescribed curriculum for the school
> day” (p. 78), and that the intervention group was given the Stabilized
> Learning System “as a supplement to the usual and customary school day
> curriculum” (p. 78). In other words, both the control group and the
> intervention group had Reading Mastery for their usual curriculum, and the
> intervention group received additional instruction via the Stabilized
> Learning System. Is this interpretation correct?
>
>
>
> Second, I’ve not been able to find other writings on the Stabilized
> Learning System. Do you know if there are other studies of the program?
>
>
>
> Thank you very much for your help with my queries. If I can provide any
> clarification, please just ask.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>

> Director of Research
>
> National Institute for Direct Instruction Eugene, Oregon
>
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>
>
>
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>> Thank you for your email. As was documented in the table notes in
>
>> Appendix
>
>> C.1 of the report, we attempted to contact the author (multiple
>
>> times), but did not receive a response.
>
>>
>
>> What Works Clearinghouse
>
>>
>
>> -----Original Message-----
>
>> From: Jean Stockard
>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>
>> Sent: Tuesday, September 04, 2012 7:12 PM
>
>> To: WhatWorks
>
>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>
>> 1757800276
>
>>
>
>> Dear WWC Staff:

>
>> I am writing with regards to the article by Herrera and associates
>
>> (1997) that was included in your recent review of Reading Mastery for
>
>> Learning Disabled students. I wondered if you had any additional
>
>> information on the Herrera et al. study that was obtained from the
>
>> authors in the course of developing your report. If so, could I
>
>> possibly have access to that information? I have been trying to
>
>> contact them, but have not been successful.
>
>>
>
>> Thank you very much for your quick responses to queries.
>
>>
>
>> Sincerely,
>
>>
>
>> Jean Stockard, Ph.D.
>
>> Director of Research
>
>> National Institute for Direct Instruction Eugene, Oregon
>
>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>
>
>>> Dear Dr. Stockard,
>
>>>
>
>>> We received your email and attachment. We provided it to the QRT.
>
>>>
>
>>> Thank you,
>
>>>
>
>>> What Works Clearinghouse

>
>>>
>
>>> -----Original Message-----
>
>>> From: Jean Stockard
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>
>>> Sent: Thursday, August 23, 2012 2:41 PM
>
>>> To: WhatWorks
>
>>> Cc: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>
>>> Subject: Re: What Works Clearinghouse (WWC 3801)
>
>>>
>
>>> Thank you for acknowledging receipt of my e-mail. I have attached a
>
>>> more lengthy analysis of both the review and related literature.
>
>>> Please ask if more information would help in your review.
>
>>>
>
>>> Sincerely,
>
>>>
>
>>>
>
>>> Jean Stockard, Ph.D.
>
>>> Director of Research
>
>>> National Institute for Direct Instruction Eugene, Oregon
>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>>
>
>>>
>
>>>> Dear Dr. Stockard,
>
>>>>
>

>>>> Thank you for contacting the What Works Clearinghouse (WWC).
>
>>>>
>
>>>> We have received your email about the Reading Mastery Intervention
>
>>>> Report.
>
>>>> The WWC Quality Review Team is reviewing your email and will
>
>>>> prepare a written response. The WWC Quality Review Team responds to
>
>>>> concerns raised about WWC reviews published on our website. These
>
>>>> quality reviews are undertaken when concerned parties present
>
>>>> evidence that a WWC review may be inaccurate. When a quality review
>
>>>> is conducted, a researcher who was not involved in the initial
>
>>>> review undertakes an independent assessment of the study in question.
>
>>>>
>
>>>> If a quality review concludes that the original review contained
>
>>>> errors, a revised version of the review will be published on the web
>
>>>> site. However, it is WWC policy that the existing report will remain
>
>>>> on the web site, and be revised only if the WWC team determines it
>
>>>> does contain errors. These quality reviews are one of tools used to
>
>>>> ensure that the standards established by the Institute of
>
>>>> Educational Sciences (IES) are upheld on every review conducted by
>
>>>> the What Works Clearinghouse.
>
>>>>
>
>>>> At this time, we are unable to provide a timeline for a QRT response.
>
>>>> Please submit further comments that you would like given to the QRT.
>
>>>> You may send them in response to this email.
>

>>>>
>
>>>> Thank you,
>
>>>>
>
>>>> What Works Clearinghouse
>
>>>>
>
>>>> The What Works Clearinghouse was established by the U.S. Department
>
>>>> of Education's Institute of Education Sciences to provide educators,
>
>>>> policymakers, researchers, and the public with a central and trusted
>
>>>> source of scientific evidence of what works in education. For more
>
>>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>> -----Original Message-----
>
>>>>
>
>>>> From:
>>>>
jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@nifdi.org%3cmailto:jstockard@nifdi.org>>
>
>>>>
[mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]><mailto:[mailto:jstockard@nifdi.org]>
>
>>>>
>
>>>> Sent: Tuesday, August 21, 2012 8:54 PM
>
>>>>
>

>>>> To:
>>>>
info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@whatworks.ed.gov%3cmailto:info@whatworks.ed.gov>>
>
>>>>
>
>>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>
>>>> Reports, Reference ID Number: 1757800276
>
>>>>
>
>>>>
>
>>>>
>
>>>>
info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@whatworks.ed.gov%3cmailto:info@whatworks.ed.gov>>,
>>>> this email was
>
>>>> automatically sent through the Contact link on the WWC website.
>
>>>>
>
>>>>
>
>>>>
>
>>>> From:
>>>>
jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@nifdi.org%3cmailto:jstockard@nifdi.org>>
>
>>>>
>
>>>>
>
>>>>
>
>>>> Message: I am writing to request a Quality Review of the recent
>
>>>> report on Reading Mastery for students with learning disabilities. I
>
>>>> believe that the report misinterpreted both of the studies that were
>
>>>> reviewed and should be removed from the website immediately.

>
>>>>
>
>>>>
>
>>>>
>
>>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>
>>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>
>>>> Instruction programs, Reading Mastery and Horizons, and found that
>
>>>> students in both programs made gains over the academic year that
>
>>>> were significantly greater than those made by students in national
>
>>>> and state level populations. Gains in both programs were similar,
>
>>>> leading the reviewers to conclude that Reading Mastery was no better
>
>>>> than its comparison program. They ignored the fact that students in
>
>>>> Reading Mastery performed significantly better than the national or
>
>>>> state norms and that the comparison program was similar on all but a
>
>>>> very few characteristics, essentially a modified version of Reading
>
>>>> Mastery. To use a medical analogy, the WWC decision would be
>
>>>> equivalent to a researcher claiming that aspirin was ineffective
>
>>>> because Tylenol performed equally well.
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman,
>
>>>> 1997) involved two groups of students, both of whom appear to have
>
>>>> received Reading Mastery as part of the schools' "usual and
>
>>>> customary school day curriculum." One group of students also

>
>>>> received 45 minutes of supplemental phonemic related instruction,
>
>>>> from their regular classroom teachers, that involved motor
>
>>>> activities to accompany practice of skills.
>
>>>> Not surprisingly, the group receiving the additional instruction had
>
>>>> significantly larger gains than those who did not have additional
>
>>>> learning time. Despite these differences in exposure and the fact
>
>>>> that both groups appear to have had Reading Mastery as their usual
>
>>>> reading curriculum, the WWC used these results to suggest that
>
>>>> Reading Mastery could have potentially negative effects. A much more
>
>>>> accurate conclusion would be that students who receive additional
>
>>>> phonics related instruction have greater skill growth.
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>> I have prepared a more extensive analysis of these concerns, which I
>
>>>> would be happy to send you. Because the problems with the analysis
>
>>>> are so serious I believe that the report should be immediately taken
>
>>>> off your website. In addition, given the very serious
>
>>>> misinterpretations of the literature described above I will continue
>
>>>> to examine WWC reports and procedures.
>
>>>>
>
>>>>
>
>>>>
>
>>>> Please acknowledge receipt of my concerns and let me know the time

>
>>>> line for your review.
>
>>>>
>
>>>>
>
>>>>
>
>>>> Sincerely,
>
>>>>
>
>>>>
>
>>>>
>
>>>> Jean Stockard, Ph.D.
>
>>>>
>
>>>> Director of Research
>
>>>>
>
>>>> National Institute for Direct Instruction
>
>>>>
>
>>>>
>
>>>>
>
>>>
>
>>>
>
>>> Jean Stockard, Ph.D.
>
>>> Director of Research
>
>>> National Institute for Direct Instruction Eugene, Oregon
>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>>
>
>>

>
>>
>
>> Jean Stockard, Ph.D.
>
>> Director of Research
>
>> National Institute for Direct Instruction Eugene, Oregon
>
>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>
>
>>
>
>>
>
>
>
>
>
>
>
>
>
>
>
>

From: Jean Stockard <jstockard@nifdi.org>
Sent: Monday, December 17, 2012 5:30 PM
To: What Works
Subject: Re: What Works Clearinghouse (QR2012012)

Dear WWC Help Desk,

Thank you for sending Dr. Constantine's letter of December 11. To help further my understanding of your procedures and policies I would like to also receive more complete explanations of the reasoning behind the decisions.

I assume that your staff has prepared documents that describe the reasoning for the decisions regarding the Herrera et al and Cooke et al studies in more detail as well as documents that summarize the reasons for rejecting the additional studies for consideration and the reasons for rejecting the views expressed by Dr. Smolkowski. Please forward that material to me at your earliest convenience.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> Attached is a response to the questions you raised in your message
> regarding the What Works Clearinghouse (WWC) intervention report on
> Reading Mastery for students with learning disabilities.
>
>
>
> Thank you,
>
>
> What Works Clearinghouse
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more
> information, please visit http://ies.ed.gov/ncee/wwc/.
>

From: Jean Stockard <jstockard@nifdi.org>
Sent: Sunday, July 28, 2013 9:08 PM
To: WhatWorks
Subject: Request for Quality Review - Reading Mastery and Students with Learning Disabilities
Attachments: RM-LD and RR-BR Analysis 7.28.13.pdf

I am writing to request a Quality Review of the report on Reading Mastery for students with learning disabilities, which was posted in mid-July, 2013. The report did not include a number of studies that should have been included and seriously misinterpreted the one analysis that was reviewed. In addition, the date on the report (July 2012) is incorrect, for it was clearly not posted until July, 2013 and it is substantially different than the report that was posted in July of 2012. The date should be corrected. In addition the 2013 report should acknowledge that it replaces a report posted in 2012 that was inaccurate.

The article that was accepted for review in the 2013 report (Cooke, Gibbs, Campbell, & Schalvis, 2004) compared two very similar Direct Instruction programs, Reading Mastery and Horizons, and found that students in both programs made gains over the academic year that were significantly greater than those made by students in national and state level populations. This important positive finding is not even mentioned in the WWC report. Instead, the review notes that gains in both programs were similar and concludes that Reading Mastery was no better than its comparison program. To use a medical analogy, the WWC decision would be equivalent to a researcher claiming that aspirin was ineffective because buffered aspirin performed equally well. This conclusion is clearly illogical, and the report should be immediately withdrawn and a correction posted prominently on your website.

Our office has prepared a more extensive analysis of these concerns, including a comparison of the procedures used with the review of Reading Mastery to those used with a recent review of Reading Recovery. The comparison of the two analyses shows both the inaccuracies in the WWC's recent report on Reading Mastery as well as inconsistencies with other reviews. I have taken the liberty of attaching this report.

Please acknowledge receipt of my concerns and let me know the time line for your review and the removal of the report from the website.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

****

EXPERT TRAINING AVAILABLE IN ALL MAJOR DI PROGRAMS AND LEVELS! Join us for the 2013 Annual Training Institute presented by the National Institute for Direct Instruction in San Antonio August 5-8th, 2013. For more information, go to www.nifdi.org/training-institute.

From: What Works
Sent: Tuesday, December 11, 2012 2:23 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse (QR2012012)
Attachments: QRT 2012012.pdf

Dear Dr. Stockard,

Attached is a response to the questions you raised in your message regarding the What Works Clearinghouse (WWC) intervention report on Reading Mastery for students with learning disabilities.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: What Works <whatworks@mathematica-mpr.com>
Sent: Wednesday, February 13, 2013 11:34 AM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (QR2012012)
Attachments: QR2012012.pdf

Dear Dr. Stockard,

Attached is a response to the questions you raised in response to Dr. Constantine's letter of December 11.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: WhatWorks
Sent: Tuesday, July 16, 2013 6:15 PM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (QR2012012)

Dear Dr. Stockard,

We are emailing you to let you know that the revised report is live on our website: http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=573.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: What Works [mailto:whatworks@mathematica-mpr.com]
Sent: Wednesday, February 13, 2013 10:34 AM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (QR2012012)

Dear Dr. Stockard,

Attached is a response to the questions you raised in response to Dr. Constantine's letter of December 11.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: What Works
Sent: Friday, September 20, 2013 9:09 AM
To: 'jstockard@nifdi.org'
Cc: 'Joy.Lesnick@ed.gov'
Subject: What Works Clearinghouse (QRT 2013001 and 2012012)
Attachments: QRT2013001 Response_9.19.2013.pdf

Dear Dr. Stockard,

Attached in the document entitled *QRT 2013001 Response_9.19.2013* is a response to your inquiries from July 28, 2013 into the What Works Clearinghouse review of Reading Mastery and Reading Recovery.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: WhatWorks
Sent: Wednesday, August 22, 2012 4:07 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse (WWC 3801)

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC).

We have received your email about the *Reading Mastery* Intervention Report. The WWC Quality Review Team is reviewing your email and will prepare a written response. The WWC Quality Review Team responds to concerns raised about WWC reviews published on our website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question.

If a quality review concludes that the original review contained errors, a revised version of the review will be published on the web site. However, it is WWC policy that the existing report will remain on the web site, and be revised only if the WWC team determines it does contain errors. These quality reviews are one of tools used to ensure that the standards established by the Institute of Educational Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.

At this time, we are unable to provide a timeline for a QRT response. Please submit further comments that you would like given to the QRT. You may send them in response to this email.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Tuesday, August 21, 2012 8:54 PM
To: info@whatworks.ed.gov
Subject: IES WWC Website: Contact Us: Publications, Intervention Reports, Reference ID Number: 1757800276

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: jstockard@nifdi.org

Message: I am writing to request a Quality Review of the recent report on Reading Mastery for students with learning disabilities. I believe that the report misinterpreted both of the studies that were reviewed and should be removed from the website immediately.

The WWC analysis was based on two articles. One (Cooke, Gibbs, Campbell, & Schalvis, 2004) compared two very similar Direct Instruction programs, Reading Mastery and Horizons, and found that students in both programs made gains over the academic year that were significantly greater than those made by students in national and state level populations. Gains in both programs were similar, leading the reviewers to conclude that Reading Mastery was no better than its comparison program. They ignored the fact that students in Reading Mastery performed significantly better than the national or state norms and that the comparison program was similar on all but a very few characteristics, essentially a modified version of Reading Mastery. To use a medical analogy, the WWC decision would be equivalent to a researcher claiming that aspirin was ineffective because Tylenol performed equally well.

The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman, 1997) involved two groups of students, both of whom appear to have received Reading Mastery as part of the schools' "usual and customary school day curriculum." One group of students also received 45 minutes of supplemental phonemic related instruction, from their regular classroom teachers, that involved motor activities to accompany practice of skills. Not surprisingly, the group receiving the additional instruction had significantly larger gains than those who did not have additional learning time. Despite these differences in exposure and the fact that both groups appear to have had Reading Mastery as their usual reading curriculum, the WWC used these results to suggest that Reading Mastery could have potentially negative effects. A much more accurate conclusion would be that students who receive additional phonics related instruction have greater skill growth.

I have prepared a more extensive analysis of these concerns, which I would be happy to send you. Because the problems with the analysis are so serious I believe that the report should be immediately taken off your website. In addition, given the very serious misinterpretations of the literature described above I will continue to examine WWC reports and procedures.

Please acknowledge receipt of my concerns and let me know the time line for your review.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction

From: WhatWorks
Sent: Thursday, August 23, 2012 4:44 PM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWC 3801)

Dear Dr. Stockard,

We received your email and attachment. We provided it to the QRT.

Thank you,

What Works Clearinghouse

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Thursday, August 23, 2012 2:41 PM
To: WhatWorks
Cc: jstockard@nifdi.org
Subject: Re: What Works Clearinghouse (WWC 3801)

Thank you for acknowledging receipt of my e-mail. I have attached a more lengthy analysis of both the review and related literature.
Please ask if more information would help in your review.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> Thank you for contacting the What Works Clearinghouse (WWC).
>
> We have received your email about the Reading Mastery Intervention Report.
> The WWC Quality Review Team is reviewing your email and will prepare
> a written response. The WWC Quality Review Team responds to concerns
> raised about WWC reviews published on our website. These quality
> reviews are undertaken when concerned parties present evidence that a
> WWC review may be inaccurate. When a quality review is conducted, a
> researcher who was not involved in the initial review undertakes an
> independent assessment of the study in question.
>
> If a quality review concludes that the original review contained
> errors, a revised version of the review will be published on the web

> site. However, it is WWC policy that the existing report will remain
> on the web site, and be revised only if the WWC team determines it
> does contain errors. These quality reviews are one of tools used to
> ensure that the standards established by the Institute of Educational
> Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.
>
> At this time, we are unable to provide a timeline for a QRT response.
> Please submit further comments that you would like given to the QRT.
> You may send them in response to this email.
>
> Thank you,
>
> What Works Clearinghouse
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more
> information, please visit http://ies.ed.gov/ncee/wwc/.
>
>
>
>
>
> -----Original Message-----
>
> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>
> Sent: Tuesday, August 21, 2012 8:54 PM
>
> To: info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>
>
> Subject: IES WWC Website: Contact Us: Publications, Intervention
> Reports, Reference ID Number: 1757800276
>
>
>
> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>, this email was
> automatically sent through the Contact link on the WWC website.
>
>
>
> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>
>
>
> Message: I am writing to request a Quality Review of the recent report

> on Reading Mastery for students with learning disabilities. I believe
> that the report misinterpreted both of the studies that were reviewed
> and should be removed from the website immediately.
>
>
>
> The WWC analysis was based on two articles. One (Cooke, Gibbs,
> Campbell, & Schalvis, 2004) compared two very similar Direct
> Instruction programs, Reading Mastery and Horizons, and found that
> students in both programs made gains over the academic year that were
> significantly greater than those made by students in national and
> state level populations. Gains in both programs were similar, leading
> the reviewers to conclude that Reading Mastery was no better than its
> comparison program. They ignored the fact that students in Reading
> Mastery performed significantly better than the national or state
> norms and that the comparison program was similar on all but a very
> few characteristics, essentially a modified version of Reading
> Mastery. To use a medical analogy, the WWC decision would be
> equivalent to a researcher claiming that aspirin was ineffective because Tylenol performed equally
well.
>
>
>
> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman,
> 1997) involved two groups of students, both of whom appear to have
> received Reading Mastery as part of the schools' "usual and customary
> school day curriculum." One group of students also received 45 minutes
> of supplemental phonemic related instruction, from their regular
> classroom teachers, that involved motor activities to accompany practice of skills.
> Not surprisingly, the group receiving the additional instruction had
> significantly larger gains than those who did not have additional
> learning time. Despite these differences in exposure and the fact that
> both groups appear to have had Reading Mastery as their usual reading
> curriculum, the WWC used these results to suggest that Reading Mastery
> could have potentially negative effects. A much more accurate
> conclusion would be that students who receive additional phonics
> related instruction have greater skill growth.
>
>
>
> I have prepared a more extensive analysis of these concerns, which I
> would be happy to send you. Because the problems with the analysis are
> so serious I believe that the report should be immediately taken off
> your website. In addition, given the very serious misinterpretations
> of the literature described above I will continue to examine WWC
> reports and procedures.
>
>

>
> Please acknowledge receipt of my concerns and let me know the time
> line for your review.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

From: WhatWorks
Sent: Wednesday, September 05, 2012 12:24 PM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear Dr. Stockard,

Thank you for your email. As was documented in the table notes in Appendix C.1 of the report, we attempted to contact the author (multiple times), but did not receive a response.

What Works Clearinghouse

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Tuesday, September 04, 2012 7:12 PM
To: WhatWorks
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear WWC Staff:
I am writing with regards to the article by Herrera and associates (1997) that was included in your recent review of Reading Mastery for Learning Disabled students. I wondered if you had any additional information on the Herrera et al. study that was obtained from the authors in the course of developing your report. If so, could I possibly have access to that information? I have been trying to contact them, but have not been successful.

Thank you very much for your quick responses to queries.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> We received your email and attachment. We provided it to the QRT.
>
> Thank you,
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Thursday, August 23, 2012 2:41 PM
> To: WhatWorks
> Cc: jstockard@nifdi.org

> Subject: Re: What Works Clearinghouse (WWC 3801)
>
> Thank you for acknowledging receipt of my e-mail. I have attached a
> more lengthy analysis of both the review and related literature.
> Please ask if more information would help in your review.
>
> Sincerely,
>
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>> Dear Dr. Stockard,
>>
>> Thank you for contacting the What Works Clearinghouse (WWC).
>>
>> We have received your email about the Reading Mastery Intervention
>> Report.
>> The WWC Quality Review Team is reviewing your email and will prepare
>> a written response. The WWC Quality Review Team responds to concerns
>> raised about WWC reviews published on our website. These quality
>> reviews are undertaken when concerned parties present evidence that a
>> WWC review may be inaccurate. When a quality review is conducted, a
>> researcher who was not involved in the initial review undertakes an
>> independent assessment of the study in question.
>>
>> If a quality review concludes that the original review contained
>> errors, a revised version of the review will be published on the web
>> site. However, it is WWC policy that the existing report will remain
>> on the web site, and be revised only if the WWC team determines it
>> does contain errors. These quality reviews are one of tools used to
>> ensure that the standards established by the Institute of Educational
>> Sciences (IES) are upheld on every review conducted by the What Works
>> Clearinghouse.
>>
>> At this time, we are unable to provide a timeline for a QRT response.
>> Please submit further comments that you would like given to the QRT.
>> You may send them in response to this email.
>>
>> Thank you,
>>
>> What Works Clearinghouse
>>
>> The What Works Clearinghouse was established by the U.S. Department
>> of Education's Institute of Education Sciences to provide educators,

>> policymakers, researchers, and the public with a central and trusted
>> source of scientific evidence of what works in education. For more
>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>
>>
>>
>>
>>
>> -----Original Message-----
>>
>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>> Sent: Tuesday, August 21, 2012 8:54 PM
>>
>> To: info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>
>>
>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>> Reports, Reference ID Number: 1757800276
>>
>>
>>
>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>, this email was
>> automatically sent through the Contact link on the WWC website.
>>
>>
>>
>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>
>>
>>
>> Message: I am writing to request a Quality Review of the recent
>> report on Reading Mastery for students with learning disabilities. I
>> believe that the report misinterpreted both of the studies that were
>> reviewed and should be removed from the website immediately.
>>
>>
>>
>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>> Campbell, & Schalvis, 2004) compared two very similar Direct
>> Instruction programs, Reading Mastery and Horizons, and found that
>> students in both programs made gains over the academic year that were
>> significantly greater than those made by students in national and
>> state level populations. Gains in both programs were similar, leading
>> the reviewers to conclude that Reading Mastery was no better than its
>> comparison program. They ignored the fact that students in Reading
>> Mastery performed significantly better than the national or state
>> norms and that the comparison program was similar on all but a very

>> few characteristics, essentially a modified version of Reading
>> Mastery. To use a medical analogy, the WWC decision would be
>> equivalent to a researcher claiming that aspirin was ineffective
>> because Tylenol performed equally well.
>>
>>
>>
>> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman,
>> 1997) involved two groups of students, both of whom appear to have
>> received Reading Mastery as part of the schools' "usual and customary
>> school day curriculum." One group of students also received 45
>> minutes of supplemental phonemic related instruction, from their
>> regular classroom teachers, that involved motor activities to
>> accompany practice of skills.
>> Not surprisingly, the group receiving the additional instruction had
>> significantly larger gains than those who did not have additional
>> learning time. Despite these differences in exposure and the fact
>> that both groups appear to have had Reading Mastery as their usual
>> reading curriculum, the WWC used these results to suggest that
>> Reading Mastery could have potentially negative effects. A much more
>> accurate conclusion would be that students who receive additional
>> phonics related instruction have greater skill growth.
>>
>>
>>
>> I have prepared a more extensive analysis of these concerns, which I
>> would be happy to send you. Because the problems with the analysis
>> are so serious I believe that the report should be immediately taken
>> off your website. In addition, given the very serious
>> misinterpretations of the literature described above I will continue
>> to examine WWC reports and procedures.
>>
>>
>>
>> Please acknowledge receipt of my concerns and let me know the time
>> line for your review.
>>
>>
>>
>> Sincerely,
>>
>>
>>
>> Jean Stockard, Ph.D.
>>
>> Director of Research
>>
>> National Institute for Direct Instruction

>>
>>
>>
>
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

From: Jean Stockard <jstockard@nifdi.org>
Sent: Wednesday, September 05, 2012 12:51 PM
To: WhatWorks
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Thank you very much - That has been my experience as well, but just wanted to check.

> Dear Dr. Stockard,
>
> Thank you for your email. As was documented in the table notes in Appendix
> C.1 of the report, we attempted to contact the author (multiple times),
> but did not receive a response.
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Tuesday, September 04, 2012 7:12 PM
> To: WhatWorks
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear WWC Staff:
> I am writing with regards to the article by Herrera and associates (1997)
> that was included in your recent review of Reading Mastery for Learning
> Disabled students. I wondered if you had any additional information on the
> Herrera et al. study that was obtained from the authors in the course of
> developing your report. If so, could I possibly have access to that
> information? I have been trying to contact them, but have not been
> successful.
>
> Thank you very much for your quick responses to queries.
>
> Sincerely,
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>> Dear Dr. Stockard,
>>
>> We received your email and attachment. We provided it to the QRT.
>>
>> Thank you,
>>

>> What Works Clearinghouse
>>
>> -----Original Message-----
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>> Sent: Thursday, August 23, 2012 2:41 PM
>> To: WhatWorks
>> Cc: jstockard@nifdi.org
>> Subject: Re: What Works Clearinghouse (WWC 3801)
>>
>> Thank you for acknowledging receipt of my e-mail. I have attached a
>> more lengthy analysis of both the review and related literature.
>> Please ask if more information would help in your review.
>>
>> Sincerely,
>>
>>
>> Jean Stockard, Ph.D.
>> Director of Research
>> National Institute for Direct Instruction Eugene, Oregon
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>> Dear Dr. Stockard,
>>>
>>> Thank you for contacting the What Works Clearinghouse (WWC).
>>>
>>> We have received your email about the Reading Mastery Intervention
>>> Report.
>>> The WWC Quality Review Team is reviewing your email and will prepare
>>> a written response. The WWC Quality Review Team responds to concerns
>>> raised about WWC reviews published on our website. These quality
>>> reviews are undertaken when concerned parties present evidence that a
>>> WWC review may be inaccurate. When a quality review is conducted, a
>>> researcher who was not involved in the initial review undertakes an
>>> independent assessment of the study in question.
>>>
>>> If a quality review concludes that the original review contained
>>> errors, a revised version of the review will be published on the web
>>> site. However, it is WWC policy that the existing report will remain
>>> on the web site, and be revised only if the WWC team determines it
>>> does contain errors. These quality reviews are one of tools used to
>>> ensure that the standards established by the Institute of Educational
>>> Sciences (IES) are upheld on every review conducted by the What Works
>>> Clearinghouse.
>>>
>>> At this time, we are unable to provide a timeline for a QRT response.
>>> Please submit further comments that you would like given to the QRT.
>>> You may send them in response to this email.

>>>
>>> Thank you,
>>>
>>> What Works Clearinghouse
>>>
>>> The What Works Clearinghouse was established by the U.S. Department
>>> of Education's Institute of Education Sciences to provide educators,
>>> policymakers, researchers, and the public with a central and trusted
>>> source of scientific evidence of what works in education. For more
>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>>
>>>
>>>
>>>
>>>
>>> -----Original Message-----
>>>
>>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>>
>>> Sent: Tuesday, August 21, 2012 8:54 PM
>>>
>>> To: info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>
>>>
>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>>> Reports, Reference ID Number: 1757800276
>>>
>>>
>>>
>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>, this email was
>>> automatically sent through the Contact link on the WWC website.
>>>
>>>
>>>
>>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>>
>>>
>>>
>>> Message: I am writing to request a Quality Review of the recent
>>> report on Reading Mastery for students with learning disabilities. I
>>> believe that the report misinterpreted both of the studies that were
>>> reviewed and should be removed from the website immediately.
>>>
>>>
>>>
>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>>> Instruction programs, Reading Mastery and Horizons, and found that

>>> students in both programs made gains over the academic year that were
>>> significantly greater than those made by students in national and
>>> state level populations. Gains in both programs were similar, leading
>>> the reviewers to conclude that Reading Mastery was no better than its
>>> comparison program. They ignored the fact that students in Reading
>>> Mastery performed significantly better than the national or state
>>> norms and that the comparison program was similar on all but a very
>>> few characteristics, essentially a modified version of Reading
>>> Mastery. To use a medical analogy, the WWC decision would be
>>> equivalent to a researcher claiming that aspirin was ineffective
>>> because Tylenol performed equally well.
>>>
>>>
>>>
>>> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman,
>>> 1997) involved two groups of students, both of whom appear to have
>>> received Reading Mastery as part of the schools' "usual and customary
>>> school day curriculum." One group of students also received 45
>>> minutes of supplemental phonemic related instruction, from their
>>> regular classroom teachers, that involved motor activities to
>>> accompany practice of skills.
>>> Not surprisingly, the group receiving the additional instruction had
>>> significantly larger gains than those who did not have additional
>>> learning time. Despite these differences in exposure and the fact
>>> that both groups appear to have had Reading Mastery as their usual
>>> reading curriculum, the WWC used these results to suggest that
>>> Reading Mastery could have potentially negative effects. A much more
>>> accurate conclusion would be that students who receive additional
>>> phonics related instruction have greater skill growth.
>>>
>>>
>>>
>>> I have prepared a more extensive analysis of these concerns, which I
>>> would be happy to send you. Because the problems with the analysis
>>> are so serious I believe that the report should be immediately taken
>>> off your website. In addition, given the very serious
>>> misinterpretations of the literature described above I will continue
>>> to examine WWC reports and procedures.
>>>
>>>
>>>
>>> Please acknowledge receipt of my concerns and let me know the time
>>> line for your review.
>>>
>>>
>>>
>>> Sincerely,
>>>

>>>
>>>
>>> Jean Stockard, Ph.D.
>>>
>>> Director of Research
>>>
>>> National Institute for Direct Instruction
>>>
>>>
>>>
>>
>>
>> Jean Stockard, Ph.D.
>> Director of Research
>> National Institute for Direct Instruction Eugene, Oregon
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

From: WhatWorks
Sent: Wednesday, October 03, 2012 9:39 AM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear Dr. Stockard,

The quality review is in process and we will expedite the inclusion of the information on the Herrera, et al. study that was provided yesterday.

What Works Clearinghouse

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Tuesday, October 02, 2012 9:59 AM
To: WhatWorks
Cc: jstockard@nifdi.org
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear WWC Staff,

Some recent events have prompted me to follow up on my request for a quality review of WWC Report on Reading Mastery for Learning Disabled Students (reference number is 1757800276), which was submitted on August 21, 2012.

First, we received confirmation from an author of one of the articles cited in the review (Dr. Philip Cooker, a co-author of the Herrera, et al.
study), that our interpretation of the work, as explained in my earlier correspondence, was correct. A copy of this correspondence is attached below.

Second, we asked highly regarded independent researchers to examine our concerns. They have confirmed our interpretation of the problems with the WWC report and note additional problems with WWC procedures and policies.

Third, our office has been told that, because of the WWC posting, schools are beginning to abandon the use of Reading Mastery with learning disabled students. In addition to the fiscal damages that will accrue to the authors and publisher from these decisions, such actions would deprive many students of access to a curriculum that has repeatedly been shown to be highly effective.

Finally, we were requested to respond to Senator Lisa Murkowski, whose office cited the WWC report in correspondence with schools. We have sent her our explanations of the serious errors in your review.

Given the very serious ramifications of your report, I again ask that you immediately remove it from your website and that you provide a clear timeline for the completion of your review including the steps taken to date. If you are not able to do so, we will need to pursue additional avenues to address this situation. Please let me know, as soon as possible, your intentions in this matter.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

Below is the correspondence from Dr. Cooker

From: "PHILIP G COOKER" <cooker@olemiss.edu>
Subject: RE: Query on article published in 1997 -
Date: Thu, September 6, 2012 9:12 am
To: "jstockard@nifdi.org" <jstockard@nifdi.org>

Hello,

sorry for the late response. I did not conduct the research, but edited and coordinated the efforts. Of the authors, I only knew Dr. Logan. It has been quite a while since the publication and I have little recollection of details. It sounds as if your interpretation is correct. Since reading is not my area of specialization, I do not have knowledge of other research involving the system that was used.

Philip G. Cooker, Ph.D.
Mississippi Board of Psychology
CE Coordinator
407 Timber Lane
Oxford, MS 38655
(662) 801-1664

From: Jean Stockard [jstockard@nifdi.org]
Sent: Tuesday, September 04, 2012 6:51 PM
To: jherreramdphd@gmail.com; PHILIP G COOKER
Subject: Query on article published in 1997 -

Dear Professors Herrera and Cooker,

This is a note that I sent a few weeks ago. I am afraid it may have gotten lost in your inbox.

I am writing for a quick clarification/check on two items in your article, “Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children,”
which was published in 1997 in Reading Improvement.

First, my understanding is that all students in the study had Reading Mastery as “the usual and customary prescribed curriculum for the school day” (p. 78), and that the intervention group was given the Stabilized Learning System “as a supplement to the usual and customary school day curriculum” (p.

78). In other words, both the control group and the intervention group had Reading Mastery for their usual curriculum, and the intervention group received additional instruction via the Stabilized Learning System. Is this interpretation correct?

Second, I've not been able to find other writings on the Stabilized Learning System. Do you know if there are other studies of the program?

Thank you very much for your help with my queries. If I can provide any clarification, please just ask.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> Thank you for your email. As was documented in the table notes in
> Appendix
> C.1 of the report, we attempted to contact the author (multiple
> times), but did not receive a response.
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Tuesday, September 04, 2012 7:12 PM
> To: WhatWorks
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear WWC Staff:
> I am writing with regards to the article by Herrera and associates
> (1997) that was included in your recent review of Reading Mastery for
> Learning Disabled students. I wondered if you had any additional
> information on the Herrera et al. study that was obtained from the
> authors in the course of developing your report. If so, could I
> possibly have access to that information? I have been trying to
> contact them, but have not been successful.
>
> Thank you very much for your quick responses to queries.
>
> Sincerely,
>

> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>> Dear Dr. Stockard,
>>
>> We received your email and attachment. We provided it to the QRT.
>>
>> Thank you,
>>
>> What Works Clearinghouse
>>
>> -----Original Message-----
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>> Sent: Thursday, August 23, 2012 2:41 PM
>> To: WhatWorks
>> Cc: jstockard@nifdi.org
>> Subject: Re: What Works Clearinghouse (WWC 3801)
>>
>> Thank you for acknowledging receipt of my e-mail. I have attached a
>> more lengthy analysis of both the review and related literature.
>> Please ask if more information would help in your review.
>>
>> Sincerely,
>>
>>
>> Jean Stockard, Ph.D.
>> Director of Research
>> National Institute for Direct Instruction Eugene, Oregon
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>> Dear Dr. Stockard,
>>>
>>> Thank you for contacting the What Works Clearinghouse (WWC).
>>>
>>> We have received your email about the Reading Mastery Intervention
>>> Report.
>>> The WWC Quality Review Team is reviewing your email and will
>>> prepare a written response. The WWC Quality Review Team responds to
>>> concerns raised about WWC reviews published on our website. These
>>> quality reviews are undertaken when concerned parties present
>>> evidence that a WWC review may be inaccurate. When a quality review
>>> is conducted, a researcher who was not involved in the initial
>>> review undertakes an independent assessment of the study in question.
>>>
>>> If a quality review concludes that the original review contained

>>> errors, a revised version of the review will be published on the web
>>> site. However, it is WWC policy that the existing report will remain
>>> on the web site, and be revised only if the WWC team determines it
>>> does contain errors. These quality reviews are one of tools used to
>>> ensure that the standards established by the Institute of
>>> Educational Sciences (IES) are upheld on every review conducted by
>>> the What Works Clearinghouse.
>>>
>>> At this time, we are unable to provide a timeline for a QRT response.
>>> Please submit further comments that you would like given to the QRT.
>>> You may send them in response to this email.
>>>
>>> Thank you,
>>>
>>> What Works Clearinghouse
>>>
>>> The What Works Clearinghouse was established by the U.S. Department
>>> of Education's Institute of Education Sciences to provide educators,
>>> policymakers, researchers, and the public with a central and trusted
>>> source of scientific evidence of what works in education. For more
>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>>
>>>
>>>
>>>
>>>
>>> -----Original Message-----
>>>
>>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>>
>>> Sent: Tuesday, August 21, 2012 8:54 PM
>>>
>>> To: info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>
>>>
>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>>> Reports, Reference ID Number: 1757800276
>>>
>>>
>>>
>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov>, this email was
>>> automatically sent through the Contact link on the WWC website.
>>>
>>>
>>>
>>> From: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>>
>>>

>>>
>>> Message: I am writing to request a Quality Review of the recent
>>> report on Reading Mastery for students with learning disabilities. I
>>> believe that the report misinterpreted both of the studies that were
>>> reviewed and should be removed from the website immediately.
>>>
>>>
>>>
>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>>> Instruction programs, Reading Mastery and Horizons, and found that
>>> students in both programs made gains over the academic year that
>>> were significantly greater than those made by students in national
>>> and state level populations. Gains in both programs were similar,
>>> leading the reviewers to conclude that Reading Mastery was no better
>>> than its comparison program. They ignored the fact that students in
>>> Reading Mastery performed significantly better than the national or
>>> state norms and that the comparison program was similar on all but a
>>> very few characteristics, essentially a modified version of Reading
>>> Mastery. To use a medical analogy, the WWC decision would be
>>> equivalent to a researcher claiming that aspirin was ineffective
>>> because Tylenol performed equally well.
>>>
>>>
>>>
>>> The second article reviewed (Herrera, Logan, Cooke, Morris, & Lyman,
>>> 1997) involved two groups of students, both of whom appear to have
>>> received Reading Mastery as part of the schools' "usual and
>>> customary school day curriculum." One group of students also
>>> received 45 minutes of supplemental phonemic related instruction,
>>> from their regular classroom teachers, that involved motor
>>> activities to accompany practice of skills.
>>> Not surprisingly, the group receiving the additional instruction had
>>> significantly larger gains than those who did not have additional
>>> learning time. Despite these differences in exposure and the fact
>>> that both groups appear to have had Reading Mastery as their usual
>>> reading curriculum, the WWC used these results to suggest that
>>> Reading Mastery could have potentially negative effects. A much more
>>> accurate conclusion would be that students who receive additional
>>> phonics related instruction have greater skill growth.
>>>
>>>
>>>
>>> I have prepared a more extensive analysis of these concerns, which I
>>> would be happy to send you. Because the problems with the analysis
>>> are so serious I believe that the report should be immediately taken
>>> off your website. In addition, given the very serious
>>> misinterpretations of the literature described above I will continue

>>> to examine WWC reports and procedures.
>>>
>>>
>>>
>>> Please acknowledge receipt of my concerns and let me know the time
>>> line for your review.
>>>
>>>
>>>
>>> Sincerely,
>>>
>>>
>>>
>>> Jean Stockard, Ph.D.
>>>
>>> Director of Research
>>>
>>> National Institute for Direct Instruction
>>>
>>>
>>>
>>
>>
>> Jean Stockard, Ph.D.
>> Director of Research
>> National Institute for Direct Instruction Eugene, Oregon
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>

From: WhatWorks
Sent: Thursday, October 11, 2012 10:11 AM
To: 'Doug Carnine'
Cc: Joy.Lesnick@ed.gov
Subject: RE: Reading Mastery report by WWC

Dear Dr. Carnine,

The quality review of the WWC Special Needs report on Reading Mastery is in process and is being expedited as quickly as possible.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: Doug Carnine [mailto:dcarnine@uoregon.edu]
Sent: Wednesday, October 10, 2012 1:43 PM
To: Joy.Lesnick@ed.gov
Cc: WhatWorks; jstockard@nifdi.org
Subject: Reading Mastery report by WWC

The following email is also attached: IES-WWC email

10-10-12

Dear Dr. Lesnick:

It is painful and depressing to have to come out of retirement, briefly, to write you about the inaccuracies of the recent report posted on the WWC website regarding the Direct Instruction program, Reading Mastery, and its effect on learning disabled students. Dr. Jean Stockard, research director of the National Institute for Direct Instruction, submitted a request for a quality review of the Reading Mastery report on August 21. Here is a summary of the problems Dr.
Stockard found with the report:

The WWC analysis was based on only two articles, a fraction of the available studies. One of the articles (Cooke, Gibbs, Campbell, & Schalvis,
2004)
compared two very similar Direct Instruction programs, Reading Mastery and Horizons that were written by the same senior author (Siegfried Engelmann). The study found that students in both programs made gains over the academic year that were significantly greater than would be expected given national norms (of the Woodcock Johnson-Revised) or the performance of other students in the state (on the state K-2 literacy test) (see page 146 of the article).
Gains in both programs were similar, leading the WWC to conclude that Reading Mastery was no better than its comparison program. They ignored the fact that students in Reading Mastery performed significantly better than the national or state norms and that the comparison program was similar on all

but a very few characteristics, essentially a modified version of Reading Mastery. To use a medical analogy, the WWC decision would be equivalent to a researcher who compared ibuprofen and aspirin finding similar positive effects for both and then claiming that aspirin was ineffective. (I realize this is a challenge in research with two treatments rather than a treatment and a control.)

The second article (Herrera, Logan, Cooke, Morris, & Lyman, 1997) involved two groups of students, both of whom appear to have received Reading Mastery as part of the schools' ³usual and customary school day curriculum.² The article was not clearly written. After reading it several times, Dr. Stockard interpreted the article to indicate that one group of students also received 45 minutes of supplemental phonemic related instruction, from their regular classroom teachers. Dr. Stockard's interpretation was tentatively confirmed in an e-mail sent by the second author of the study. (Correspondence of Dr. Stockard with the WWC indicated that they had unsuccessfully tried to contact only the senior author). Not surprisingly, the group receiving the additional instruction had significantly larger gains than those who did not have additional learning time. Despite these differences in exposure and the fact that both groups had Reading Mastery as their usual reading curriculum, the authors of the WWC report used these results to conclude that Reading Mastery could have potentially negative effects. The accurate conclusion would be that students who receive additional reading-related instruction had greater growth in reading related skills (as measured by a district developed curriculum based measure of vocabulary and reading, see pp.
78-79). To again use a medical analogy, this would be equivalent to comparing the impact of taking one aspirin or one aspirin and one Tylenol and, upon finding that one Tylenol and one aspirin resulted in better pain relief, concluding that one aspirin had a ³potentially negative² effect.

There have been earlier concerns about the quality of the reviews of Direct Instruction programs conducted by WWC. This quote is from a September 8, 2008 communication from Mathematica in response to an inquiry by Dr. Stockard about the exclusion of a study on the effects of a Reading Mastery implementation sponsored by the Rodeo Institute for Teacher Effectiveness (RITE).
Coleen Carlson and David Francis conducted the study. ³A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program² (Dynarski, p. 5). The Carlson and Francis study involved thousands of students, hundreds of teachers, and dozens of schools and used sophisticated multivariate techniques to analyze the data. The study was excluded because the teachers received training, which might suggest that any evaluation of a literacy program in which professional development is being conducted at the school should be excluded because of the potential confound. Alternatively the study should have been reviewed in the category of school reform model, as is the case for SFA.
When I contrast the excluded Carlson and Francis study with the Herrera, Logan, Cooke, Morris, & Lyman, (1997) study in which students received Reading Mastery and an additional 45 minutes of reading instruction, I can only say I am perplexed. Does not 45 minutes of extra reading instruction constitute a confound? According to the 2011 WWC guidelines this study should not have met standards for inclusion: ³A study may fail to meet WWC standards if the measures of effect cannot be attributed solely to the intervention-the intervention was combined with another intervention² (pp.15).

I have attached Dr. Stockard's most recent appeal to WWC, where you can read the reasons for the request that the review be immediately removed from the WWC website until the review can be completed. I am going to suggest that Dr. Jean
Stockard send you the earlier submitted appeals that have resulted in inadequate responses.

As you may know, several meta-analyses in several peer-reviewed journals have consistently found strong support for the efficacy of Direct Instruction programs, including Reading Mastery.
Could the current review be biased because of an inadequate literature search? The 2011 WWC guidelines state their intervention reports [3]summarize all studies published during a specific time period that examine the effectiveness of an intervention[2] (pp.1). However several relevant studies were excluded from the report. Even if some of the studies did not meet the WWC standards they should have been referenced in the report for not meeting standards.

If you have questions about the specifics of this letter or other documents that have been submitted, please submit them to Dr. Jean Stockard, whose contact information follows.

Sincerely,

Doug Carnine, Professor Emeritus
University of Oregon
Phone: 541 485 3781

Cc: Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
Phone: 541-505-5710 or toll-free 877-485-1973 jstockard@nifdi.org

From: WhatWorks
Sent: Monday, November 05, 2012 5:45 PM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear Dr. Stockard,

Thank you for your email. You will be receiving a letter summarizing the findings of the quality review. We anticipate sending the letter via email within the next week.

Thank you,

What Works Clearinghouse

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Friday, November 02, 2012 3:09 PM
To: WhatWorks
Cc: jstockard@nifdi.org
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear What Works Clearinghouse Help Desk,

I am writing in regards to my request for a quality review of the WWC Report on Reading Mastery for Learning Disabled Students (reference number is 1757800276), which was submitted on August 21, 2012. I sent additional information regarding the request a few weeks later, and Dr. Keith Smolkowski, an independent researcher, sent an extensive memo regarding the review last week.

The WWC Procedures and Standards Handbook (p. 5) notes that if a Quality Review results in a validation of the WWC's original assessment the person requesting the review is to receive a response. If the original assessment is not validated, reports are made to others within the WWC and IES. In this case, there is, however, no stated provision for notifying the person who asked for the review.

More than two months have elapsed since my original request for a Quality Review. I must assume, given the length of time since my request was submitted and the relatively simple job of reviewing the two articles involved, that my concerns with the WWC report on Reading Mastery and Learning Disabled students have been validated but that, per the stated procedures, I have not been informed of this decision.

As in my earlier correspondences, which are appended to this e-mail, I ask that you immediately remove the erroneous report from the web and clearly indicate that previous postings provided incorrect information. In addition to producing possible fiscal damages for the authors and publisher of the Reading Mastery program, continued dissemination gives very erroneous information to the public.

If I am mistaken in my assumption regarding the outcome of the review, please send me the details of your decision.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> The quality review is in process and we will expedite the inclusion of
> the information on the Herrera, et al. study that was provided yesterday.
>
>
>
> What Works Clearinghouse
>
>
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Tuesday, October 02, 2012 9:59 AM
> To: WhatWorks
> Cc: jstockard@nifdi.org
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
>
>
> Dear WWC Staff,
>
>
>
> Some recent events have prompted me to follow up on my request for a
> quality review of WWC Report on Reading Mastery for Learning Disabled
> Students (reference number is 1757800276), which was submitted on
> August 21, 2012.
>
>
>
> First, we received confirmation from an author of one of the articles
> cited in the review (Dr. Philip Cooker, a co-author of the Herrera, et
> al.
>
> study), that our interpretation of the work, as explained in my
> earlier correspondence, was correct. A copy of this correspondence is
> attached below.

>
>
>
> Second, we asked highly regarded independent researchers to examine
> our concerns. They have confirmed our interpretation of the problems
> with the WWC report and note additional problems with WWC procedures and policies.
>
>
>
> Third, our office has been told that, because of the WWC posting,
> schools are beginning to abandon the use of Reading Mastery with
> learning disabled students. In addition to the fiscal damages that
> will accrue to the authors and publisher from these decisions, such
> actions would deprive many students of access to a curriculum that has
> repeatedly been shown to be highly effective.
>
>
>
> Finally, we were requested to respond to Senator Lisa Murkowski, whose
> office cited the WWC report in correspondence with schools. We have
> sent her our explanations of the serious errors in your review.
>
>
>
> Given the very serious ramifications of your report, I again ask that
> you immediately remove it from your website and that you provide a
> clear timeline for the completion of your review including the steps
> taken to date. If you are not able to do so, we will need to pursue
> additional avenues to address this situation. Please let me know, as
> soon as possible, your intentions in this matter.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction Eugene, Oregon
>
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>
> Below is the correspondence from Dr. Cooker

>
>
>
> From: "PHILIP G COOKER"
> <cooker@olemiss.edu<mailto:cooker@olemiss.edu>>
>
> Subject: RE: Query on article published in 1997 -
>
> Date: Thu, September 6, 2012 9:12 am
>
> To: "jstockard@nifdi.org<mailto:jstockard@nifdi.org>"
> <jstockard@nifdi.org<mailto:jstockard@nifdi.org>>
>
>
>
> _____
>
>
>
> Hello,
>
>
>
> sorry for the late response. I did not conduct the research, but
> edited and coordinated the efforts. Of the authors, I only knew Dr.
> Logan. It has been quite a while since the publication and I have
> little recollection of details. It sounds as if your interpretation is
> correct. Since reading is not my area of specialization, I do not have
> knowledge of other research involving the system that was used.
>
>
>
> Philip G. Cooker, Ph.D.
>
> Mississippi Board of Psychology
>
> CE Coordinator
>
> 407 Timber Lane
>
> Oxford, MS 38655
>
> (662) 801-1664
>
>
>
> _____
>

> From: Jean Stockard [jstockard@nifdi.org]
>
> Sent: Tuesday, September 04, 2012 6:51 PM
>
> To: jherreramdphd@gmail.com<mailto:jherreramdphd@gmail.com>; PHILIP G
> COOKER
>
> Subject: Query on article published in 1997 -
>
>
>
> Dear Professors Herrera and Cooker,
>
>
>
> This is a note that I sent a few weeks ago. I am afraid it may have
> gotten lost in your inbox.
>
>
>
> I am writing for a quick clarification/check on two items in your
> article, “Phonological awareness and phonetic-graphic conversion: A
> study of the effects of two intervention paradigms with learning disabled children,”
>
> which was published in 1997 in Reading Improvement.
>
>
>
> First, my understanding is that all students in the study had Reading
> Mastery as “the usual and customary prescribed curriculum for the
> school day” (p. 78), and that the intervention group was given the
> Stabilized Learning System “as a supplement to the usual and customary
> school day curriculum” (p. 78). In other words, both the control group
> and the intervention group had Reading Mastery for their usual
> curriculum, and the intervention group received additional instruction
> via the Stabilized Learning System. Is this interpretation correct?
>
>
>
> Second, I’ve not been able to find other writings on the Stabilized
> Learning System. Do you know if there are other studies of the program?
>
>
>
> Thank you very much for your help with my queries. If I can provide
> any clarification, please just ask.
>
>

>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction Eugene, Oregon
>
> phone: 541-505-5710 or toll-free 877-485-1973
>
>
>
>
>
>
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>> Thank you for your email. As was documented in the table notes in
>
>> Appendix
>
>> C.1 of the report, we attempted to contact the author (multiple
>
>> times), but did not receive a response.
>
>>
>
>> What Works Clearinghouse
>
>>
>
>> -----Original Message-----
>
>> From: Jean Stockard
>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>
>> Sent: Tuesday, September 04, 2012 7:12 PM
>
>> To: WhatWorks
>

>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>
>> 1757800276
>
>>
>
>> Dear WWC Staff:
>
>> I am writing with regards to the article by Herrera and associates
>
>> (1997) that was included in your recent review of Reading Mastery for
>
>> Learning Disabled students. I wondered if you had any additional
>
>> information on the Herrera et al. study that was obtained from the
>
>> authors in the course of developing your report. If so, could I
>
>> possibly have access to that information? I have been trying to
>
>> contact them, but have not been successful.
>
>>
>
>> Thank you very much for your quick responses to queries.
>
>>
>
>> Sincerely,
>
>>
>
>> Jean Stockard, Ph.D.
>
>> Director of Research
>
>> National Institute for Direct Instruction Eugene, Oregon
>
>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>
>
>>> Dear Dr. Stockard,
>
>>>
>
>>> We received your email and attachment. We provided it to the QRT.
>

>>>
>
>>> Thank you,
>
>>>
>
>>> What Works Clearinghouse
>
>>>
>
>>> -----Original Message-----
>
>>> From: Jean Stockard
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>
>>> Sent: Thursday, August 23, 2012 2:41 PM
>
>>> To: WhatWorks
>
>>> Cc: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>
>>> Subject: Re: What Works Clearinghouse (WWC 3801)
>
>>>
>
>>> Thank you for acknowledging receipt of my e-mail. I have attached a
>
>>> more lengthy analysis of both the review and related literature.
>
>>> Please ask if more information would help in your review.
>
>>>
>
>>> Sincerely,
>
>>>
>
>>>
>
>>> Jean Stockard, Ph.D.
>
>>> Director of Research
>
>>> National Institute for Direct Instruction Eugene, Oregon
>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>>

>
>>>
>
>>>> Dear Dr. Stockard,
>
>>>>
>
>>>> Thank you for contacting the What Works Clearinghouse (WWC).
>
>>>>
>
>>>> We have received your email about the Reading Mastery Intervention
>
>>>> Report.
>
>>>> The WWC Quality Review Team is reviewing your email and will
>
>>>> prepare a written response. The WWC Quality Review Team responds to
>
>>>> concerns raised about WWC reviews published on our website. These
>
>>>> quality reviews are undertaken when concerned parties present
>
>>>> evidence that a WWC review may be inaccurate. When a quality review
>
>>>> is conducted, a researcher who was not involved in the initial
>
>>>> review undertakes an independent assessment of the study in question.
>
>>>>
>
>>>> If a quality review concludes that the original review contained
>
>>>> errors, a revised version of the review will be published on the
>>>> web
>
>>>> site. However, it is WWC policy that the existing report will
>>>> remain
>
>>>> on the web site, and be revised only if the WWC team determines it
>
>>>> does contain errors. These quality reviews are one of tools used to
>
>>>> ensure that the standards established by the Institute of
>
>>>> Educational Sciences (IES) are upheld on every review conducted by
>
>>>> the What Works Clearinghouse.

>
>>>>
>
>>>> At this time, we are unable to provide a timeline for a QRT response.
>
>>>> Please submit further comments that you would like given to the QRT.
>
>>>> You may send them in response to this email.
>
>>>>
>
>>>> Thank you,
>
>>>>
>
>>>> What Works Clearinghouse
>
>>>>
>
>>>> The What Works Clearinghouse was established by the U.S. Department
>
>>>> of Education's Institute of Education Sciences to provide
>>>> educators,
>
>>>> policymakers, researchers, and the public with a central and
>>>> trusted
>
>>>> source of scientific evidence of what works in education. For more
>
>>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>> -----Original Message-----
>
>>>>
>
>>>> From:
>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@nif
>>>> di.org%3cmailto:jstockard@nifdi.org>>

>
>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]><m
>>>> ailto:[mailto:jstockard@nifdi.org]>
>
>>>>
>
>>>> Sent: Tuesday, August 21, 2012 8:54 PM
>
>>>>
>
>>>> To:
>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@what
>>>> works.ed.gov%3cmailto:info@whatworks.ed.gov>>
>
>>>>
>
>>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>
>>>> Reports, Reference ID Number: 1757800276
>
>>>>
>
>>>>
>
>>>>
>
>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@what
>>>> works.ed.gov%3cmailto:info@whatworks.ed.gov>>,
>>>> this email was
>
>>>> automatically sent through the Contact link on the WWC website.
>
>>>>
>
>>>>
>
>>>>
>
>>>> From:
>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@nif
>>>> di.org%3cmailto:jstockard@nifdi.org>>
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>

>>>> Message: I am writing to request a Quality Review of the recent
>
>>>> report on Reading Mastery for students with learning disabilities.
>>>> I
>
>>>> believe that the report misinterpreted both of the studies that
>>>> were
>
>>>> reviewed and should be removed from the website immediately.
>
>>>>
>
>>>>
>
>>>>
>
>>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>
>>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>
>>>> Instruction programs, Reading Mastery and Horizons, and found that
>
>>>> students in both programs made gains over the academic year that
>
>>>> were significantly greater than those made by students in national
>
>>>> and state level populations. Gains in both programs were similar,
>
>>>> leading the reviewers to conclude that Reading Mastery was no
>>>> better
>
>>>> than its comparison program. They ignored the fact that students in
>
>>>> Reading Mastery performed significantly better than the national or
>
>>>> state norms and that the comparison program was similar on all but
>>>> a
>
>>>> very few characteristics, essentially a modified version of Reading
>
>>>> Mastery. To use a medical analogy, the WWC decision would be
>
>>>> equivalent to a researcher claiming that aspirin was ineffective
>
>>>> because Tylenol performed equally well.
>
>>>>
>

>>>>
>
>>>>
>
>>>> The second article reviewed (Herrera, Logan, Cooke, Morris, &
>>>> Lyman,
>
>>>> 1997) involved two groups of students, both of whom appear to have
>
>>>> received Reading Mastery as part of the schools' "usual and
>
>>>> customary school day curriculum." One group of students also
>
>>>> received 45 minutes of supplemental phonemic related instruction,
>
>>>> from their regular classroom teachers, that involved motor
>
>>>> activities to accompany practice of skills.
>
>>>> Not surprisingly, the group receiving the additional instruction
>>>> had
>
>>>> significantly larger gains than those who did not have additional
>
>>>> learning time. Despite these differences in exposure and the fact
>
>>>> that both groups appear to have had Reading Mastery as their usual
>
>>>> reading curriculum, the WWC used these results to suggest that
>
>>>> Reading Mastery could have potentially negative effects. A much
>>>> more
>
>>>> accurate conclusion would be that students who receive additional
>
>>>> phonics related instruction have greater skill growth.
>
>>>>
>
>>>>
>
>>>>
>
>>>>
>
>>>> I have prepared a more extensive analysis of these concerns, which
>>>> I
>
>>>> would be happy to send you. Because the problems with the analysis
>

>>>> are so serious I believe that the report should be immediately
>>>> taken
>
>>>> off your website. In addition, given the very serious
>
>>>> misinterpretations of the literature described above I will
>>>> continue
>
>>>> to examine WWC reports and procedures.
>
>>>>
>
>>>>
>
>>>>
>
>>>> Please acknowledge receipt of my concerns and let me know the time
>
>>>> line for your review.
>
>>>>
>
>>>>
>
>>>>
>
>>>> Sincerely,
>
>>>>
>
>>>>
>
>>>>
>
>>>> Jean Stockard, Ph.D.
>
>>>>
>
>>>> Director of Research
>
>>>>
>
>>>> National Institute for Direct Instruction
>
>>>>
>
>>>>
>

>>>>
>
>>>
>
>>>
>
>>> Jean Stockard, Ph.D.
>
>>> Director of Research
>
>>> National Institute for Direct Instruction Eugene, Oregon
>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>>
>
>>
>
>>
>
>> Jean Stockard, Ph.D.
>
>> Director of Research
>
>> National Institute for Direct Instruction Eugene, Oregon
>
>> phone: 541-505-5710 or toll-free 877-485-1973
>
>>
>
>>
>
>>
>
>
>
>
>
>
>
>
>
>
>
>

From: Jean Stockard <jstockard@nifdi.org>
Sent: Monday, November 19, 2012 11:47 AM
To: WhatWorks
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear WWC Help Desk,

On November 5 you indicated that I would "be receiving a letter summarizing the findings of the quality review....via email within the next week." (The message is below.) To date (November 19), I have not received such a letter. Because I have been out of the country for a while, it is possible that the communication did not come through. If that is the case, could you please resend it? If it has not yet been sent, can you please let me know when it will come? As you know, I am very concerned about the posting of erroneous and damaging material on your website.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> Thank you for your email. You will be receiving a letter summarizing the
> findings of the quality review. We anticipate sending the letter via
> email within the next week.
>
> Thank you,
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Friday, November 02, 2012 3:09 PM
> To: WhatWorks
> Cc: jstockard@nifdi.org
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear What Works Clearinghouse Help Desk,
>
> I am writing in regards to my request for a quality review of the WWC
> Report on Reading Mastery for Learning Disabled Students (reference number
> is 1757800276), which was submitted on August 21, 2012. I sent additional

> information regarding the request a few weeks later, and Dr. Keith
> Smolkowski, an independent researcher, sent an extensive memo regarding
> the review last week.
>
> The WWC Procedures and Standards Handbook (p. 5) notes that if a Quality
> Review results in a validation of the WWC's original assessment the person
> requesting the review is to receive a response. If the original assessment
> is not validated, reports are made to others within the WWC and IES. In
> this case, there is, however, no stated provision for notifying the person
> who asked for the review.
>
> More than two months have elapsed since my original request for a Quality
> Review. I must assume, given the length of time since my request was
> submitted and the relatively simple job of reviewing the two articles
> involved, that my concerns with the WWC report on Reading Mastery and
> Learning Disabled students have been validated but that, per the stated
> procedures, I have not been informed of this decision.
>
> As in my earlier correspondences, which are appended to this e-mail, I ask
> that you immediately remove the erroneous report from the web and clearly
> indicate that previous postings provided incorrect information. In
> addition to producing possible fiscal damages for the authors and
> publisher of the Reading Mastery program, continued dissemination gives
> very erroneous information to the public.
>
> If I am mistaken in my assumption regarding the outcome of the review,
> please send me the details of your decision.
>
> Sincerely,
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>> Dear Dr. Stockard,
>>
>>
>>
>> The quality review is in process and we will expedite the inclusion of
>> the information on the Herrera, et al. study that was provided
>> yesterday.
>>
>>
>>
>> What Works Clearinghouse
>>
>>

>>
>> -----Original Message-----
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>> Sent: Tuesday, October 02, 2012 9:59 AM
>> To: WhatWorks
>> Cc: jstockard@nifdi.org
>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>> 1757800276
>>
>>
>>
>> Dear WWC Staff,
>>
>>
>>
>> Some recent events have prompted me to follow up on my request for a
>> quality review of WWC Report on Reading Mastery for Learning Disabled
>> Students (reference number is 1757800276), which was submitted on
>> August 21, 2012.
>>
>>
>>
>> First, we received confirmation from an author of one of the articles
>> cited in the review (Dr. Philip Cooker, a co-author of the Herrera, et
>> al.
>>
>> study), that our interpretation of the work, as explained in my
>> earlier correspondence, was correct. A copy of this correspondence is
>> attached below.
>>
>>
>>
>> Second, we asked highly regarded independent researchers to examine
>> our concerns. They have confirmed our interpretation of the problems
>> with the WWC report and note additional problems with WWC procedures and
>> policies.
>>
>>
>>
>> Third, our office has been told that, because of the WWC posting,
>> schools are beginning to abandon the use of Reading Mastery with
>> learning disabled students. In addition to the fiscal damages that
>> will accrue to the authors and publisher from these decisions, such
>> actions would deprive many students of access to a curriculum that has
>> repeatedly been shown to be highly effective.
>>
>>
>>

>> Finally, we were requested to respond to Senator Lisa Murkowski, whose
>> office cited the WWC report in correspondence with schools. We have
>> sent her our explanations of the serious errors in your review.
>>
>>
>>
>> Given the very serious ramifications of your report, I again ask that
>> you immediately remove it from your website and that you provide a
>> clear timeline for the completion of your review including the steps
>> taken to date. If you are not able to do so, we will need to pursue
>> additional avenues to address this situation. Please let me know, as
>> soon as possible, your intentions in this matter.
>>
>>
>>
>> Sincerely,
>>
>>
>>
>> Jean Stockard, Ph.D.
>>
>> Director of Research
>>
>> National Institute for Direct Instruction Eugene, Oregon
>>
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>
>> Below is the correspondence from Dr. Cooker
>>
>>
>>
>> From: "PHILIP G COOKER"
>> <cooker@olemiss.edu<mailto:cooker@olemiss.edu>>
>>
>> Subject: RE: Query on article published in 1997 -
>>
>> Date: Thu, September 6, 2012 9:12 am
>>
>> To: "jstockard@nifdi.org<mailto:jstockard@nifdi.org>"
>> <jstockard@nifdi.org<mailto:jstockard@nifdi.org>>
>>
>>
>>
>> _____
>>
>>

>>
>> Hello,
>>
>>
>>
>> sorry for the late response. I did not conduct the research, but
>> edited and coordinated the efforts. Of the authors, I only knew Dr.
>> Logan. It has been quite a while since the publication and I have
>> little recollection of details. It sounds as if your interpretation is
>> correct. Since reading is not my area of specialization, I do not have
>> knowledge of other research involving the system that was used.
>>
>>
>>
>> Philip G. Cooker, Ph.D.
>>
>> Mississippi Board of Psychology
>>
>> CE Coordinator
>>
>> 407 Timber Lane
>>
>> Oxford, MS 38655
>>
>> (662) 801-1664
>>
>>
>>
>> _____
>>
>> From: Jean Stockard [jstockard@nifdi.org]
>>
>> Sent: Tuesday, September 04, 2012 6:51 PM
>>
>> To: (b)(6) PHILIP G
>> COOKER
>>
>> Subject: Query on article published in 1997 -
>>
>>
>>
>> Dear Professors Herrera and Cooker,
>>
>>
>>
>> This is a note that I sent a few weeks ago. I am afraid it may have
>> gotten lost in your inbox.
>>

>>
>>
>> I am writing for a quick clarification/check on two items in your
>> article, “Phonological awareness and phonetic-graphic conversion: A
>> study of the effects of two intervention paradigms with learning
>> disabled children,”
>>
>> which was published in 1997 in Reading Improvement.
>>
>>
>>
>> First, my understanding is that all students in the study had Reading
>> Mastery as “the usual and customary prescribed curriculum for the
>> school day” (p. 78), and that the intervention group was given the
>> Stabilized Learning System “as a supplement to the usual and customary
>> school day curriculum” (p. 78). In other words, both the control group
>> and the intervention group had Reading Mastery for their usual
>> curriculum, and the intervention group received additional instruction
>> via the Stabilized Learning System. Is this interpretation correct?
>>
>>
>>
>> Second, I’ve not been able to find other writings on the Stabilized
>> Learning System. Do you know if there are other studies of the program?
>>
>>
>>
>> Thank you very much for your help with my queries. If I can provide
>> any clarification, please just ask.
>>
>>
>>
>> Sincerely,
>>
>>
>>
>> Jean Stockard, Ph.D.
>>
>> Director of Research
>>
>> National Institute for Direct Instruction Eugene, Oregon
>>
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>
>>
>>

>>
>>
>>
>>
>>> Dear Dr. Stockard,
>>
>>>
>>
>>> Thank you for your email. As was documented in the table notes in
>>
>>> Appendix
>>
>>> C.1 of the report, we attempted to contact the author (multiple
>>
>>> times), but did not receive a response.
>>
>>>
>>
>>> What Works Clearinghouse
>>
>>>
>>
>>> -----Original Message-----
>>
>>> From: Jean Stockard
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>>> Sent: Tuesday, September 04, 2012 7:12 PM
>>
>>> To: WhatWorks
>>
>>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>>
>>> 1757800276
>>
>>>
>>
>>> Dear WWC Staff:
>>
>>> I am writing with regards to the article by Herrera and associates
>>
>>> (1997) that was included in your recent review of Reading Mastery for
>>
>>> Learning Disabled students. I wondered if you had any additional
>>
>>> information on the Herrera et al. study that was obtained from the
>>
>>> authors in the course of developing your report. If so, could I

>>
>>> possibly have access to that information? I have been trying to
>>
>>> contact them, but have not been successful.
>>
>>>
>>
>>> Thank you very much for your quick responses to queries.
>>
>>>
>>
>>> Sincerely,
>>
>>>
>>
>>> Jean Stockard, Ph.D.
>>
>>> Director of Research
>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>
>>
>>>> Dear Dr. Stockard,
>>
>>>>
>>
>>>> We received your email and attachment. We provided it to the QRT.
>>
>>>>
>>
>>>> Thank you,
>>
>>>>
>>
>>>> What Works Clearinghouse
>>
>>>>
>>
>>>> -----Original Message-----
>>
>>>> From: Jean Stockard
>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>>>> Sent: Thursday, August 23, 2012 2:41 PM
>>

>>>> To: WhatWorks
>>
>>>> Cc: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>
>>>> Subject: Re: What Works Clearinghouse (WWC 3801)
>>
>>>>
>>
>>>> Thank you for acknowledging receipt of my e-mail. I have attached a
>>
>>>> more lengthy analysis of both the review and related literature.
>>
>>>> Please ask if more information would help in your review.
>>
>>>>
>>
>>>> Sincerely,
>>
>>>>
>>
>>>>
>>
>>>> Jean Stockard, Ph.D.
>>
>>>> Director of Research
>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>>
>>
>>>>
>>
>>>>> Dear Dr. Stockard,
>>
>>>>>
>>
>>>>> Thank you for contacting the What Works Clearinghouse (WWC).
>>
>>>>>
>>
>>>>> We have received your email about the Reading Mastery Intervention
>>
>>>>> Report.
>>
>>>>> The WWC Quality Review Team is reviewing your email and will
>>

>>>>> prepare a written response. The WWC Quality Review Team responds to
>>
>>>>> concerns raised about WWC reviews published on our website. These
>>
>>>>> quality reviews are undertaken when concerned parties present
>>
>>>>> evidence that a WWC review may be inaccurate. When a quality review
>>
>>>>> is conducted, a researcher who was not involved in the initial
>>
>>>>> review undertakes an independent assessment of the study in
>>>>> question.
>>
>>>>>
>>
>>>>> If a quality review concludes that the original review contained
>>
>>>>> errors, a revised version of the review will be published on the
>>>>> web
>>
>>>>> site. However, it is WWC policy that the existing report will
>>>>> remain
>>
>>>>> on the web site, and be revised only if the WWC team determines it
>>
>>>>> does contain errors. These quality reviews are one of tools used to
>>
>>>>> ensure that the standards established by the Institute of
>>
>>>>> Educational Sciences (IES) are upheld on every review conducted by
>>
>>>>> the What Works Clearinghouse.
>>
>>>>>
>>
>>>>> At this time, we are unable to provide a timeline for a QRT
>>>>> response.
>>
>>>>> Please submit further comments that you would like given to the QRT.
>>
>>>>> You may send them in response to this email.
>>
>>>>>
>>
>>>>> Thank you,
>>
>>>>>
>>

>>>>> What Works Clearinghouse
>>
>>>>>
>>
>>>>> The What Works Clearinghouse was established by the U.S. Department
>>
>>>>> of Education's Institute of Education Sciences to provide
>>>>> educators,
>>
>>>>> policymakers, researchers, and the public with a central and
>>>>> trusted
>>
>>>>> source of scientific evidence of what works in education. For more
>>
>>>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> -----Original Message-----
>>
>>>>>
>>
>>>>> From:
>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@nif
>>>>> di.org%3cmailto:jstockard@nifdi.org>>
>>
>>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]><m
>>>>> ailto:[mailto:jstockard@nifdi.org]>
>>
>>>>>
>>
>>>>> Sent: Tuesday, August 21, 2012 8:54 PM
>>
>>>>>
>>
>>>>> To:
>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@what
>>>>> works.ed.gov%3cmailto:info@whatworks.ed.gov>>
>>
>>>>>

>>
>>>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>>
>>>>> Reports, Reference ID Number: 1757800276
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@what
>>>>> works.ed.gov%3cmailto:info@whatworks.ed.gov>>,
>>>>> this email was
>>
>>>>> automatically sent through the Contact link on the WWC website.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> From:
>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@nif
>>>>> di.org%3cmailto:jstockard@nifdi.org>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Message: I am writing to request a Quality Review of the recent
>>
>>>>> report on Reading Mastery for students with learning disabilities.
>>>>> I
>>
>>>>> believe that the report misinterpreted both of the studies that
>>>>> were
>>
>>>>> reviewed and should be removed from the website immediately.
>>
>>>>>
>>
>>>>>
>>
>>>>>

>>
>>>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>>
>>>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>>
>>>>> Instruction programs, Reading Mastery and Horizons, and found that
>>
>>>>> students in both programs made gains over the academic year that
>>
>>>>> were significantly greater than those made by students in national
>>
>>>>> and state level populations. Gains in both programs were similar,
>>
>>>>> leading the reviewers to conclude that Reading Mastery was no
>>>>> better
>>
>>>>> than its comparison program. They ignored the fact that students in
>>
>>>>> Reading Mastery performed significantly better than the national or
>>
>>>>> state norms and that the comparison program was similar on all but
>>>>> a
>>
>>>>> very few characteristics, essentially a modified version of Reading
>>
>>>>> Mastery. To use a medical analogy, the WWC decision would be
>>
>>>>> equivalent to a researcher claiming that aspirin was ineffective
>>
>>>>> because Tylenol performed equally well.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> The second article reviewed (Herrera, Logan, Cooke, Morris, &
>>>>> Lyman,
>>
>>>>> 1997) involved two groups of students, both of whom appear to have
>>
>>>>> received Reading Mastery as part of the schools' "usual and
>>
>>>>> customary school day curriculum." One group of students also
>>
>>>>> received 45 minutes of supplemental phonemic related instruction,
>>

>>>>> from their regular classroom teachers, that involved motor
>>
>>>>> activities to accompany practice of skills.
>>
>>>>> Not surprisingly, the group receiving the additional instruction
>>>>> had
>>
>>>>> significantly larger gains than those who did not have additional
>>
>>>>> learning time. Despite these differences in exposure and the fact
>>
>>>>> that both groups appear to have had Reading Mastery as their usual
>>
>>>>> reading curriculum, the WWC used these results to suggest that
>>
>>>>> Reading Mastery could have potentially negative effects. A much
>>>>> more
>>
>>>>> accurate conclusion would be that students who receive additional
>>
>>>>> phonics related instruction have greater skill growth.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> I have prepared a more extensive analysis of these concerns, which
>>>>> I
>>
>>>>> would be happy to send you. Because the problems with the analysis
>>
>>>>> are so serious I believe that the report should be immediately
>>>>> taken
>>
>>>>> off your website. In addition, given the very serious
>>
>>>>> misinterpretations of the literature described above I will
>>>>> continue
>>
>>>>> to examine WWC reports and procedures.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>

>>
>>>>> Please acknowledge receipt of my concerns and let me know the time
>>
>>>>> line for your review.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Sincerely,
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Jean Stockard, Ph.D.
>>
>>>>>
>>
>>>>> Director of Research
>>
>>>>>
>>
>>>>> National Institute for Direct Instruction
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>
>>
>>>>
>>
>>>> Jean Stockard, Ph.D.
>>
>>>> Director of Research
>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>>

>>
>>>
>>
>>>
>>
>>> Jean Stockard, Ph.D.
>>
>>> Director of Research
>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>
>>
>>>
>>
>>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>
>
>
>

From: WhatWorks
Sent: Wednesday, November 21, 2012 3:03 PM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear Dr. Stockard,

Unfortunately, we are unable to provide any additional information right now. We hope to be able to provide an update next week, as staff return to the office.

Thank you,

What Works Clearinghouse

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Monday, November 19, 2012 10:47 AM
To: WhatWorks
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear WWC Help Desk,

On November 5 you indicated that I would "be receiving a letter summarizing the findings of the quality review....via email within the next week." (The message is below.) To date (November 19), I have not received such a letter. Because I have been out of the country for a while, it is possible that the communication did not come through. If that is the case, could you please resend it? If it has not yet been sent, can you please let me know when it will come? As you know, I am very concerned about the posting of erroneous and damaging material on your website.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> Thank you for your email. You will be receiving a letter summarizing
> the findings of the quality review. We anticipate sending the letter
> via email within the next week.
>
> Thank you,
>
> What Works Clearinghouse
>
> -----Original Message-----

> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Friday, November 02, 2012 3:09 PM
> To: WhatWorks
> Cc: jstockard@nifdi.org
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear What Works Clearinghouse Help Desk,
>
> I am writing in regards to my request for a quality review of the WWC
> Report on Reading Mastery for Learning Disabled Students (reference
> number is 1757800276), which was submitted on August 21, 2012. I sent
> additional information regarding the request a few weeks later, and
> Dr. Keith Smolkowski, an independent researcher, sent an extensive
> memo regarding the review last week.
>
> The WWC Procedures and Standards Handbook (p. 5) notes that if a
> Quality Review results in a validation of the WWC's original
> assessment the person requesting the review is to receive a response.
> If the original assessment is not validated, reports are made to
> others within the WWC and IES. In this case, there is, however, no
> stated provision for notifying the person who asked for the review.
>
> More than two months have elapsed since my original request for a
> Quality Review. I must assume, given the length of time since my
> request was submitted and the relatively simple job of reviewing the
> two articles involved, that my concerns with the WWC report on Reading
> Mastery and Learning Disabled students have been validated but that,
> per the stated procedures, I have not been informed of this decision.
>
> As in my earlier correspondences, which are appended to this e-mail, I
> ask that you immediately remove the erroneous report from the web and
> clearly indicate that previous postings provided incorrect
> information. In addition to producing possible fiscal damages for the
> authors and publisher of the Reading Mastery program, continued
> dissemination gives very erroneous information to the public.
>
> If I am mistaken in my assumption regarding the outcome of the review,
> please send me the details of your decision.
>
> Sincerely,
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>> Dear Dr. Stockard,

>>
>>
>>
>> The quality review is in process and we will expedite the inclusion
>> of the information on the Herrera, et al. study that was provided
>> yesterday.
>>
>>
>>
>> What Works Clearinghouse
>>
>>
>>
>> -----Original Message-----
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>> Sent: Tuesday, October 02, 2012 9:59 AM
>> To: WhatWorks
>> Cc: jstockard@nifdi.org
>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>> 1757800276
>>
>>
>>
>> Dear WWC Staff,
>>
>>
>>
>> Some recent events have prompted me to follow up on my request for a
>> quality review of WWC Report on Reading Mastery for Learning Disabled
>> Students (reference number is 1757800276), which was submitted on
>> August 21, 2012.
>>
>>
>>
>> First, we received confirmation from an author of one of the articles
>> cited in the review (Dr. Philip Cooker, a co-author of the Herrera,
>> et al.
>>
>> study), that our interpretation of the work, as explained in my
>> earlier correspondence, was correct. A copy of this correspondence is
>> attached below.
>>
>>
>>
>> Second, we asked highly regarded independent researchers to examine
>> our concerns. They have confirmed our interpretation of the problems
>> with the WWC report and note additional problems with WWC procedures
>> and policies.

>>
>>
>>
>> Third, our office has been told that, because of the WWC posting,
>> schools are beginning to abandon the use of Reading Mastery with
>> learning disabled students. In addition to the fiscal damages that
>> will accrue to the authors and publisher from these decisions, such
>> actions would deprive many students of access to a curriculum that
>> has repeatedly been shown to be highly effective.
>>
>>
>>
>> Finally, we were requested to respond to Senator Lisa Murkowski,
>> whose office cited the WWC report in correspondence with schools. We
>> have sent her our explanations of the serious errors in your review.
>>
>>
>>
>> Given the very serious ramifications of your report, I again ask that
>> you immediately remove it from your website and that you provide a
>> clear timeline for the completion of your review including the steps
>> taken to date. If you are not able to do so, we will need to pursue
>> additional avenues to address this situation. Please let me know, as
>> soon as possible, your intentions in this matter.
>>
>>
>>
>> Sincerely,
>>
>>
>>
>> Jean Stockard, Ph.D.
>>
>> Director of Research
>>
>> National Institute for Direct Instruction Eugene, Oregon
>>
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>
>> Below is the correspondence from Dr. Cooker
>>
>>
>>
>> From: "PHILIP G COOKER"
>> <cooker@olemiss.edu<mailto:cooker@olemiss.edu>>
>>

>> Subject: RE: Query on article published in 1997 -
>>
>> Date: Thu, September 6, 2012 9:12 am
>>
>> To: "jstockard@nifdi.org<mailto:jstockard@nifdi.org>"
>> <jstockard@nifdi.org<mailto:jstockard@nifdi.org>>
>>
>>
>>
>> _____
>>
>>
>>
>> Hello,
>>
>>
>>
>> sorry for the late response. I did not conduct the research, but
>> edited and coordinated the efforts. Of the authors, I only knew Dr.
>> Logan. It has been quite a while since the publication and I have
>> little recollection of details. It sounds as if your interpretation
>> is correct. Since reading is not my area of specialization, I do not
>> have knowledge of other research involving the system that was used.
>>
>>
>>
>> Philip G. Cooker, Ph.D.
>>
>> Mississippi Board of Psychology
>>
>> CE Coordinator
>>
>> 407 Timber Lane
>>
>> Oxford, MS 38655
>>
>> (662) 801-1664
>>
>>
>>
>> _____
>>
>> From: Jean Stockard [jstockard@nifdi.org]
>>
>> Sent: Tuesday, September 04, 2012 6:51 PM
>>
>> To: (b)(6) PHILIP G
>> COOKER

>>
>> Subject: Query on article published in 1997 -
>>
>>
>>
>> Dear Professors Herrera and Cooker,
>>
>>
>>
>> This is a note that I sent a few weeks ago. I am afraid it may have
>> gotten lost in your inbox.
>>
>>
>>
>> I am writing for a quick clarification/check on two items in your
>> article, “Phonological awareness and phonetic-graphic conversion: A
>> study of the effects of two intervention paradigms with learning
>> disabled children,”
>>
>> which was published in 1997 in Reading Improvement.
>>
>>
>>
>> First, my understanding is that all students in the study had Reading
>> Mastery as “the usual and customary prescribed curriculum for the
>> school day” (p. 78), and that the intervention group was given the
>> Stabilized Learning System “as a supplement to the usual and
>> customary school day curriculum” (p. 78). In other words, both the
>> control group and the intervention group had Reading Mastery for
>> their usual curriculum, and the intervention group received
>> additional instruction via the Stabilized Learning System. Is this interpretation correct?
>>
>>
>>
>> Second, I’ve not been able to find other writings on the Stabilized
>> Learning System. Do you know if there are other studies of the program?
>>
>>
>>
>> Thank you very much for your help with my queries. If I can provide
>> any clarification, please just ask.
>>
>>
>>
>> Sincerely,
>>
>>
>>

>> Jean Stockard, Ph.D.
>>
>> Director of Research
>>
>> National Institute for Direct Instruction Eugene, Oregon
>>
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>> Dear Dr. Stockard,
>>
>>>
>>
>>> Thank you for your email. As was documented in the table notes in
>>
>>> Appendix
>>
>>> C.1 of the report, we attempted to contact the author (multiple
>>
>>> times), but did not receive a response.
>>
>>>
>>
>>> What Works Clearinghouse
>>
>>>
>>
>>> -----Original Message-----
>>
>>> From: Jean Stockard
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>>> Sent: Tuesday, September 04, 2012 7:12 PM
>>
>>> To: WhatWorks
>>
>>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>>
>>> 1757800276
>>
>>>

>>
>>> Dear WWC Staff:
>>
>>> I am writing with regards to the article by Herrera and associates
>>
>>> (1997) that was included in your recent review of Reading Mastery
>>> for
>>
>>> Learning Disabled students. I wondered if you had any additional
>>
>>> information on the Herrera et al. study that was obtained from the
>>
>>> authors in the course of developing your report. If so, could I
>>
>>> possibly have access to that information? I have been trying to
>>
>>> contact them, but have not been successful.
>>
>>>
>>
>>> Thank you very much for your quick responses to queries.
>>
>>>
>>
>>> Sincerely,
>>
>>>
>>
>>> Jean Stockard, Ph.D.
>>
>>> Director of Research
>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>
>>
>>>> Dear Dr. Stockard,
>>
>>>>
>>
>>>> We received your email and attachment. We provided it to the QRT.
>>
>>>>
>>
>>>> Thank you,
>>

>>>>
>>
>>>> What Works Clearinghouse
>>
>>>>
>>
>>>> -----Original Message-----
>>
>>>> From: Jean Stockard
>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>>>> Sent: Thursday, August 23, 2012 2:41 PM
>>
>>>> To: WhatWorks
>>
>>>> Cc: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>
>>>> Subject: Re: What Works Clearinghouse (WWC 3801)
>>
>>>>
>>
>>>> Thank you for acknowledging receipt of my e-mail. I have attached a
>>
>>>> more lengthy analysis of both the review and related literature.
>>
>>>> Please ask if more information would help in your review.
>>
>>>>
>>
>>>> Sincerely,
>>
>>>>
>>
>>>>
>>
>>>> Jean Stockard, Ph.D.
>>
>>>> Director of Research
>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>>
>>
>>>>
>>
>>>>>> Dear Dr. Stockard,

>>
>>>>>
>>
>>>>> Thank you for contacting the What Works Clearinghouse (WWC).
>>
>>>>>
>>
>>>>> We have received your email about the Reading Mastery Intervention
>>
>>>>> Report.
>>
>>>>> The WWC Quality Review Team is reviewing your email and will
>>
>>>>> prepare a written response. The WWC Quality Review Team responds
>>>>> to
>>
>>>>> concerns raised about WWC reviews published on our website. These
>>
>>>>> quality reviews are undertaken when concerned parties present
>>
>>>>> evidence that a WWC review may be inaccurate. When a quality
>>>>> review
>>
>>>>> is conducted, a researcher who was not involved in the initial
>>
>>>>> review undertakes an independent assessment of the study in
>>>>> question.
>>
>>>>>
>>
>>>>> If a quality review concludes that the original review contained
>>
>>>>> errors, a revised version of the review will be published on the
>>>>> web
>>
>>>>> site. However, it is WWC policy that the existing report will
>>>>> remain
>>
>>>>> on the web site, and be revised only if the WWC team determines it
>>
>>>>> does contain errors. These quality reviews are one of tools used
>>>>> to
>>
>>>>> ensure that the standards established by the Institute of
>>
>>>>> Educational Sciences (IES) are upheld on every review conducted by
>>
>>>>> the What Works Clearinghouse.

>>
>>>>>
>>
>>>>> At this time, we are unable to provide a timeline for a QRT
>>>>> response.
>>
>>>>> Please submit further comments that you would like given to the QRT.
>>
>>>>> You may send them in response to this email.
>>
>>>>>
>>
>>>>> Thank you,
>>
>>>>>
>>
>>>>> What Works Clearinghouse
>>
>>>>>
>>
>>>>> The What Works Clearinghouse was established by the U.S.
>>>>> Department
>>
>>>>> of Education's Institute of Education Sciences to provide
>>>>> educators,
>>
>>>>> policymakers, researchers, and the public with a central and
>>>>> trusted
>>
>>>>> source of scientific evidence of what works in education. For more
>>
>>>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> -----Original Message-----
>>
>>>>>
>>
>>>>> From:

>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@ni
>>>>> f di.org%3cmailto:jstockard@nifdi.org>>
>>
>>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]><
>>>>> m ailto:[mailto:jstockard@nifdi.org]>
>>
>>>>>
>>
>>>>> Sent: Tuesday, August 21, 2012 8:54 PM
>>
>>>>>
>>
>>>>> To:
>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@wha
>>>>> t works.ed.gov%3cmailto:info@whatworks.ed.gov>>
>>
>>>>>
>>
>>>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>>
>>>>> Reports, Reference ID Number: 1757800276
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@wha
>>>>> t works.ed.gov%3cmailto:info@whatworks.ed.gov>>,
>>>>> this email was
>>
>>>>> automatically sent through the Contact link on the WWC website.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> From:
>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@ni
>>>>> f di.org%3cmailto:jstockard@nifdi.org>>
>>
>>>>>
>>
>>>>>
>>

>>>>>
>>
>>>>> Message: I am writing to request a Quality Review of the recent
>>
>>>>> report on Reading Mastery for students with learning disabilities.
>>>>> I
>>
>>>>> believe that the report misinterpreted both of the studies that
>>>>> were
>>
>>>>> reviewed and should be removed from the website immediately.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>>
>>>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>>
>>>>> Instruction programs, Reading Mastery and Horizons, and found that
>>
>>>>> students in both programs made gains over the academic year that
>>
>>>>> were significantly greater than those made by students in national
>>
>>>>> and state level populations. Gains in both programs were similar,
>>
>>>>> leading the reviewers to conclude that Reading Mastery was no
>>>>> better
>>
>>>>> than its comparison program. They ignored the fact that students
>>>>> in
>>
>>>>> Reading Mastery performed significantly better than the national
>>>>> or
>>
>>>>> state norms and that the comparison program was similar on all but
>>>>> a
>>
>>>>> very few characteristics, essentially a modified version of
>>>>> Reading
>>
>>>>> Mastery. To use a medical analogy, the WWC decision would be
>>
>>>>> equivalent to a researcher claiming that aspirin was ineffective

>>
>>>>> because Tylenol performed equally well.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> The second article reviewed (Herrera, Logan, Cooke, Morris, &
>>>>> Lyman,
>>
>>>>> 1997) involved two groups of students, both of whom appear to have
>>
>>>>> received Reading Mastery as part of the schools' "usual and
>>
>>>>> customary school day curriculum." One group of students also
>>
>>>>> received 45 minutes of supplemental phonemic related instruction,
>>
>>>>> from their regular classroom teachers, that involved motor
>>
>>>>> activities to accompany practice of skills.
>>
>>>>> Not surprisingly, the group receiving the additional instruction
>>>>> had
>>
>>>>> significantly larger gains than those who did not have additional
>>
>>>>> learning time. Despite these differences in exposure and the fact
>>
>>>>> that both groups appear to have had Reading Mastery as their usual
>>
>>>>> reading curriculum, the WWC used these results to suggest that
>>
>>>>> Reading Mastery could have potentially negative effects. A much
>>>>> more
>>
>>>>> accurate conclusion would be that students who receive additional
>>
>>>>> phonics related instruction have greater skill growth.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>

>>>>> I have prepared a more extensive analysis of these concerns, which
>>>>> I
>>
>>>>> would be happy to send you. Because the problems with the analysis
>>
>>>>> are so serious I believe that the report should be immediately
>>>>> taken
>>
>>>>> off your website. In addition, given the very serious
>>
>>>>> misinterpretations of the literature described above I will
>>>>> continue
>>
>>>>> to examine WWC reports and procedures.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Please acknowledge receipt of my concerns and let me know the time
>>
>>>>> line for your review.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Sincerely,
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Jean Stockard, Ph.D.
>>
>>>>>
>>
>>>>> Director of Research
>>
>>>>>
>>
>>>>> National Institute for Direct Instruction

>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>
>>
>>>>
>>
>>>> Jean Stockard, Ph.D.
>>
>>>> Director of Research
>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>>
>>
>>>
>>
>>>
>>
>>> Jean Stockard, Ph.D.
>>
>>> Director of Research
>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>
>>
>>>
>>
>>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>

>>
>
>
>
>

From: WhatWorks
Sent: Friday, November 30, 2012 10:35 AM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear Dr. Stockard,

The issue is in the final approval stages. The response will be provided soon.

Thank you,

What Works Clearinghouse

-----Original Message-----
From: WhatWorks
Sent: Wednesday, November 21, 2012 2:03 PM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear Dr. Stockard,

Unfortunately, we are unable to provide any additional information right now. We hope to be able to provide an update next week, as staff return to the office.

Thank you,

What Works Clearinghouse

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Monday, November 19, 2012 10:47 AM
To: WhatWorks
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear WWC Help Desk,

On November 5 you indicated that I would "be receiving a letter summarizing the findings of the quality review....via email within the next week." (The message is below.) To date (November 19), I have not received such a letter. Because I have been out of the country for a while, it is possible that the communication did not come through. If that is the case, could you please resend it? If it has not yet been sent, can you please let me know when it will come? As you know, I am very concerned about the posting of erroneous and damaging material on your website.

Sincerely,

Jean Stockard, Ph.D.
Director of Research

National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> Thank you for your email. You will be receiving a letter summarizing
> the findings of the quality review. We anticipate sending the letter
> via email within the next week.
>
> Thank you,
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Friday, November 02, 2012 3:09 PM
> To: WhatWorks
> Cc: jstockard@nifdi.org
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear What Works Clearinghouse Help Desk,
>
> I am writing in regards to my request for a quality review of the WWC
> Report on Reading Mastery for Learning Disabled Students (reference
> number is 1757800276), which was submitted on August 21, 2012. I sent
> additional information regarding the request a few weeks later, and
> Dr. Keith Smolkowski, an independent researcher, sent an extensive
> memo regarding the review last week.
>
> The WWC Procedures and Standards Handbook (p. 5) notes that if a
> Quality Review results in a validation of the WWC's original
> assessment the person requesting the review is to receive a response.
> If the original assessment is not validated, reports are made to
> others within the WWC and IES. In this case, there is, however, no
> stated provision for notifying the person who asked for the review.
>
> More than two months have elapsed since my original request for a
> Quality Review. I must assume, given the length of time since my
> request was submitted and the relatively simple job of reviewing the
> two articles involved, that my concerns with the WWC report on Reading
> Mastery and Learning Disabled students have been validated but that,
> per the stated procedures, I have not been informed of this decision.
>
> As in my earlier correspondences, which are appended to this e-mail, I
> ask that you immediately remove the erroneous report from the web and
> clearly indicate that previous postings provided incorrect
> information. In addition to producing possible fiscal damages for the

> authors and publisher of the Reading Mastery program, continued
> dissemination gives very erroneous information to the public.
>
> If I am mistaken in my assumption regarding the outcome of the review,
> please send me the details of your decision.
>
> Sincerely,
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>> Dear Dr. Stockard,
>>
>>
>>
>> The quality review is in process and we will expedite the inclusion
>> of the information on the Herrera, et al. study that was provided
>> yesterday.
>>
>>
>>
>> What Works Clearinghouse
>>
>>
>>
>> -----Original Message-----
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>> Sent: Tuesday, October 02, 2012 9:59 AM
>> To: WhatWorks
>> Cc: jstockard@nifdi.org
>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>> 1757800276
>>
>>
>>
>> Dear WWC Staff,
>>
>>
>>
>> Some recent events have prompted me to follow up on my request for a
>> quality review of WWC Report on Reading Mastery for Learning Disabled
>> Students (reference number is 1757800276), which was submitted on
>> August 21, 2012.
>>
>>
>>

>> First, we received confirmation from an author of one of the articles
>> cited in the review (Dr. Philip Cooker, a co-author of the Herrera,
>> et al.
>>
>> study), that our interpretation of the work, as explained in my
>> earlier correspondence, was correct. A copy of this correspondence is
>> attached below.
>>
>>
>>
>> Second, we asked highly regarded independent researchers to examine
>> our concerns. They have confirmed our interpretation of the problems
>> with the WWC report and note additional problems with WWC procedures
>> and policies.
>>
>>
>>
>> Third, our office has been told that, because of the WWC posting,
>> schools are beginning to abandon the use of Reading Mastery with
>> learning disabled students. In addition to the fiscal damages that
>> will accrue to the authors and publisher from these decisions, such
>> actions would deprive many students of access to a curriculum that
>> has repeatedly been shown to be highly effective.
>>
>>
>>
>> Finally, we were requested to respond to Senator Lisa Murkowski,
>> whose office cited the WWC report in correspondence with schools. We
>> have sent her our explanations of the serious errors in your review.
>>
>>
>>
>> Given the very serious ramifications of your report, I again ask that
>> you immediately remove it from your website and that you provide a
>> clear timeline for the completion of your review including the steps
>> taken to date. If you are not able to do so, we will need to pursue
>> additional avenues to address this situation. Please let me know, as
>> soon as possible, your intentions in this matter.
>>
>>
>>
>> Sincerely,
>>
>>
>>
>> Jean Stockard, Ph.D.
>>
>> Director of Research

>>
>> National Institute for Direct Instruction Eugene, Oregon
>>
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>
>> Below is the correspondence from Dr. Cooker
>>
>>
>>
>> From: "PHILIP G COOKER"
>> <cooker@olemiss.edu<mailto:cooker@olemiss.edu>>
>>
>> Subject: RE: Query on article published in 1997 -
>>
>> Date: Thu, September 6, 2012 9:12 am
>>
>> To: "jstockard@nifdi.org<mailto:jstockard@nifdi.org>"
>> <jstockard@nifdi.org<mailto:jstockard@nifdi.org>>
>>
>>
>>
>> _____
>>
>>
>>
>> Hello,
>>
>>
>>
>> sorry for the late response. I did not conduct the research, but
>> edited and coordinated the efforts. Of the authors, I only knew Dr.
>> Logan. It has been quite a while since the publication and I have
>> little recollection of details. It sounds as if your interpretation
>> is correct. Since reading is not my area of specialization, I do not
>> have knowledge of other research involving the system that was used.
>>
>>
>>
>> Philip G. Cooker, Ph.D.
>>
>> Mississippi Board of Psychology
>>
>> CE Coordinator
>>
>> 407 Timber Lane
>>

>> Oxford, MS 38655
>>
>> (662) 801-1664
>>
>>
>>
>> _____
>>
>> From: Jean Stockard [jstockard@nifdi.org]
>>
>> Sent: Tuesday, September 04, 2012 6:51 PM
>>
>> To: [redacted] PHILIP G
>> COOKER
>>
>> Subject: Query on article published in 1997 -
>>
>>
>>
>> Dear Professors Herrera and Cooker,
>>
>>
>>
>> This is a note that I sent a few weeks ago. I am afraid it may have
>> gotten lost in your inbox.
>>
>>
>>
>> I am writing for a quick clarification/check on two items in your
>> article, “Phonological awareness and phonetic-graphic conversion: A
>> study of the effects of two intervention paradigms with learning
>> disabled children,”
>>
>> which was published in 1997 in Reading Improvement.
>>
>>
>>
>> First, my understanding is that all students in the study had Reading
>> Mastery as “the usual and customary prescribed curriculum for the
>> school day” (p. 78), and that the intervention group was given the
>> Stabilized Learning System “as a supplement to the usual and
>> customary school day curriculum” (p. 78). In other words, both the
>> control group and the intervention group had Reading Mastery for
>> their usual curriculum, and the intervention group received
>> additional instruction via the Stabilized Learning System. Is this interpretation correct?
>>
>>
>>

>> Second, I've not been able to find other writings on the Stabilized
>> Learning System. Do you know if there are other studies of the program?
>>
>>
>>
>> Thank you very much for your help with my queries. If I can provide
>> any clarification, please just ask.
>>
>>
>>
>> Sincerely,
>>
>>
>>
>> Jean Stockard, Ph.D.
>>
>> Director of Research
>>
>> National Institute for Direct Instruction Eugene, Oregon
>>
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>> Dear Dr. Stockard,
>>
>>>
>>
>>> Thank you for your email. As was documented in the table notes in
>>
>>> Appendix
>>
>>> C.1 of the report, we attempted to contact the author (multiple
>>
>>> times), but did not receive a response.
>>
>>>
>>
>>> What Works Clearinghouse
>>
>>>
>>

>>> -----Original Message-----
>>
>>> From: Jean Stockard
>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>>> Sent: Tuesday, September 04, 2012 7:12 PM
>>
>>> To: WhatWorks
>>
>>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>>
>>> 1757800276
>>
>>>
>>
>>> Dear WWC Staff:
>>
>>> I am writing with regards to the article by Herrera and associates
>>
>>> (1997) that was included in your recent review of Reading Mastery
>>> for
>>
>>> Learning Disabled students. I wondered if you had any additional
>>
>>> information on the Herrera et al. study that was obtained from the
>>
>>> authors in the course of developing your report. If so, could I
>>
>>> possibly have access to that information? I have been trying to
>>
>>> contact them, but have not been successful.
>>
>>>
>>
>>> Thank you very much for your quick responses to queries.
>>
>>>
>>
>>> Sincerely,
>>
>>>
>>
>>> Jean Stockard, Ph.D.
>>
>>> Director of Research
>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>

>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>
>>
>>>> Dear Dr. Stockard,
>>
>>>>
>>
>>>> We received your email and attachment. We provided it to the QRT.
>>
>>>>
>>
>>>> Thank you,
>>
>>>>
>>
>>>> What Works Clearinghouse
>>
>>>>
>>
>>>> -----Original Message-----
>>
>>>> From: Jean Stockard
>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>
>>>> Sent: Thursday, August 23, 2012 2:41 PM
>>
>>>> To: WhatWorks
>>
>>>> Cc: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>
>>>> Subject: Re: What Works Clearinghouse (WWC 3801)
>>
>>>>
>>
>>>> Thank you for acknowledging receipt of my e-mail. I have attached a
>>
>>>> more lengthy analysis of both the review and related literature.
>>
>>>> Please ask if more information would help in your review.
>>
>>>>
>>
>>>> Sincerely,
>>
>>>>
>>
>>>>

>>
>>>> Jean Stockard, Ph.D.
>>
>>>> Director of Research
>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>>
>>
>>>>
>>
>>>>> Dear Dr. Stockard,
>>
>>>>>
>>
>>>>> Thank you for contacting the What Works Clearinghouse (WWC).
>>
>>>>>
>>
>>>>> We have received your email about the Reading Mastery Intervention
>>
>>>>> Report.
>>
>>>>> The WWC Quality Review Team is reviewing your email and will
>>
>>>>> prepare a written response. The WWC Quality Review Team responds
>>>>> to
>>
>>>>> concerns raised about WWC reviews published on our website. These
>>
>>>>> quality reviews are undertaken when concerned parties present
>>
>>>>> evidence that a WWC review may be inaccurate. When a quality
>>>>> review
>>
>>>>> is conducted, a researcher who was not involved in the initial
>>
>>>>> review undertakes an independent assessment of the study in
>>>>> question.
>>
>>>>>
>>
>>>>> If a quality review concludes that the original review contained
>>
>>>>> errors, a revised version of the review will be published on the
>>>>> web

>>
>>>>> site. However, it is WWC policy that the existing report will
>>>>> remain
>>
>>>>> on the web site, and be revised only if the WWC team determines it
>>
>>>>> does contain errors. These quality reviews are one of tools used
>>>>> to
>>
>>>>> ensure that the standards established by the Institute of
>>
>>>>> Educational Sciences (IES) are upheld on every review conducted by
>>
>>>>> the What Works Clearinghouse.
>>
>>>>>
>>
>>>>> At this time, we are unable to provide a timeline for a QRT
>>>>> response.
>>
>>>>> Please submit further comments that you would like given to the QRT.
>>
>>>>> You may send them in response to this email.
>>
>>>>>
>>
>>>>> Thank you,
>>
>>>>>
>>
>>>>> What Works Clearinghouse
>>
>>>>>
>>
>>>>> The What Works Clearinghouse was established by the U.S.
>>>>> Department
>>
>>>>> of Education's Institute of Education Sciences to provide
>>>>> educators,
>>
>>>>> policymakers, researchers, and the public with a central and
>>>>> trusted
>>
>>>>> source of scientific evidence of what works in education. For more
>>
>>>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>
>>>>>

>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> -----Original Message-----
>>
>>>>>
>>
>>>>> From:
>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@ni
>>>>> f di.org%3cmailto:jstockard@nifdi.org>>
>>
>>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]><
>>>>> m ailto:[mailto:jstockard@nifdi.org]>
>>
>>>>>
>>
>>>>> Sent: Tuesday, August 21, 2012 8:54 PM
>>
>>>>>
>>
>>>>> To:
>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@wha
>>>>> t works.ed.gov%3cmailto:info@whatworks.ed.gov>>
>>
>>>>>
>>
>>>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>>
>>>>> Reports, Reference ID Number: 1757800276
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@wha
>>>>> t works.ed.gov%3cmailto:info@whatworks.ed.gov>>,
>>>>> this email was
>>
>>>>> automatically sent through the Contact link on the WWC website.
>>

>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> From:
>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@ni
>>>>> f di.org%3cmailto:jstockard@nifdi.org>>
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Message: I am writing to request a Quality Review of the recent
>>
>>>>> report on Reading Mastery for students with learning disabilities.
>>>>> I
>>
>>>>> believe that the report misinterpreted both of the studies that
>>>>> were
>>
>>>>> reviewed and should be removed from the website immediately.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>>
>>>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>>
>>>>> Instruction programs, Reading Mastery and Horizons, and found that
>>
>>>>> students in both programs made gains over the academic year that
>>
>>>>> were significantly greater than those made by students in national
>>
>>>>> and state level populations. Gains in both programs were similar,
>>
>>>>> leading the reviewers to conclude that Reading Mastery was no
>>>>> better
>>
>>>>> than its comparison program. They ignored the fact that students

>>>>> in
>>
>>>>> Reading Mastery performed significantly better than the national
>>>>> or
>>
>>>>> state norms and that the comparison program was similar on all but
>>>>> a
>>
>>>>> very few characteristics, essentially a modified version of
>>>>> Reading
>>
>>>>> Mastery. To use a medical analogy, the WWC decision would be
>>
>>>>> equivalent to a researcher claiming that aspirin was ineffective
>>
>>>>> because Tylenol performed equally well.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> The second article reviewed (Herrera, Logan, Cooke, Morris, &
>>>>> Lyman,
>>
>>>>> 1997) involved two groups of students, both of whom appear to have
>>
>>>>> received Reading Mastery as part of the schools' "usual and
>>
>>>>> customary school day curriculum." One group of students also
>>
>>>>> received 45 minutes of supplemental phonemic related instruction,
>>
>>>>> from their regular classroom teachers, that involved motor
>>
>>>>> activities to accompany practice of skills.
>>
>>>>> Not surprisingly, the group receiving the additional instruction
>>>>> had
>>
>>>>> significantly larger gains than those who did not have additional
>>
>>>>> learning time. Despite these differences in exposure and the fact
>>
>>>>> that both groups appear to have had Reading Mastery as their usual
>>
>>>>> reading curriculum, the WWC used these results to suggest that

>>
>>>>> Reading Mastery could have potentially negative effects. A much
>>>>> more
>>
>>>>> accurate conclusion would be that students who receive additional
>>
>>>>> phonics related instruction have greater skill growth.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> I have prepared a more extensive analysis of these concerns, which
>>>>> I
>>
>>>>> would be happy to send you. Because the problems with the analysis
>>
>>>>> are so serious I believe that the report should be immediately
>>>>> taken
>>
>>>>> off your website. In addition, given the very serious
>>
>>>>> misinterpretations of the literature described above I will
>>>>> continue
>>
>>>>> to examine WWC reports and procedures.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Please acknowledge receipt of my concerns and let me know the time
>>
>>>>> line for your review.
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>> Sincerely,
>>
>>>>>

>>
>>>>>
>>
>>>>>
>>
>>>>> Jean Stockard, Ph.D.
>>
>>>>>
>>
>>>>> Director of Research
>>
>>>>>
>>
>>>>> National Institute for Direct Instruction
>>
>>>>>
>>
>>>>>
>>
>>>>>
>>
>>>>
>>
>>>>
>>
>>>> Jean Stockard, Ph.D.
>>
>>>> Director of Research
>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>>
>>
>>>
>>
>>>
>>
>>> Jean Stockard, Ph.D.
>>
>>> Director of Research
>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>>

>>
>>>
>>
>>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>
>
>
>

From: Jean Stockard <jstockard@nifdi.org>
Sent: Friday, December 07, 2012 11:13 AM
To: WhatWorks
Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number: 1757800276

Dear WWC Help Desk.

I am again writing regarding the quality review I requested of the WWC Report on Reading Mastery for students with Learning Disabilities. The request was given the reference number 1757800276.

As you and your staff no doubt know, it has been more than a month since I was promised "a letter summarizing the findings of the quality review…via email within the next week." Although it appears that the erroneous report has been removed from your website, it is still appearing in ERIC. The numerous blog postings about the report have also not been corrected.

I have been waiting to contact the blog authors and ERIC about your apparent decision to remove the report until I received your explanatory letter. I have done so out of professional courtesy, hoping that your report would indicate a decision to adhere to common scientific standards of research and federal legal requirements regarding transparency of actions and policies. However, your review team's delayed actions have led me to rethink this approach. If I have not been given the explanatory letter by next Friday (December 14), I will ask our staff to send communications to all who publicized the report, telling them of the removal from the website and the reasons that I believe promoted this action.

Thank you for your continued attention to this matter.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> The issue is in the final approval stages. The response will be provided
> soon.
>
> Thank you,
>
> What Works Clearinghouse

>
> -----Original Message-----
> From: WhatWorks
> Sent: Wednesday, November 21, 2012 2:03 PM
> To: 'jstockard@nifdi.org'
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear Dr. Stockard,
>
> Unfortunately, we are unable to provide any additional information right
> now. We hope to be able to provide an update next week, as staff return to
> the office.
>
> Thank you,
>
> What Works Clearinghouse
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Monday, November 19, 2012 10:47 AM
> To: WhatWorks
> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
> 1757800276
>
> Dear WWC Help Desk,
>
> On November 5 you indicated that I would “be receiving a letter
> summarizing the findings of the quality review....via email within the
> next week.” (The message is below.) To date (November 19), I have not
> received such a letter. Because I have been out of the country for a
> while, it is possible that the communication did not come through. If that
> is the case, could you please resend it? If it has not yet been sent, can
> you please let me know when it will come? As you know, I am very concerned
> about the posting of erroneous and damaging material on your website.
>
> Sincerely,
>
> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction Eugene, Oregon
> phone: 541-505-5710 or toll-free 877-485-1973
>
>> Dear Dr. Stockard,
>>
>> Thank you for your email. You will be receiving a letter summarizing
>> the findings of the quality review. We anticipate sending the letter
>> via email within the next week.

>>
>> Thank you,
>>
>> What Works Clearinghouse
>>
>> -----Original Message-----
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>> Sent: Friday, November 02, 2012 3:09 PM
>> To: WhatWorks
>> Cc: jstockard@nifdi.org
>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>> 1757800276
>>
>> Dear What Works Clearinghouse Help Desk,
>>
>> I am writing in regards to my request for a quality review of the WWC
>> Report on Reading Mastery for Learning Disabled Students (reference
>> number is 1757800276), which was submitted on August 21, 2012. I sent
>> additional information regarding the request a few weeks later, and
>> Dr. Keith Smolkowski, an independent researcher, sent an extensive
>> memo regarding the review last week.
>>
>> The WWC Procedures and Standards Handbook (p. 5) notes that if a
>> Quality Review results in a validation of the WWC's original
>> assessment the person requesting the review is to receive a response.
>> If the original assessment is not validated, reports are made to
>> others within the WWC and IES. In this case, there is, however, no
>> stated provision for notifying the person who asked for the review.
>>
>> More than two months have elapsed since my original request for a
>> Quality Review. I must assume, given the length of time since my
>> request was submitted and the relatively simple job of reviewing the
>> two articles involved, that my concerns with the WWC report on Reading
>> Mastery and Learning Disabled students have been validated but that,
>> per the stated procedures, I have not been informed of this decision.
>>
>> As in my earlier correspondences, which are appended to this e-mail, I
>> ask that you immediately remove the erroneous report from the web and
>> clearly indicate that previous postings provided incorrect
>> information. In addition to producing possible fiscal damages for the
>> authors and publisher of the Reading Mastery program, continued
>> dissemination gives very erroneous information to the public.
>>
>> If I am mistaken in my assumption regarding the outcome of the review,
>> please send me the details of your decision.
>>
>> Sincerely,
>>

>> Jean Stockard, Ph.D.
>> Director of Research
>> National Institute for Direct Instruction Eugene, Oregon
>> phone: 541-505-5710 or toll-free 877-485-1973
>>
>>> Dear Dr. Stockard,
>>>
>>>
>>>
>>> The quality review is in process and we will expedite the inclusion
>>> of the information on the Herrera, et al. study that was provided
>>> yesterday.
>>>
>>>
>>>
>>> What Works Clearinghouse
>>>
>>>
>>>
>>> -----Original Message-----
>>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>>> Sent: Tuesday, October 02, 2012 9:59 AM
>>> To: WhatWorks
>>> Cc: jstockard@nifdi.org
>>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID Number:
>>> 1757800276
>>>
>>>
>>>
>>> Dear WWC Staff,
>>>
>>>
>>>
>>> Some recent events have prompted me to follow up on my request for a
>>> quality review of WWC Report on Reading Mastery for Learning Disabled
>>> Students (reference number is 1757800276), which was submitted on
>>> August 21, 2012.
>>>
>>>
>>>
>>> First, we received confirmation from an author of one of the articles
>>> cited in the review (Dr. Philip Cooker, a co-author of the Herrera,
>>> et al.
>>>
>>> study), that our interpretation of the work, as explained in my
>>> earlier correspondence, was correct. A copy of this correspondence is
>>> attached below.
>>>

>>>
>>>
>>> Second, we asked highly regarded independent researchers to examine
>>> our concerns. They have confirmed our interpretation of the problems
>>> with the WWC report and note additional problems with WWC procedures
>>> and policies.
>>>
>>>
>>>
>>> Third, our office has been told that, because of the WWC posting,
>>> schools are beginning to abandon the use of Reading Mastery with
>>> learning disabled students. In addition to the fiscal damages that
>>> will accrue to the authors and publisher from these decisions, such
>>> actions would deprive many students of access to a curriculum that
>>> has repeatedly been shown to be highly effective.
>>>
>>>
>>>
>>> Finally, we were requested to respond to Senator Lisa Murkowski,
>>> whose office cited the WWC report in correspondence with schools. We
>>> have sent her our explanations of the serious errors in your review.
>>>
>>>
>>>
>>> Given the very serious ramifications of your report, I again ask that
>>> you immediately remove it from your website and that you provide a
>>> clear timeline for the completion of your review including the steps
>>> taken to date. If you are not able to do so, we will need to pursue
>>> additional avenues to address this situation. Please let me know, as
>>> soon as possible, your intentions in this matter.
>>>
>>>
>>>
>>> Sincerely,
>>>
>>>
>>>
>>> Jean Stockard, Ph.D.
>>>
>>> Director of Research
>>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>>>
>>>
>>>
>>> Below is the correspondence from Dr. Cooker

>>>
>>>
>>>
>>> From: "PHILIP G COOKER"
>>> <cooker@olemiss.edu<mailto:cooker@olemiss.edu>>
>>>
>>> Subject: RE: Query on article published in 1997 -
>>>
>>> Date: Thu, September 6, 2012 9:12 am
>>>
>>> To: "jstockard@nifdi.org<mailto:jstockard@nifdi.org>"
>>> <jstockard@nifdi.org<mailto:jstockard@nifdi.org>>
>>>
>>>
>>>
>>> _____
>>>
>>>
>>>
>>> Hello,
>>>
>>>
>>>
>>> sorry for the late response. I did not conduct the research, but
>>> edited and coordinated the efforts. Of the authors, I only knew Dr.
>>> Logan. It has been quite a while since the publication and I have
>>> little recollection of details. It sounds as if your interpretation
>>> is correct. Since reading is not my area of specialization, I do not
>>> have knowledge of other research involving the system that was used.
>>>
>>>
>>>
>>> Philip G. Cooker, Ph.D.
>>>
>>> Mississippi Board of Psychology
>>>
>>> CE Coordinator
>>>
>>> 407 Timber Lane
>>>
>>> Oxford, MS 38655
>>>
>>> (662) 801-1664
>>>
>>>
>>>
>>> _____
>>>

>>> From: Jean Stockard [jstockard@nifdi.org]
>>>
>>> Sent: Tuesday, September 04, 2012 6:51 PM
>>>
>>> To: (b)(6) ; PHILIP G
>>> COOKER
>>>
>>> Subject: Query on article published in 1997 -
>>>
>>>
>>>
>>> Dear Professors Herrera and Cooker,
>>>
>>>
>>>
>>> This is a note that I sent a few weeks ago. I am afraid it may have
>>> gotten lost in your inbox.
>>>
>>>
>>>
>>> I am writing for a quick clarification/check on two items in your
>>> article, “Phonological awareness and phonetic-graphic conversion: A
>>> study of the effects of two intervention paradigms with learning
>>> disabled children,”
>>>
>>> which was published in 1997 in Reading Improvement.
>>>
>>>
>>>
>>> First, my understanding is that all students in the study had Reading
>>> Mastery as “the usual and customary prescribed curriculum for the
>>> school day” (p. 78), and that the intervention group was given the
>>> Stabilized Learning System “as a supplement to the usual and
>>> customary school day curriculum” (p. 78). In other words, both the
>>> control group and the intervention group had Reading Mastery for
>>> their usual curriculum, and the intervention group received
>>> additional instruction via the Stabilized Learning System. Is this
>>> interpretation correct?
>>>
>>>
>>>
>>> Second, I’ve not been able to find other writings on the Stabilized
>>> Learning System. Do you know if there are other studies of the
>>> program?
>>>
>>>
>>>
>>> Thank you very much for your help with my queries. If I can provide

>>> any clarification, please just ask.
>>>
>>>
>>>
>>> Sincerely,
>>>
>>>
>>>
>>> Jean Stockard, Ph.D.
>>>
>>> Director of Research
>>>
>>> National Institute for Direct Instruction Eugene, Oregon
>>>
>>> phone: 541-505-5710 or toll-free 877-485-1973
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>> Dear Dr. Stockard,
>>>
>>>>
>>>
>>>> Thank you for your email. As was documented in the table notes in
>>>
>>>> Appendix
>>>
>>>> C.1 of the report, we attempted to contact the author (multiple
>>>
>>>> times), but did not receive a response.
>>>
>>>>
>>>
>>>> What Works Clearinghouse
>>>
>>>>
>>>
>>>> -----Original Message-----
>>>
>>>> From: Jean Stockard
>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>>
>>>> Sent: Tuesday, September 04, 2012 7:12 PM

>>>
>>>> To: WhatWorks
>>>
>>>> Subject: RE: What Works Clearinghouse (WWC 3801), Reference ID
>>>> Number:
>>>
>>>> 1757800276
>>>
>>>>
>>>
>>>> Dear WWC Staff:
>>>
>>>> I am writing with regards to the article by Herrera and associates
>>>
>>>> (1997) that was included in your recent review of Reading Mastery
>>>> for
>>>
>>>> Learning Disabled students. I wondered if you had any additional
>>>
>>>> information on the Herrera et al. study that was obtained from the
>>>
>>>> authors in the course of developing your report. If so, could I
>>>
>>>> possibly have access to that information? I have been trying to
>>>
>>>> contact them, but have not been successful.
>>>
>>>>
>>>
>>>> Thank you very much for your quick responses to queries.
>>>
>>>>
>>>
>>>> Sincerely,
>>>
>>>>
>>>
>>>> Jean Stockard, Ph.D.
>>>
>>>> Director of Research
>>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>>
>>>>
>>>
>>>>> Dear Dr. Stockard,

>>>
>>>>>
>>>
>>>>> We received your email and attachment. We provided it to the QRT.
>>>
>>>>>
>>>
>>>>> Thank you,
>>>
>>>>>
>>>
>>>>> What Works Clearinghouse
>>>
>>>>>
>>>
>>>>> -----Original Message-----
>>>
>>>>> From: Jean Stockard
>>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]>
>>>
>>>>> Sent: Thursday, August 23, 2012 2:41 PM
>>>
>>>>> To: WhatWorks
>>>
>>>>> Cc: jstockard@nifdi.org<mailto:jstockard@nifdi.org>
>>>
>>>>> Subject: Re: What Works Clearinghouse (WWC 3801)
>>>
>>>>>
>>>
>>>>> Thank you for acknowledging receipt of my e-mail. I have attached a
>>>
>>>>> more lengthy analysis of both the review and related literature.
>>>
>>>>> Please ask if more information would help in your review.
>>>
>>>>>
>>>
>>>>> Sincerely,
>>>
>>>>>
>>>
>>>>>
>>>
>>>>> Jean Stockard, Ph.D.
>>>
>>>>> Director of Research
>>>

>>>>> National Institute for Direct Instruction Eugene, Oregon
>>>
>>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>>
>>>>>
>>>
>>>>>
>>>
>>>>>> Dear Dr. Stockard,
>>>
>>>>>>
>>>
>>>>>> Thank you for contacting the What Works Clearinghouse (WWC).
>>>
>>>>>>
>>>
>>>>>> We have received your email about the Reading Mastery Intervention
>>>
>>>>>> Report.
>>>
>>>>>> The WWC Quality Review Team is reviewing your email and will
>>>
>>>>>> prepare a written response. The WWC Quality Review Team responds
>>>>>> to
>>>
>>>>>> concerns raised about WWC reviews published on our website. These
>>>
>>>>>> quality reviews are undertaken when concerned parties present
>>>
>>>>>> evidence that a WWC review may be inaccurate. When a quality
>>>>>> review
>>>
>>>>>> is conducted, a researcher who was not involved in the initial
>>>
>>>>>> review undertakes an independent assessment of the study in
>>>>>> question.
>>>
>>>>>>
>>>
>>>>>> If a quality review concludes that the original review contained
>>>
>>>>>> errors, a revised version of the review will be published on the
>>>>>> web
>>>
>>>>>> site. However, it is WWC policy that the existing report will
>>>>>> remain
>>>
>>>>>> on the web site, and be revised only if the WWC team determines it

>>>
>>>>>> does contain errors. These quality reviews are one of tools used
>>>>>> to
>>>
>>>>>> ensure that the standards established by the Institute of
>>>
>>>>>> Educational Sciences (IES) are upheld on every review conducted by
>>>
>>>>>> the What Works Clearinghouse.
>>>
>>>>>>
>>>
>>>>>> At this time, we are unable to provide a timeline for a QRT
>>>>>> response.
>>>
>>>>>> Please submit further comments that you would like given to the
>>>>>> QRT.
>>>
>>>>>> You may send them in response to this email.
>>>
>>>>>>
>>>
>>>>>> Thank you,
>>>
>>>>>>
>>>
>>>>>> What Works Clearinghouse
>>>
>>>>>>
>>>
>>>>>> The What Works Clearinghouse was established by the U.S.
>>>>>> Department
>>>
>>>>>> of Education's Institute of Education Sciences to provide
>>>>>> educators,
>>>
>>>>>> policymakers, researchers, and the public with a central and
>>>>>> trusted
>>>
>>>>>> source of scientific evidence of what works in education. For more
>>>
>>>>>> information, please visit http://ies.ed.gov/ncee/wwc/.
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>

>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> -----Original Message-----
>>>
>>>>>>
>>>
>>>>>> From:
>>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@ni
>>>>>> f di.org%3cmailto:jstockard@nifdi.org>>
>>>
>>>>>> [mailto:jstockard@nifdi.org]<mailto:[mailto:jstockard@nifdi.org]><
>>>>>> m ailto:[mailto:jstockard@nifdi.org]>
>>>
>>>>>>
>>>
>>>>>> Sent: Tuesday, August 21, 2012 8:54 PM
>>>
>>>>>>
>>>
>>>>>> To:
>>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@wha
>>>>>> t works.ed.gov%3cmailto:info@whatworks.ed.gov>>
>>>
>>>>>>
>>>
>>>>>> Subject: IES WWC Website: Contact Us: Publications, Intervention
>>>
>>>>>> Reports, Reference ID Number: 1757800276
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> info@whatworks.ed.gov<mailto:info@whatworks.ed.gov<mailto:info@wha
>>>>>> t works.ed.gov%3cmailto:info@whatworks.ed.gov>>,
>>>>>> this email was
>>>
>>>>>> automatically sent through the Contact link on the WWC website.
>>>
>>>>>>
>>>
>>>>>>
>>>

>>>>>>
>>>
>>>>>> From:
>>>>>> jstockard@nifdi.org<mailto:jstockard@nifdi.org<mailto:jstockard@ni
>>>>>> f di.org%3cmailto:jstockard@nifdi.org>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> Message: I am writing to request a Quality Review of the recent
>>>
>>>>>> report on Reading Mastery for students with learning disabilities.
>>>>>> I
>>>
>>>>>> believe that the report misinterpreted both of the studies that
>>>>>> were
>>>
>>>>>> reviewed and should be removed from the website immediately.
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> The WWC analysis was based on two articles. One (Cooke, Gibbs,
>>>
>>>>>> Campbell, & Schalvis, 2004) compared two very similar Direct
>>>
>>>>>> Instruction programs, Reading Mastery and Horizons, and found that
>>>
>>>>>> students in both programs made gains over the academic year that
>>>
>>>>>> were significantly greater than those made by students in national
>>>
>>>>>> and state level populations. Gains in both programs were similar,
>>>
>>>>>> leading the reviewers to conclude that Reading Mastery was no
>>>>>> better
>>>
>>>>>> than its comparison program. They ignored the fact that students
>>>>>> in
>>>
>>>>>> Reading Mastery performed significantly better than the national
>>>>>> or

>>>
>>>>>> state norms and that the comparison program was similar on all but
>>>>>> a
>>>
>>>>>> very few characteristics, essentially a modified version of
>>>>>> Reading
>>>
>>>>>> Mastery. To use a medical analogy, the WWC decision would be
>>>
>>>>>> equivalent to a researcher claiming that aspirin was ineffective
>>>
>>>>>> because Tylenol performed equally well.
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> The second article reviewed (Herrera, Logan, Cooke, Morris, &
>>>>>> Lyman,
>>>
>>>>>> 1997) involved two groups of students, both of whom appear to have
>>>
>>>>>> received Reading Mastery as part of the schools' "usual and
>>>
>>>>>> customary school day curriculum." One group of students also
>>>
>>>>>> received 45 minutes of supplemental phonemic related instruction,
>>>
>>>>>> from their regular classroom teachers, that involved motor
>>>
>>>>>> activities to accompany practice of skills.
>>>
>>>>>> Not surprisingly, the group receiving the additional instruction
>>>>>> had
>>>
>>>>>> significantly larger gains than those who did not have additional
>>>
>>>>>> learning time. Despite these differences in exposure and the fact
>>>
>>>>>> that both groups appear to have had Reading Mastery as their usual
>>>
>>>>>> reading curriculum, the WWC used these results to suggest that
>>>
>>>>>> Reading Mastery could have potentially negative effects. A much
>>>>>> more
>>>

>>>>>> accurate conclusion would be that students who receive additional
>>>
>>>>>> phonics related instruction have greater skill growth.
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> I have prepared a more extensive analysis of these concerns, which
>>>>>> I
>>>
>>>>>> would be happy to send you. Because the problems with the analysis
>>>
>>>>>> are so serious I believe that the report should be immediately
>>>>>> taken
>>>
>>>>>> off your website. In addition, given the very serious
>>>
>>>>>> misinterpretations of the literature described above I will
>>>>>> continue
>>>
>>>>>> to examine WWC reports and procedures.
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> Please acknowledge receipt of my concerns and let me know the time
>>>
>>>>>> line for your review.
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>> Sincerely,
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>

>>>
>>>>>> Jean Stockard, Ph.D.
>>>
>>>>>>
>>>
>>>>>> Director of Research
>>>
>>>>>>
>>>
>>>>>> National Institute for Direct Instruction
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>>
>>>
>>>>>
>>>
>>>>>
>>>
>>>>> Jean Stockard, Ph.D.
>>>
>>>>> Director of Research
>>>
>>>>> National Institute for Direct Instruction Eugene, Oregon
>>>
>>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>>
>>>>>
>>>
>>>>
>>>
>>>>
>>>
>>>> Jean Stockard, Ph.D.
>>>
>>>> Director of Research
>>>
>>>> National Institute for Direct Instruction Eugene, Oregon
>>>
>>>> phone: 541-505-5710 or toll-free 877-485-1973
>>>
>>>>
>>>
>>>>
>>>
>>>>

>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>
>>
>>
>>
>
>
>
>

From: Jean Stockard <jstockard@nifdi.org>
Sent: Friday, February 08, 2013 4:51 PM
To: What Works
Subject: [Fwd: Re: What Works Clearinghouse (QR2012012)]

Dear What Works Help Desk,
I've attached below a message that I sent in mid December asking for additional information about material that you sent to me at that time. Could you please let me know when that information will be sent to me? If you would prefer, I could submit my request under the Freedom of Information Act, but it would, no doubt, be much simpler for all of us if you just sent it to me now.

Thank you for your attention to this matter.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973
---------- Original Message ----------
Subject: Re: What Works Clearinghouse (QR2012012)
From: "Jean Stockard" <jstockard@nifdi.org>
Date: Mon, December 17, 2012 3:29 pm
To: "What Works" <whatworks@mathematica-mpr.com>

Dear WWC Help Desk,

Thank you for sending Dr. Constantine's letter of December 11. To help further my understanding of your procedures and policies I would like to also receive more complete explanations of the reasoning behind the decisions.

I assume that your staff has prepared documents that describe the reasoning for the decisions regarding the Herrera et al and Cooke et al studies in more detail as well as documents that summarize the reasons for rejecting the additional studies for consideration and the reasons for rejecting the views expressed by Dr. Smolkowski. Please forward that material to me at your earliest convenience.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction

Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> Attached is a response to the questions you raised in your message
> regarding the What Works Clearinghouse (WWC) intervention report on
> Reading Mastery for students with learning disabilities.
>
>
>
> Thank you,
>
>
> What Works Clearinghouse
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more
> information, please visit http://ies.ed.gov/ncee/wwc/.
>

From: WhatWorks
Sent: Monday, February 11, 2013 4:20 PM
To: 'jstockard@nifdi.org'
Subject: RE: [Fwd: Re: What Works Clearinghouse (QR2012012)] (WWCPC 3801)

Dear Dr. Stockard,

We are finalizing a response to your most recent round of questions and will send that to you this week. There is no need to make a freedom of information request as all information you request is or will soon be publically available on our website.

What Works Clearinghouse

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Friday, February 08, 2013 4:51 PM
To: What Works
Subject: [Fwd: Re: What Works Clearinghouse (QR2012012)]

Dear What Works Help Desk,
I've attached below a message that I sent in mid December asking for additional information about material that you sent to me at that time.
Could you please let me know when that information will be sent to me? If you would prefer, I could submit my request under the Freedom of Information Act, but it would, no doubt, be much simpler for all of us if you just sent it to me now.

Thank you for your attention to this matter.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973
---------- Original Message ----------
Subject: Re: What Works Clearinghouse (QR2012012)
From: "Jean Stockard" <jstockard@nifdi.org>
Date: Mon, December 17, 2012 3:29 pm
To: "What Works" <whatworks@mathematica-mpr.com>

Dear WWC Help Desk,

Thank you for sending Dr. Constantine's letter of December 11. To help further my understanding of your procedures and policies I would like to also receive more complete explanations of the reasoning behind the decisions.

I assume that your staff has prepared documents that describe the reasoning for the decisions regarding the Herrera et al and Cooke et al studies in more detail as well as documents that summarize the reasons for rejecting the additional studies for consideration and the reasons for rejecting the views expressed by Dr. Smolkowski. Please forward that material to me at your earliest convenience.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> Attached is a response to the questions you raised in your message
> regarding the What Works Clearinghouse (WWC) intervention report on
> Reading Mastery for students with learning disabilities.
>
>
>
> Thank you,
>
>
> What Works Clearinghouse
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more
> information, please visit http://ies.ed.gov/ncee/wwc/.
>

From: WhatWorks
Sent: Monday, July 29, 2013 3:05 PM
To: 'jstockard@nifdi.org'
Subject: RE: Request for Quality Review - Reading Mastery and Students with Learning Disabilities

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC).

We have received your email about the *Reading Mastery* for students with learning disabilities Intervention Report. The WWC Quality Review Team is reviewing your email and will prepare a written response.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Sunday, July 28, 2013 8:08 PM
To: WhatWorks
Subject: Request for Quality Review - Reading Mastery and Students with Learning Disabilities

I am writing to request a Quality Review of the report on Reading Mastery for students with learning disabilities, which was posted in mid-July, 2013. The report did not include a number of studies that should have been included and seriously misinterpreted the one analysis that was reviewed.
In addition, the date on the report (July 2012) is incorrect, for it was clearly not posted until July, 2013 and it is substantially different than the report that was posted in July of 2012. The date should be corrected.
In addition the 2013 report should acknowledge that it replaces a report posted in 2012 that was inaccurate.

The article that was accepted for review in the 2013 report (Cooke, Gibbs, Campbell, & Schalvis, 2004) compared two very similar Direct Instruction programs, Reading Mastery and Horizons, and found that students in both programs made gains over the academic year that were significantly greater than those made by students in national and state level populations. This important positive finding is not even mentioned in the WWC report.
Instead, the review notes that gains in both programs were similar and concludes that Reading Mastery was no better than its comparison program.
To use a medical analogy, the WWC decision would be equivalent to a researcher claiming that aspirin was ineffective because buffered aspirin performed equally well. This conclusion is clearly illogical, and the report should be immediately withdrawn and a correction posted prominently on your website.

Our office has prepared a more extensive analysis of these concerns, including a comparison of the procedures used with the review of Reading Mastery to those used with a recent review of Reading Recovery. The comparison of the two analyses shows both the inaccuracies in the WWC's recent report on Reading Mastery as well as inconsistencies with other reviews. I have taken the liberty of attaching this report.

Please acknowledge receipt of my concerns and let me know the time line for your review and the removal of the report from the website.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-505-5710 or toll-free 877-485-1973

****

EXPERT TRAINING AVAILABLE IN ALL MAJOR DI PROGRAMS AND LEVELS! Join us for the 2013 Annual Training Institute presented by the National Institute for Direct Instruction in San Antonio August 5-8th, 2013. For more information, go to www.nifdi.org/training-institute.

NATIONAL INSTITUTE FOR DIRECT INSTRUCTION
P.O. Box 11248 Eugene, OR 97440 Phone: 541-485-1973 FAX: 541-683-7543

January 2, 2009

Eric A. Hanushek, Ph.D., Chair
National Board for Education Sciences
Paul and Jean Hanna Senior Fellow
Hoover Institution
Stanford University
Stanford, CA 94305-6010

Senator Robert C. Byrd, Chair
U.S. Senate Committee on Appropriations
United States Senate
311 Hart Building
Washington, DC 20510

Dear Dr. Hanushek and Senator Byrd:

As you no doubt know, in July 2008, the Senate Appropriations Committee requested that the National Board for Education Sciences convene a panel of experts to assess the What Works Clearinghouse's (WWC) review processes. In late October the committee completed its examination and called for a "full review." I am writing to give strong support to this recommendation.

In the past few months, in my role as the Director of Research for the National Institute for Direct Instruction (NIFDI), I have reviewed many of the reports developed by the WWC and have found that there are numerous, serious technical problems with these documents. These problems occur in the identification of studies for review as well as in the interpretation and reporting of results. The errors are so severe that I believe that the ratings provided by the Clearinghouse give very misleading pictures of the body of educational research. The enclosed report is a summary of the errors found in just one of these reports. It also includes a description of our attempts to communicate these problems to WWC personnel. Both I and the NIFDI president, Kurt Engelmann, have requested that the WWC remove the erroneous reports from their website. Unfortunately, these requests have been ignored.

NIFDI strongly embraces the goal of providing accurate summaries of the scientific literature to the public. Unfortunately, my own review of their work, as well as that of numerous other scholars, suggests that the WWC appears to have failed in this attempt. Thus, I am very pleased that the expert committee requested a full review of the WWC procedures, and urge you to accommodate that recommendation. If I may provide any additional information, please do not hesitate to contact me via our toll-free number, 1 877 485-1973, or email: jstockard@nifdi.org.

Sincerely,

Jean Stockard

Jean Stockard, Ph.D.
Director of Research

cc: (b)(6)

Appendix B
Statistical Analyses of Case Studies

This appendix presents the statistical analyses of all of the separate analyses used in the third section of the paper. The first section gives the results for analyses that used the cohort control group (CCG) design; the second gives results for analyses that used the cohort control group with historical comparisons (CCGH) design; the third provides a summary overview of the results including effect sizes and confidence intervals for all of the analyses, at both the comparison and site level; and the fourth section presents data included in the SRA report that were not appropriate for analysis with either the CCG or CCGH design.

Analyses Using the Cohort Control Group Design

Table B-1 describes the 7 sites included in the SRA report that had data appropriate for analysis with the CCG design. They represent all parts of the country, schools with varying grade levels and focused on students at different grades. Table B-2 summarizes the socio-demographic characteristics of the schools, and this also indicates substantial variation. While 5 of the seven sites had a majority of Caucasian students, Hispanic students comprised 90 percent of one school and Native American students comprised 100 percent of another.

Separate analyses are presented below for each of the schools listed in Tables B-1 and B-2 with the exception of School 1. The full analyses for that school were given in the text. A summary table at the end of this section includes the effect sizes and results of inferential tests for all of the analyses.

Site # 3: An Arizona Elementary School

Table B-3 reports data from Wilson Primary School, which serves pre-K through third grade students in Phoenix, Arizona. Ninety percent of the students were Hispanic, and 95 percent of the students qualified for free and reduced lunch. The school began using Direct Instruction programs in the fall of 1998. Data were presented in the SRA publication for Stanford Achievement Test/9 scores for third grade students from the spring of 1998 through the spring of 2001. A cohort control group design can be used to examine these data, comparing scores of the cohort tested before the program was implemented to scores after implementation.

The data in Table B-3 indicate that third graders' scores increased markedly with implementation of the new curriculum. While the average student scored at the 17th percentile before implementation, the average student scored at or just below the national average in later years. The effect size that summarizes the change from 1998 to 2001 was .85, substantially larger than the criterion for educational importance, and the t-test assessing this change was also statistically significant.

Site Number 8: A Florida Elementary School

Clay Hill Elementary School served almost 600 students in grades K-6. Ninety-seven percent of the students were Caucasian, and slightly less than half (47%) qualified for free or reduced lunch. The school began using Corrective Reading with all students in grades 4-6 during the 1996-97 school year, initially using it for 90 minutes a day, but in 1997-98 increasing the dosage to 120 minutes. Reading Mastery was used with children in the earlier grades. The SRA publication provided data on the national percentile scores on the California Test of Basic Skills for students in grades 3-6 from 1996, before beginning the program, through 1999, three years later. For the cohort control group design scores of the cohorts who took the test in 1996, before implementation of the program, were compared to those who took the test in 1999, three years after implementation.

Percentiles, corresponding NCE scores, the related effect sizes, t-test results, and lower bound of the uni-directional confidence interval are given in Table B-4. Changes over time were positive for all grades, but the effect size surpassed .25 only for second graders. The t-test reached traditional levels of statistical significance and the confidence interval surpassed zero at only that level as well.

Site 14 - A Kentucky School District

Barren County School District, based in Glasgow, Kentucky, implemented Reading Mastery in five elementary schools in the 2000-2001 school year. The district serves 4100 students, is 97 percent Caucasian, and has a 42 percent free and reduced lunch rate. Data were provided on CTBS scores for grade 3 students from 1999 through 2005, two years before the implementation and 5 years afterwards, for the five schools (Data were also provided for scores on the Kentucky Core Content Test for fifth graders, but I was unable to find state-level data for comparison, so omitted these from analysis.) Results are summarized in Table B-5.

The national average on the norm-referenced CTBS is 50. Before the implementation of RM, in 1999 and 2000, most of the schools had scores that close to or above the national average. However, after implementation of the program, the scores increased markedly and reached their highest point when the third graders would have been exposed to the program since kindergarten (the 2004 and 2005 cohorts). The change in CTBS scores over time can be analyzed using the cohort control group design. Analyses are conducted separately for the five schools, thus adding a natural replication. All of the effect sizes surpassed the usual levels of educational significance, ranging from .71 to 1.70; all of the t-tests were statistically significant, and the minimum value on all confidence intervals for the effect size surpassed the usual level seen as indicating educational significance.

Site 15: A Rural Minnesota School

The Nay Ah Shing School in Onamia, Minnesota served 216 students in grades K-12. All of the students are Native American. At the start of the 2004-05 school year they began their implementation of

the Direct Instruction program Reading Mastery in grades K-3. DIBELS data were provided for students in these grades from the start of the 2004 year (as implementation was beginning) and the end of the 2005-06 year. Thus, a cohort comparison design can be used to compare the percentage of students at benchmark at the beginning of the school year before significant exposure to RM to the percentage at benchmark at the end of two years of implementation. At that point both the kindergarten and first grade cohorts would have been exposed to RM throughout their schooling careers, while the 2nd and 3rd graders would have had 2 years of exposure (grades 1-2 or grades 2-3). Table B-6 gives the percentage of students who reached benchmark at these two time periods, the associated effect sizes, t-tests, and minimum value for the confidence interval for the comparison between cohorts for each grade level. All effect sizes surpassed the usual criterion of educational importance, but effect sizes were larger and results reached standard levels of statistical significance only for the two lowest grades where students had received the most exposure to RM.

Site 16: Special Education Students in a Tennessee District

The Putnam County School District, based in Cookeville, Tennessee, serves almost 10,000 students in grades K-12. Almost half of the students (47%) qualify for free or reduced lunch. Ninety-one percent are Caucasian, and six percent are Hispanic. The district began implementing DI programs (RM for grades K-3 and Corrective Reading for grades 4-8) in all special education classrooms in the 2003-2004 school year. The report gave the percentage of special education students who read at the proficient level, as measured by the state of Tennessee's assessment program (TCAP), in 2002-03 before implementation, through 2005-06. These data and the associated statistics are given in Table B-7. Because no comparable state level data could be found for the 2002-03 year, before implementation began, a cohort control group design was used, comparing the percentage in 2002-3 with the percentage in 2005-06. As shown in the last lines of Table B-7, the associated effect size was large (over 1.0), the inferential test was statistically significant, and the minimum value for the confidence interval was also above 1.0.

Site 23 – An Elementary School in Washington State

Evergreen Elementary in Spokane, Washington served 652 students in grades K-6. Slightly less than a third of the students qualified for free or reduced lunch. Almost 90 % were Caucasian, five percent were Asian, and there were no ELL students. The school began using Reading Mastery in first grade classrooms in the mid 1990s and started using it in grades K-3 in 1998-99. The SRA report provided the percentile rank of grade three students on the ITBS reading test from 2000 to 2005. These data and the accompanying statistics are summarized in Table B-8. Note that all cohorts had some exposure to DI, thus the analysis examines the impact of greater dosage and more consistent exposure throughout the school and curriculum.

The SRA report quoted school officials as saying that the scores for third graders in 2005 "dropped a bit because we had an unusually high number of at-risk readers in that group." Using the language from the CCSS research design literature, they suggested that this cohort was not equivalent to other cohorts. Thus, data for the third grade cohort in 2000 was compared to the third grade cohorts in both 2004 and 2005. The change from 2000 to 2005 was not statistically significant and did not meet the standards of educational importance, while the change from 2002 to 2004 did.

Summary

Table B-9 summarizes the information provided in the analyses above. The first panel of the table gives statistics for each of the 21 separate analyses, and the bottom panel aggregates the information across each of the seven sites. As would be expected, the range of effect sizes is larger with the group of individual comparisons (0.08 to 1.66) than with the values averaged to the site level (.23 to 1.14). The average effect size is, however, slightly larger with the analysis with sites as the unit of analysis (.66 versus .71). On average, the minimum value for the 95 percent (unidimensional) confidence interval was .29 for the set of individual comparisons and .39 for the site level analysis.

Cohort Control Group Design with Historical Comparisons (CCG-H)

Table B-10 summarizes characteristics of the sites that had data appropriate for analysis with the cohort control group with historical comparisons (CCG-H) design. There were 14 sites and 41 separate comparisons. As with the CCG analyses, these sites represented schools and districts from around the country and schools with varying grades. Most of the analyses involved students in grades three and four, with a few also including students in grades 5 and 6. This no doubt reflects the grades at which state assessments are first routinely given. Table B-11 summarizes the socio-demographic characteristics of the students. There was substantial variation across the 15 sites in racial-ethnic composition, in the percentage of students eligible for free or reduced lunch, and in the percentage of ELL students. Results associated with site 1 were reported in the text, and the results for other sites are summarized in the sections below.

Site 2: An Alabama Elementary School

Elba Elementary School is a K-6 school in Alabama. Almost three-fourths of the students received free or reduced lunch. Slightly more than half of the students were African American, and most of the rest were Caucasian. The school began its implementation of Direct Instruction programs (Reading Mastery and Corrective Reading) in the fall of 2005. Table B-12 below reports the percentage of students in grades 3-6 who met or exceeded state reading standards on the Alabama Reading and Mathematics Test from the spring of 2005, before implementation, to the spring of 2007. Also reported are the percentage of students in the state who reached these levels. These data fit the logic of the cohort control group with historical comparisons design. The table includes z-scores (standard scores) that compare the

percentage in the school to that for the state. The final columns give the effect sizes describing the magnitude of the change from 2005 to 2007, adjusting for the changes over that time in the state, as well as the value of t-tests examining the null hypothesis that this change equals zero, and the lower bound of the uni-dimensional 95 percent confidence interval for each grade.

Changes over time, relative to those in the state, were positive for the third, fifth, and sixth grade cohorts, although only those for the fifth grade approached the level usually termed educationally important and none reach standard levels of statistical significance. The results with the fourth grade were opposite to what would be expected, with declining percentages meeting the standards over the time period. The SRA publication reports that a local school official explained the decrease in scores as due to "teacher turnover and a large number of incoming students who couldn't read or read poorly." Using the language of research design this suggests that the cohort groups are not equivalent. In other words, the internal validity of this design is threatened because the composition of the cohorts varied from one time period to the next.

Site 5: A Colorado Elementary School

Table B-13 presents data for Ivywild Elementary School in Colorado Springs. The school served 135 students in grades K through 5. Over 90 percent of the students received free or reduced lunch. Half were Hispanic and most of the other students were Caucasian. Forty-four percent of the students were English Language learners. The school began using the Direct Instruction program Reading Mastery in 2003-04. Table B-13 reports scores on the Colorado Student Assessment Tests from the spring of 2003, before implementation, to the spring of 2005. The percentage of students scoring proficient and advanced as well as the percentage scoring at an unsatisfactory level are reported for each of the grades tested (grades 3-5) for students in the school and students in the state. Results indicate that students in Ivywild were more likely to score at the proficient or advanced level and less likely to score at the unsatisfactory level over time, even with controls for changes within the state as a whole. Patterns of change differed slightly for the three grade levels. Significant changes occurred with both measures for third graders, but only for the percentage scoring at the unsatisfactory level for fourth graders. Neither of the analyses was statistically significant for fifth graders, although both effect sizes surpassed the criterion of educationally important.

Site 6: A Delaware Charter School

East Side Charter School in Wilmington, Delaware served 160 students. Over 90 percent of the students were African American and 85 percent qualified for free or reduced lunch. The school opened in 1997 and began working with Direct Instruction programs in 1998. Table B-14 reports the percentage of third grade East Side students meeting or exceeding state standards from 1999, the first year for which data were available, through 2005. Comparable data are also given for the state. In 1999, only 20 percent

of East Side third graders met or exceeded state reading standards, but this percentage increased markedly in later years, when at least two-thirds of the students met or exceeded standards. The change in East Side relative to the state equaled .79 of a standard deviation in magnitude, this difference was statistically significant ($p = .01$), and the lower limit of the 95 percent (one-tail) confidence interval for the effect size is well above zero. (Note that in 1999 the third graders would have been exposed to RM in third grade. By 2002 students continuing in the school would have been exposed since kindergarten. Thus the comparison is between those with full exposure to the model (since kindergarten) and minimal exposure (only in third grade).)

Site 7 - Alta Vista Elementary School, Florida

Alta Vista Elementary School is a PreK-5 school in Sarasota, Florida serving 723 students. Three quarters of the students qualified for free and reduced lunch. Twenty-six percent were African American, 32% were Caucasian, and 34% were Hispanic. About one-third of the students were English Language Learners. The school began to implement Reading Mastery in the fall of 1999, but did not fully do so throughout all grades until 2002-2003.

Table B-15 reports the percentage of students in grades three to five at Alta Vista and in the state as a whole who scored at the proficient or advanced level on the Florida state reading assessment test (FCAT) in 2002 and 2006. Data were provided for the total group of students and for sub-groups based on race-ethnicity, income, and special education. Comparisons were made between the school scores and the state scores within each sub-group and also between each subgroup and the overall score for the state. For instance, the percentage of African Americans scoring at proficient or advanced at Alta Vista was compared to the percentage of African Americans scoring at that level as well as to the percentage of all students scoring at that level in each year. The effect sizes (d's) indicate the change for each group relative to their own sub-group and to the total group over the four year period. All the d scores were positive, ranging from .09 for all students to .34 for African American students; most of the t-tests were statistically significant; and most of the minimum confidence interval values were above zero. Gains for minority and low income students were more significant in comparison to changes for the total group than for changes within their own group.

Site 10 - Gulf Elementary School

Gulf Elementary is a K-5 school in Cape Coral, Florida with almost 1600 students. Almost three-fourths of the students (73%) were Caucasian, and 17% were Hispanic. Only four percent were ELL students and 21 percent received free or reduced price lunch. Even though the students were more likely than those in the state to score at or above the proficient level on the Florida Assessments, the SRA publication reported that teachers felt they could do better and implemented Direct Instruction in the 1996-1997 school year. Table B-16 reports changes in FCAT scores over time as students who had been

exposed to RM throughout their school career moved through the grades. Using a cohort-control group design, data from the earliest and latest cohorts are compared. The first two cohorts had RM beginning in 2nd or 1st grade, while the remaining cohorts had RM beginning in kindergarten. Engelmann and Engelmann (2007) suggest that it takes several years for a school to fully incorporate the changes involved in implementing Direct Instruction and that stabilization only occurs after a number of years. Thus, the data in Table B-16 can also be seen as examining changes over time as this stabilization occurred. The data show that students in Gulf Elementary were more likely than those in the state to score at level 3 or above at all of the years included. However, the gap between the proportion for the state and Gulf Elementary increased over time. The effect size associated with the change from 1999 to 2005 (controlling for increases in the state) was .19 of a standard deviation, and the change was statistically significant.

Site 11 - Parkway Elementary

Parkway Elementary in Miami served 554 students in grades pre-K through 5. Ninety-four percent of the students were African American, and 91% received free or reduced lunch. The school began using RM sometime in the late 1990s (exact year was not given). Table B-17 reports the percentage of fourth graders scoring at Level 3 or above on the Florida state test from 1999 through 2005. It can be seen that this percentage grew from 29 in 1999 to 60 in 2005. Because the percentage meeting this level rose throughout the state, the Parkway students still had, on average, lower scores than those in the state as a whole, but their scores were increasing relative to the increase in the state. The associated effect size for the changes from 1999 through 2005, relative to the change in the state, was .22 and the t-ratio was 1.37 ($p = .09$).

Site 13 - Highland Elementary School in Kentucky

Highland Elementary is a Pre-K to 5th grade school of 250 students in Hopkinsville, Kentucky. Ninety-two percent of the students received free or reduced lunch. Sixty-one percent were African American, and 38% were Caucasian. Reading Mastery was implemented in grades K-5 in October 2005. Table B-18 reports the percentage of fourth graders scoring at the proficient or higher range on the Kentucky Core Content Test administered in the spring of 2005, 2006, and 2007. In 2005, before RM was implemented, the fourth graders had scores that were .84 of a standard deviation below those for students in the state as a whole, but by 2007 their scores were .20 of a standard deviation above those for the state. The associated effect size was 1.04, indicating that, relative to changes in the state, scores in Highland increased by more than a standard deviation. This change was highly statistically significant.

Site 17: Brunswick County, NC Schools

The SRA report included information on schools from Brunswick County, North Carolina that began using Reading Mastery, and other DI programs as needed, in the 2004-2005 school year.

Information was provided on the percentage of students scoring at or about the proficient level in reading with the End-of-Grade tests in grades 3, 4, and 5 from spring of 2004 through spring of 2007 for two schools. In 2004 none of the cohorts would have been exposed to RM, but by 2007 they would have had three years of exposure (beginning in grade 1 to grade 4). Table B-19 reports these data for the two schools and each grade level, as well as for the state as a whole. Over time, students in all grades and both schools were more likely to score at the proficient or higher level in reading. The effect sizes surpassed the .25 level for both the third and fourth grade comparisons, and all but a comparison for fifth graders in one of the schools was statistically significant.

Site 18 - Cleveland, Ohio – Louisa May Alcott Elementary

Louisa May Alcott Elementary in Cleveland, Ohio served 208 students in grades K-6. One hundred percent of the students qualified for free or reduced lunch. Slightly more than a third (35%) were African American, close to half (47%) were Caucasian, and most of the rest were Hispanic (12%). The school fully adopted the DI programs Reading Mastery and Language for Learning in 1998-1999. Table B-20 reports the percentage of fourth graders passing the state reading tests from the spring of 1998, when none would have been exposed, through 2006. Beginning in 2003 fourth graders continuing in the school would have been exposed to RM and LL since kindergarten, thus having the maximum exposure. The data indicate dramatic increases in the percentage of fourth graders meeting state standards, from less than a third before implementation to well over three-quarters in 2003 and later. There was a noticeable drop in 2006, which the school officials attributed to an influx of new students. Using the language of experimental design, this would suggest that the 2006 cohort was not strictly comparable to earlier cohorts. Thus comparisons of 1998 and 2006 as well as 1998 and 2005 are included. Both comparisons indicate substantial effect sizes and statistically significant results, although results are smaller for the former comparison.

Site 19 - An Oregon School District

According to the SRA material, at the time of publication Milton-Freewater Unified School District served slightly more than 2,000 students in grades K-12. About half of the students were Hispanic and almost all of the rest were Caucasian. Almost three-fourths (72%) of the students received free or reduced price lunch. The district adopted the Direct Instruction program Reading Mastery as its core reading curriculum in the fall of the 2003-2004 school year. Table B-21 reports the percentage of third grade students who met or exceeded state reading standards in the spring testing in 2003, the year before the program was implemented, through 2006, after three years of implementation. Data are presented for both the total group of students and for Hispanic students in the school district as well as in the state as a whole. Thus these data may be seen as involving a matched cohort design with comparative historical data.

Three types of statistical comparisons are included in Table B-21: one compares the scores of the total group of district students to all students in the state, another compares the scores of Hispanic students to Hispanic students in the state, and the third compares the scores of Hispanic students to all students in the state. In 2003, before the program was implemented the effect sizes were negative (for comparisons to the total group) or zero (comparisons between the Latino groups). In other words, before implementation the third grade Hispanic students in the district met or exceeded state reading standards at the same proportion as third grade Hispanic students throughout the state, but in comparison to the total population, the district's students were less likely to meet the standards. Over time, however, cohorts had increasingly higher scores. By 2006 after three years of implementation, 95% of all students (both the total group and the Hispanic students) met or exceeded standards. The percentage of students meeting standards also increased in the state, but not at the same pace. The effect sizes reflecting these changes would be considered large, ranging from .65 for the comparison between the total groups, to .81 for comparisons involving the Hispanic students. All of the associated t-tests are statistically significant and minimal values of the confidence intervals are well above zero.

Site 23 - Evergreen Elementary in Spokane, Washington

Evergreen Elementary in Spokane, Washington served 652 students in grades K-6. Slightly less than a third of the students qualified for free or reduced lunch. Almost 90 % were Caucasian, five percent were Asian, and there were no ELL students. The school began using Reading Mastery in first grade classrooms in the mid 1990s and started using it in grades K-3 in 1998-99. In addition to ITBS data described in the previous major section, the SRA report provided data regarding the percentage of fourth grade students who met or exceeded state reading standards on the Washington state assessment. These data and the accompanying statistics are summarized in Table B-22. Note that all cohorts had some exposure to DI, thus the analysis examines the impact of greater dosage and more consistent exposure throughout the school and curriculum. It can be seen that all cohorts of Evergreen students were more likely than students in the state as a whole to score at the proficient or higher level. There was little change in the difference beween the school and the state over the time period.

Site 24 – A Wisconsin Elementary School

Honey Creek Continuous Progress Elementary is a Pre-K-5 school with 374 students in Milwaukee, Wisconsin. Forty-four percent of students received free or reduced lunch. Slightly less than ten percent were African American, fifteen percent were Hispanic, and 73 percent were Caucasian. The school began using Direct Instruction programs in the fall of 1998. The report provided data on the percentage of Grade 4 students who scored proficient or advanced on the Wisconsin Knowledge and Concepts Exam in 1997, before implementation, and from 1999 to 2005, after implementation. These data, as well as comparative data for the state, are in Table B-23. Before the implementation and in the

first years, the fourth grade students were less likely to meet proficiency than others in the state, but this changed in later years, beginning with the cohort that had experienced RM since first grade. The effect size associated with the change met the usual level of educational importance and the t-test was statistically significant.

Summary

Table B-24 summarizes the information provided in the analyses in this section. The first panel of the table gives statistics for each of the 40 separate analyses, and the bottom panel aggregates the information across the thirteen sites. As would be expected, the range of effect sizes is larger with the group of individual comparisons (-.26 to 1.04) than with the values averaged to the site level (.03 to 1.04). As with the analyses with the CCG designs, the average effect size is slightly larger with the analysis with sites as the unit of analysis (.43 versus .36). On average, the minimum value for the 95 percent (unidimensional) confidence interval was .07 for the set of individual comparisons and .13 for the site level analysis.

Other Comparisons

As noted above, data from eight of the sites described in the SRA report did not fit the criteria of the CCG or CCG-H designs. Two of the sites (numbers 4 and 22) had implemented DI programs for a number of years and did not have pretest data. Another site (number 9) had data on changing patterns of DIBELS scores from one year to another, two (numbers 16 and 25) had implemented the program with only some students, but reported data for the entire school, and three (numbers 12, 20, and 26) reported data for only small numbers of special education students. The sections below discuss each of these sets of data and possible ways to analyze them.

Sites 4 and 22 – Maintaining Effects

If a site has implemented DI for a number of years they may be interested in using assessment data to ensure that they can maintain improvements that have been made. Two of the descriptions included in the SRA publication can illustrate this situation. The first is site four, Richfield Elementary in Corning, California, a K-8 school that began using DI materials in the mid 1990s. All of the data in the SRA report came from years after that time, and are summarized in Table B-25. This table gives the percentage of second grade students in the school that scored at the proficient or advanced level from 2001 to 2005 in the school and in the state. Data are given for all students and for those classified as economically disadvantaged using the state's criteria. The SRA write-up quoted a school official as noting that the scores "dipped slightly in 2005 because of an influx of students." Thus, comparisons are made between 2001 and 2004 as well as 2001 and 2005. Three contrasts are included within each pair of years: 1) the total group of students in the school compared to the total group for the state, 2) disadvantaged students in the school compared to disadvantaged students in the state, and 3)

disadvantaged students in the school compared to the total group within the state. All of the effect sizes for the comparisons between 2001 and 2004 met usual criteria of educational significance and the t-test comparison of the disadvantaged students to the total group also met usual standards of statistical significance. The differences between 2001 and 2005 were smaller in magnitude and the comparison (d and t) for the total group of students was negative. Thus, even though the school had implemented DI for several years before the first year of data, scores continued to improve. This may reflect the stabilization process described by Engelmann and Engelmann (2007), in which teachers and students become fully familiar with the program and it has been institutionalized at all grades over the entire career span of students.

The second site for which there was no pre-implementation data was Cache County School District, located in North Logan Utah (site 22). The district serves more than 13,000 students in grades K-12. Slightly less than a third qualified for free or reduced lunch. More than 90 percent were Caucasian and most of the remaining students (6%) were Hispanic. The school began using DI materials in the 1970s and has continued doing so in the ensuing decades. The SRA report gives the percentage of district students who scored at or above the proficient level in the Utah Criterion-Referenced test for students in Grades 1-5 in both 2004 and 2005 as well as DIBELS scores for one cohort from kindergarten through third grade. These data are given in Tables B-26 and B-27.

Because the district has been using DI materials for decades, the administration is probably not interested in determining if the implementation will alter previous performance. Instead, the district is probably interested in the extent to which students in the district perform relative to those in the state as a whole and if they are maintaining an advantage over time. The data in Table B-26 address that issue by giving the percentage of Cache County students scoring at the proficient level and the corresponding percentage in the state for grades 1-5 and both 2004 and 2005. The effect sizes were calculated using the population (state) data to obtain the standard deviation and indicate the extent to which Cache County students differed from those in the state in standard deviation terms. It can be seen that the effect sizes are all positive and range from .12 of a standard deviation, for fourth graders, to over .30 for students in the primary grades. The z-scores test the null hypothesis that the proportion in the district equals the proportion in the state and all of the results are statistically significant.

Table B-27 summarizes DIBELS scores of one cohort of CACHE County elementary students from kindergarten through third grade. The percentage of students at risk of future difficulties at mid-year and the percentage at low risk or “on target” at the end of each year are reported. The table also gives data from a set of national percentiles with the DIBELS system (Good, et al, 2002). These can be used to compare to the Cache County data in the manner that was used in Table B-26. Thus the table includes the effect size (standard deviation score) comparing the district percent to the national data and the

accompanying inferential test. (Signs were changed to match other comparisons so that positive scores indicate that those receiving DI had higher achievement.) Results indicate that the Cache County students were significantly less likely to be at risk and significantly more likely to have established skills than students in the national sample with all but the end of year kindergarten measure.

Site 9 - Changing DIBELS Scores Over Time

Edgewood Academy in Fort Myers, Florida, is a Pre-K-5 school with 700 students. Almost ninety percent (88%) of the students received free or reduced lunch, 34 percent were African American, twenty percent were Caucasian, and 40 percent were Hispanic. Almost 30 percent (29%) were ELL students. The school began using Reading Mastery in grades K-5 in August, 2006. Before that time they had found that the percentage of students who were assessed, using DIBELS, as being at grade level or low risk declined from the fall to mid-year testing. After implementing RM, however, this pattern altered. The relevant data, from 2004-05 and 2006-07 are summarized in the first panel of Table B-28.

The second panel of Table B-28 reports the difference between fall and mid-year testing and associated statistics for each grade. It can be seen that in 2004-05, before implementing RM, the percentage of students scoring at grade level declined during the first part of the school year in all but third grade. However, in 2006-07, the percentage of students at grade level increased in all grades.

For the most accurate statistical test for these data one would want to examine the changes for individual students, rather than looking at the percentages for the total group. At the same time, one can use these data to get an estimate of the magnitude of the change over time, by examining the difference in the two proportional differences. Essentially one can test the null hypothesis that the difference between fall and winter in 04-05 equals the difference in 06-07 (see Blalock, 1979, pp. 234-236 for an example). The corresponding effect sizes and tests of this hypothesis are shown in the last three columns of the second panel of Table B-28. All of the effect sizes are positive and the results for grades one, two and four reach traditional levels of statistical significance and educational importance.

Sites 16 and 25 – Implementation with a Subset of Students

Two of the SRA reports involved sites where DI programs had been implemented with students believed to be at risk, such as special education students, but data were reported for the entire school. The first (Site 16) involved Iredell Statesville Schools in Statesville, North Carolina. The district serves 21,000 students in Pre-K through twelfth grade. Slightly more than a third (35%) have free or reduced lunch, eighteen percent are African American, 72 percent are Caucasian , and seven percent are Hispanic. There were no reported ELL students. The district began using DI programs (Reading Mastery and Corrective Reading) in special education classrooms in the fall of 2003 and adopted the programs for at-risk students in regular education in the next school year (2004-05). The SRA report gives the percentage of students in grades three to eight who attained adequate yearly progress, based on the North Carolina

End-of-Grade Test, from 2002-03, before DI was implemented, to 2005-06. Data for both students with disabilities and all students were given and are reproduced in Table B-29 along with the corresponding data for the state as a whole.

These data were analyzed with the same techniques used with the cohort control group model with historical comparisons. Three comparisons were conducted: 1) students with disabilities in the district with students with disabilities in the state, 2) students with disabilities in the district versus the total group of students in the state, and 3) the total group of students in the district with the total group of students in the state. Note that, unlike the comparisons in the earlier section, these cohorts are not independent of each other because the grade range overlaps from one year to another. (For instance, students in grade 3 in the first year would be in the comparison for all the subsequent years.) However, the results presented can be instructive, in that they show that the students in the district scored more highly relative to others in the state over the time period and that the increase was especially marked when those with disabilities were compared with the total group of students in the state.

The second set of data that involved schools where DI programs were given only to a subset involved site number 25, including two elementary schools from Monroe, Wisconsin. One of the schools, Northside Elementary, served 339 students in grades Pre-K to 5. Ninety-five percent of the students were Caucasian and fifteen percent qualified for free or reduced lunch. The other school, Parkside Elementary, served 358 students, 93 percent of whom were Caucasian. Thirty percent of the Parkside students qualified for free or reduced price lunches. The schools implemented Reading Mastery with at-risk and special education students in grades 1-5 in the mid 1990s and added the program to the kindergartens at the start of the 2004-05 school year. The SRA publication reported the percentage of grade 3 students who scored at the proficient or advanced level in reading on the Wisconsin Reading and Comprehension Test in the spring of 2000 to 2005. Although the data do not fit the model of either the CCG or CCG-H design, they can be compared to data obtained for the state as a whole and examine the extent to which the schools' third graders have scores that exceed those of other students in the state. These results are given in Table B-30 below.

It can be seen that the students in the school were more likely than those in the state to score at the proficient or advanced level at each year, although the difference was at its smallest point in 2005, the last year of data collection. Thus, over time, even though the percentages increased in the two schools, the increase was not as large as in the state as a whole and thus the effect sizes were slightly negative and the t-ratios were not significant.

Sites 12, 20, and 26: Reports Regarding Small Numbers of SPED Students

Three of the reports involved data regarding only small numbers of special education students. The first such report (site number 12) came from Marmaton Valley Elementary School in Moran, Kansas,

which uses Reading Mastery as its core curriculum school wide. The report gave the scores of 16 special education students on the Kansas State Reading Assessment. During the 2006-07 academic year 13 of the 16 students identified for special education met or exceeded the state standards, while the other three approached the standards. These numbers were compared to those that would be expected given the results for the state as a whole, using a standard chi-square test. Because the chi-square requires that a minimum expected frequency equal five or larger, the comparisons involved the number of students who a) met or exceeded versus others and b) the number of students who exceeded the standards versus the others. These data are shown in Table B-30 and indicate that, although more students met or exceeded the standards in the school than in the state as a whole, the difference was not so large that it would be judged statistically significant.

Site number 20 was the Clover School District in Clover, South Carolina, which served 5,600 students from K to 12, the majority of whom (83%) were Caucasian. Slightly less than a third qualified for free or reduced lunch. The district began using DI materials with their special education students in the 2005-2006 academic year, and the SRA report described the growth of these students during that year in comparison to general education students. Academic growth was measured by the scores on the state-aligned computerized adaptive test from the Measures of Academic Progress system. The district was especially interested in comparing the growth of the special education students to all students and thus the extent to which the special education students might be able to eventually catch up with their peers.

Table B-32 summarizes these data. It includes the average growth of special education students and all students in the Clover school district as well as average growth for students in the nation, which was obtained from the MAP website. I was not able to find standard deviations associated with either the total district or total national figures. Thus, inferential tests were not possible. Instead, Table B-32 includes ratios of the growth of special education students in the district to the growth of all students in the district and the national sample. It can be seen that growth was greater in the special education population in grades five to eight, but less so in the lower grades.

The last site in the analysis was East High School in Wauwatosa, Wisconsin, a school serving almost 1200 students in grades nine through 12. Seven percent of the students were eligible for free or reduced lunch, 80 percent were Caucasian, and eleven percent were African American. The school used Corrective Reading and other DI programs for students felt to be struggling readers, and the SRA report described the average gain over three school years for students in the program, based on scores on the Woodcock-Johnson III. These data are summarized in Table B-33. It can be seen that students in two years, 2003-04 and 2005-06, had gains that were greater than would be expected, while those in the 2004-05 year were less than would be expected. The average across years, weighted by the number of students in each year, is slightly larger than would be expected. Note that the samples of students are probably not

independent, for students may have been in a sample for more than one year, so application of inferential statistics, without more data, would be difficult.

Table B-1

Characteristics of SRA/McGraw Hill Reports on DI Interventions Appropriate for Analysis with CCG Design

Report #	# of Comparisons	State	Grades in School	# of Students	Assessment	Grades for Analysis
1	3	Alaska	K-5	300	CBM	K
3	1	Arizona	PreK-3	813	NR	3
8	5	Florida	K-6	593	NR	K-6
14	5	Kentucky (5 schools)	Pre-k - 6	318 (average)	NR	3
15	4	Minnesota	K-12	216	CBM	K-3
21	1	Tennessee	K-12	9918 (district)	SA	3-8, SPED only
23	3	Washington	K-6	652	NR	3

Note: Two sites in the list (numbers 1 and 23) also had data that were analyzed with the CCG-H design. CBM tests were DIBELS, NR refers to norm referenced tests, and SA refers to state assessment.

Table B-2

Socio-Demographic Characteristics of Schools in CCG Comparisons

Report #	Free or Reduced Lunch	African American	Caucasian	Hispanic	Asian-American	Native Amer./Al.Native	ELL
1	64	0	75	0	0	25	0
3	95	4	3	90	1	2	52
8	47	1	97	1	0	1	1
14	42	1	97	1	1	0	0
15	100	0	0	0	0	100	0
21	47	2	91	6	1	0	5
23	30	4	89	2	5	0	0

Note: All numbers are percentage of students listed in the report as having a given characteristic.

Table B-3

Spring SAT/9 Scores, 1998 to 2001, Arizona Third Graders, - Site 3

Year	Percentile	NCE Score	Z-score
1998	17	29.9	-0.95
1999	50	50.0	0.00
2000	49	49.5	-0.02
2001	46	47.9	-0.10
Effect size, change 1998 to 2001		0.85	
T-test, change 1998 to 2001		7.64	
prob.		<.0001	
Lower limit of 95% c.i. (one-tail)		0.67	

Note: The standard deviation for NCE scores is 21.06. Z-scores were calculated for each year by comparing the NCE score to the national average of 50.0 and dividing by 21.06. Effect size reflects the change in NCE scores over time relative to the s.d. of 21.06. Direct Instruction was first implemented in the 1998-1999 year.T-test was computed with an on-line calculator, with sample size of 160 estimated for each school year.

Table B-4

CTBS Scores, 1996 and 1999, Site 8, A Florida Elementary School

	1996		1999					
Grade	% ile	NCE	% ile	NCE	d	t	prob.	C.I. Min.
Grade 2	28	37.7	50	50.0	0.58	3.80	0.0001	0.33
Grade 3	42	45.7	45	47.4	0.08	0.50	0.31	-0.18
Grade 4	44	46.8	47	48.4	0.08	0.49	0.31	-0.18
Grade 5	40	44.7	49	49.5	0.23	1.49	0.07	-0.02
Grade 6	48	48.9	55	52.6	0.18	1.15	0.13	-0.08

Note: Percentiles were translated to NCE scores for calculations. By definition, the standard deviation for NCE scores is 21.06. Effect sizes were calculated by subtracting the NCE scores for the two cohorts (1996 and 1999) and dividing by the standard deviation of 21.06. An on-line calculator (http://www.dimensionresearch.com/resources/calculators/ttest.html) was used to calculate the t-scores with an estimated class size of 85 students for each year. Thus degrees of freedom = 164. All t-test probabilities are one-tailed.

Table B-5

CTBS Scores of Third Grade Students, 1999-2005, 5 Elementary Schools in Kentucky (site 14)

Year	Austin Tracy	Hiseville	Park City	Red Cross	Temple Hill	Years of RM
1999	54	46	38	63	58	0
2000	56	50	59	64	55	0
2001	57	55	49	69	59	1
2002	41	58	60	62	62	2
2003	70	62	39	71	68	3
2004	77	67	55	72	71	4
2005	69	63	73	77	73	4
Effect size change 1999 to 2005	0.71	0.81	1.66	0.66	0.71	
t	2.94	3.91	7.24	4.23	3.19	
df	66	92	74	160	78	
prob.	0.0023	<.0001	<.0001	<.0001	0.001	
C.I. Min.	0.31	0.47	1.28	0.41	0.34	

Note: The standard deviation for the CTBS scores is 21.06. Effect sizes were calculated by subtracting the NCE scores for the two cohorts and dividing by the standard deviation of 21.06. An on-line calculator (http://www.dimensionresearch.com/resources/calculators/ttest.html) was used to calculate the t-scores with class sizes estimated given the school size as of 2009-10. All probability values are one-tailed.

Table B-6

Changes in Percentage of Students Meeting DIBELS Benchmarks, K-3, Minnesota School (site 15)

Grade	% Meeting, Fall, 2004	% Meeting, Spring 2006	effect size	t-test change	prob.	Min. C.I.
Kindergarten	10	70	1.22	3.21	0.002	0.60
Grade 1	21	92	1.43	3.86	0.0003	0.82
Grade 2	40	55	0.30	0.53	0.30	-0.63
Grade 3	35	67	0.64	1.52	0.07	-0.05

Note: Effect sizes were calculated by subtracting the proportions for the two cohorts and dividing by the common standard deviation. An on-line calculator (http://www.dimensionresearch.com/resources/calculators/ztest.html) was used to calculate the z-scores. Class sizes were estimated given the school size as the base. All probability values are one-tailed and adjusted for the small sample size.

Table B-7

SPED Students, Grades 3-8, Reading at Proficient Level or Higher, 2002-03 through 2005-06, Tennessee School District (Site 21)

Year	Percent
2002-03	37
2003-04	58
2004-05	80
2005-06	85
Effect Size	1.14
z-value	20.83
df	1798
prob.	<.0001
C.I. Min	1.05

Note: Sample size based on district enrollment and state reports was estimated at 900 each year.

Table B-8

ITBS Reading Scores, 2000 to 2005, Site 23, a Washington State Elementary School (Site 23)

Year	% ile	NCE	Grades had DI	
2000	58	54.3	1 to 3	
2001	67	59.3	K to 3	
2002	70	61.0	K to 3	
2003	73	62.9	K to 3	
2004	71	61.7	K to 3	
2005	64	57.5	K to 3	
Comparison	d	t	prob.	C.I. Min.
2000 vs. 2005	0.16	1.04	0.15	-0.09
2000 vs. 2004	0.35	2.40	0.01	0.11

Note: DF = 184, reflecting an estimated class size of 93 for each year.

Table B-9

Summary of Results with Analyses Using Cohort Control Group Designs, 7 Sites and 21 Comparisons

Site #	Grade	Gr. began RM	Years DI at School	effect size	z	prob.	Min. for C.I.
1	K	K	4	0.62	2.77	0.006	0.25
2	K	K	4	0.48	2.17	0.002	0.12
3	K	K	4	0.87	4.15	<.0001	0.53
3	3	1	3	0.85	7.64	<.0001	0.67
8	2	K	3	0.58	3.8	0.0001	0.33
8	3	1	3	0.08	0.5	0.31	-0.18
8	4	2	3	0.08	0.49	0.31	-0.19
8	5	3	3	0.23	1.49	0.07	-0.02
8	6	4	3	0.18	1.15	0.13	-0.08
14	3	K	5	0.71	2.94	0.0023	0.31
14	3	K	5	0.81	3.91	<.0001	0.47
14	3	K	5	1.66	7.24	<.0001	1.28
14	3	K	5	0.66	4.23	<.0001	0.4
14	3	K	5	0.71	3.19	0.001	0.34
15	K	K	2	1.22	3.21	0.002	0.59
15	1	K	2	1.43	3.86	0.003	0.82
15	2	1	2	0.3	0.53	0.3	-0.63
15	3	2	2	0.64	1.52	0.07	-0.05
21	3-8, SPED		2	1.14	20.83	<.0001	1.05
23	3	K	7	0.16	1.04	0.15	-0.09
23	3	K	6	0.35	2.4	0.01	0.11
Min.				0.08			-0.63
Max.				1.66			1.28
Aver.				0.66			0.29

Site-Level Summaries

Site #	Average effect size	Aver. Min. for C.I.
1	0.66	0.3
3	0.85	0.67
8	0.23	-0.03
14	0.91	0.56
15	0.9	0.18
21	1.14	1.05
23	0.25	0.01
Min.	0.23	-0.03
Max.	1.14	1.05
Aver.	0.71	0.39

Table B-10

Characteristics of SRA/McGraw Hill Reports on DI Interventions Appropriate for Analysis with CCG-H Design

Report #	# of Comparisons	State	Grades in School	# of Students	Grades for Analysis
1	1	Alaska	K-5	300	3
2	4	Alabama	K - 6	470	3 to 6
5	6	Colorado	K - 5	135	3 to 5
6	1	Delaware	PreK-7	160	3
7	11	Florida	Pre-K-5	723	3 to 5
10	1	Florida	K - 5	1577	4
11	1	Florida	PreK - 5	554	4
13	1	Kentucky	PreK - 5	250	4
17	6	N. Car.	PreK - 12 (district)	675	3 to 5
18	2	Ohio	K-6	208	4
19	3	Oregon	K-12	2059	3
23	1	Washington	K-6	652	4
24	1	Wisconsin	PreK-5	374	4
25	2	Wisconsin	PreK-5	339	3

Note: Sites 1 and 23 also had analyses that were appropriate for the CCG design. All sites had state assessment data.

Table B-11

Socio-Demographic Characteristics of Schools in CCG Comparisons

Report #	Free or Reduced Lunch	African American	Caucasian	Hispanic	Asian-American	Native Amer./Al.Native	ELL
1	64	0	75	0	0	25	0
2	72	55	42	3	0	0	0
5	92	4	47	49	0	0	44
6	85	92	2	6	0	0	0
7	76	26	32	34	7	1	32
10	21	3	73	17	1	6	4
11	91	94	1	3	1	1	3
13	92	61	38	1	0	0	0
16	35	18	72	7	3	0	0
17	51	23	70	6	0	0	0
18	100	38	47	12	0	0	0
19	72	0	49	50	0	0	58
23	30	4	89	2	5	0	0
24	44	8	73	15	3	1	0
25-a	15	2	95	3	0	0	0
25-b	30	3	93	3	0	1	0
min	15	0	1	0	0	0	0
max	100	94	95	50	7	25	58
mean	60.6	26.9	56.1	13.2	1.3	2.2	8.8

Note: Socio-demographic data were provided for both of the schools included for site number 25. All numbers are percentage of students listed in the report as having a given characteristic.

Table B-12

Percent Meeting or Exceeding State Reading Content Standards, Grades 3 to 6, An Alabama Elementary and State of Alabama, 2005 to 2007 (Site # 2)

Grade 3

Year	School	State	n	Z relative to state	Effect Size (d)	t	prob.	Min. C.I.
2005	80	81.07	66.00	-0.03				
2006	86	83.65	59	0.06				
2007	88	84.9	65	0.09	0.11	0.65	0.26	-0.17

Grade 4

Year	School	State	n	Z relative to state	Effect Size (d)	t	prob.	Min. C.I.
2005	92	83.34	51	0.23				
2006	84	84.4	76	-0.01				
2007	84	85.01	56	-0.03	-0.26	-1.35	0.91	-0.58

Grade 5

Year	School	State	n	Z relative to state	Effect Size (d)	t	prob.	Min. C.I.
2005	71	80.93	79	-0.25				
2006	82	80.71	55	0.03				
2007	83	84.6	67	-0.04	0.21	1.26	0.10	-0.06

Grade 6

Year	School	State	n	Z relative to state	Effect Size (d)	t	prob.	Min. C.I.
2005	88	81.65	75	0.16				
2006	90	83.08	73	0.18				
2007	94	84.87	59	0.25	0.09	0.52	0.30	-0.20

Note: Z-scores reflect the difference between the school and state values in standard-score form. D values are the effect of the change from 2005 to 2007 as a percentage of the standard deviation. T-values test the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom = $(n_1 + n_2) - 2$, where n_i equals the sample size in the school for each year. All probability values are one-tail.

Table B-13

Students Scoring Proficient or Advanced and Unsatisfactory In Reading on Colorado Student Assessment Program, Ivywild Elementary and State, 2003-2005 (Site # 5)

	Percent Proficient or Advanced			Percent Unsatisfactory		
Grade 3	School	State	Z-score	School	State	Z-score
2003	43	75	-0.74	30	7	0.90
2004	25	73	-1.08	8	8	0.00
2005	67	72	-0.11	5	9	-0.14
Grade 4						
2003	23	63	-0.83	32	12	0.62
2004	23	63	-0.83	32	11	0.67
2005	24	64	-0.83	14	13	0.03
Grade 5						
2003	35	66	-0.65	35	13	0.65
2004	34	69	-0.76	22	11	0.35
2005	55	69	-0.30	23	11	0.38

Effect Sizes and T-Tests, Changes 2003 to 2005, by Grade

	Percent Proficient or Advanced				Percent Unsatisfactory			
	D	t	prob.	Min. C.I.	d	T	prob.	Min. C.I.
Grade 3	0.63	2.08	0.02	0.13	1.04	3.45	0.001	0.55
Grade 4	0.00	-0.02	0.51	-0.50	0.59	1.94	0.03	0.09
Grade 5	0.35	1.17	0.13	-0.14	0.27	0.90	0.19	-0.23

Note: Z-scores reflect the difference between the school and state values in standard-score form. D values are the effect of the change from 2003 to 2005 as a percentage of the standard deviation. T-values test the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom equal 40. All probability values are one-tail reflecting the directional hypotheses. Signs have been changed on the statistics regarding unsatisfactory progress to match the direction in other tables, with a positive values indicating increased achievement for DI students over time.

Table B-14

East Side Charter School, Percent Third Graders Meeting or Exceeding State Reading Standards, 1999 to 2005 (Site # 6)

Year	School	State	Z relative to state
1999	20	68.62	-1.05
2000	83	76.80	0.15
2001	65	74.11	-0.21
2002	69	79.31	-0.25
2003	100	79.30	0.51
2004	77	82.36	-0.14
2005	75	84.45	-0.26
Effect Size - Change 1999 to 2000		0.79	
t-test		2.36	
Probability		0.01	
Min. C.I.		0.24	

Note: Z-scores reflect the difference between the school and state values in standard-score form. Effect size is Cohen's d describing the effect of the change from 1999 to 2005 as a percentage of the population standard deviation. T-value tests the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom equal 34. Probability value is one-tail reflecting the directional hypotheses.

Table B-15

Students in Grades 3-5 Scoring Proficient or Advanced in Reading on FCAT, 2002 and 2006, Alta Vista Elementary and State of Florida, Total Group and Sub-Groups (Site # 7)

Percentage of Students Scoring Proficient or Advanced and Z-Scores, 2002				
	Alta Vista	State	Z - same group	Z - Total Group
African American	26	37	-0.22	-0.60
Hispanic	38	46	-0.16	-0.36
Low Income	42	43	-0.03	-0.28
Exceptional Student Education	22	26	-0.09	-0.68
Caucasian	63	68	-0.11	0.14
All	51	56	-0.10	-0.10

Percentage of Students Scoring Proficient or Advanced and Z-Scores, 2006				
	Alta Vista	State	Z - same group	Z - Total Group
African American	57	53	0.07	-0.26
Hispanic	65	63	0.03	-0.09
Low Income	66	58	0.16	-0.06
Exceptional Student Education	48	39	0.18	-0.45
Caucasian	84	79	0.12	0.32
All	69	69	-0.01	-0.01

Effect Sizes and t-tests (Within Group)					
	d	t	prob.	df	Min. C.I.
African American	0.29	1.99	0.02	180	0.05
Hispanic	0.20	1.55	0.06	252	-0.01
Low Income	0.18	1.99	0.02	472	0.03
Exceptional Student	0.27	1.46	0.07	116	-0.03
Caucasian	0.23	1.74	0.04	228	0.01
All	0.09	1.25	0.11	718	-0.03

Effect Sizes and t-tests (Comparisons to Total)					
	d	t	prob.	df	Min. C.I.
African American	0.34	2.33	0.01	180	0.10
Hispanic	0.28	2.20	0.01	252	0.07
Low Income	0.22	2.36	0.01	472	0.07

Exceptional Student	0.23	1.25	0.11	116	-0.07
Caucasian	0.18	1.39	0.08	228	-0.03
All	0.09	1.25	0.11	718	-0.03

Note: Z-scores reflect the difference between the school and state values in standard-score form. Effect size is Cohen's d describing the effect of the change from 2002 to 2006 as a percentage of the population standard deviation. T-value tests the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Probability value is one-tail reflecting the directional hypotheses.

Table B-16

Students Scoring at 3 and above, FCAT Gulf Elementary and State of Florida, 1999 to 2005 (Site # 10)

Percentages, Gulf Elementary and State, Z-Scores and Year Began DI

Year	%	State %	z relative to state	Year began DI
1999	59	52	0.14	2nd
2000	64	52	0.24	1st
2001	69	53	0.32	K
2002	74	55	0.38	K
2003	76	60	0.33	K
2004	89	70	0.41	K
2005	86	71	0.33	K

Change 1999 to 2005	
effect size	0.19
t-ratio	2.18
prob.	0.01
Min. C.I.	0.05

Note: Z-scores reflect the difference between the school and state values in standard-score form. D values are the effect of the change from 1999 to 2005 as a percentage of the standard deviation. T-value tests the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom = (n1 + n2) - 2, where ni equals the sample size in the school for each year. Probability value is one-tail.

Table B-17

Students Scoring at 3 and above, FCAT Parkway Elementary and State of Florida, 1999 to 2005 (Site # 11)

Percentages, Parkway Elementary and State, and Z-Scores			
Year	%	State %	z relative to state
1999	29	52	-0.46
2000	43	52	-0.18
2001	38	53	-0.30
2002	47	55	-0.16
2003	52	60	-0.16
2004	58	70	-0.26
2005	60	71	-0.24
Change 1999 to 2005			
effect size		0.22	
t-ratio		1.37	
prob.		0.09	
Min. C.I.		-0.04	

Note: Z-scores reflect the difference between the school and state values in standard-score form. D values are the effect of the change from 1999 to 2005 as a percentage of the standard deviation. T-value tests the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom = (n1 + n2) - 2 =158, where ni equals the sample size in the school for each year. Probability value is one-tail.

Table B-18

Highland Elementary School and State of Kentucky, KCCT Scores
Fourth Graders, 2005 to 2007 (Site # 13)

	% Proficient or Higher		
Year	School	State	Z from state
2005	19	60	-0.84
2006	39	58	-0.38
2007	63	53	0.2
Effect size (change 05 to 07)			1.04
t-ratio			4.4
Prob.			<.0001
Min. C.I.			0.65

Note: Z-scores reflect the difference between the school and state values in standard-score form. Effect size reflects the change from 2005 to 2007 as a percentage of the standard deviation. T-value tests the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom = (n1 + n2) - 2 =70, where ni equals the sample size in the school for fourth graders for each year. Probability value is one-tail.

Table B-19

Brunswick County Schools, Percentage of Students Scoring At or Above Proficient in Reading, End of Grade Test, 2004 to 2007 - and State Data (Site # 17)

Grade 3			
	Belville	Jessie Mae Monroe	State
2003-04	67	76	83.4
2004-05	70	83	83.4
2005-06	76	82	84.9
2006-07	87	90	83.9
Effect size change	0.53	0.36	
t-test for change	5.57	2.06	
prob.	<.0001	0.02	
Min. C.I.	0.37	0.07	

Grade 4			
	Belville	Jessie Mae Monroe	State
2003-04	69	72	83.7
2004-05	71	91	83.5
2005-06	87	84	85.4
2006-07	86	88	87.6
Effect size change	0.35	0.33	
t-test for change	3.71	1.86	
prob.	0.0001	0.03	
Min. C.I.	0.19	0.04	

Grade 5			
	Belville	Jessie Mae Monroe	State
2003-04	83	84	89.5
2004-05	82	91	90.1
2005-06	91	90	90.5
2006-07	92	87	91.7
Effect size change	0.22	0.01	
t-test for change	2.36	0.05	
prob.	0.01	0.67	
Min. C.I.	0.07	-0.28	

Note: Effect sizes (d values) are the effect of the change from 2004 to 2007 in the district, controlling for differences in the state, as a percentage of the standard deviation. T-values test the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom = 448 for Belville and 126 for JM Monroe. All probability values are one-tail.

Table B-20
Fourth Graders Scoring Proficient or Higher, Louisa May Alcott Elementary and State of Ohio, 1998 to 2006 (Site # 18)

Percentages		
	LMA School	State
1998	31	58.4
1999	37	58.4
2000	30	56.9
2001	41	54.1
2002	74	64.6
2003	86	66.3
2004	93	70.8
2005	93	76.6
2006	80	76.8

Effect Sizes and Tests of Significance		
	1998 to 2005	1998 to 2006
Effect Size	0.94	0.63
t-change	5.17	3.46
prob.	<.0001	0.0005
Min. C.I.	0.64	0.33

Note: Effect sizes (d values) are the effect of the change from 1998 to 2006 in the school, controlling for differences in the state, as a percentage of the standard deviation. T-values test the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Degrees of freedom = 58. All probability values are one-tail. State data were not found for 1998, so data for 1999 were substituted.

Table B-21

Students Meeting or Exceeding State Standards, 2003 to 2006, by Ethnicity, Milton-Freewater, Oregon (Site # 19)

Percentage Students Meeting Standards by Group and Year

	District		State	
Year	Hispanic	Total	Hispanic	Total
2003 (pre-RM)	66	60	60	82
2004	73	69	63	82
2005	83	82	71	86
2006	95	95	75	87

Effect Sizes and T-Tests Regarding Changes in District Relative to State

	Z-Scores		
	Total to Total	Hispanic to Hispanic	Hispanic to Total
2003 (pre-RM)	-0.42	0.00	-0.57
2004	-0.23	0.16	-0.34
2005	-0.09	0.38	-0.12
2006	0.24	0.80	0.24
Effect Size, 2003 -06	0.65	0.80	0.81
t-tests	5.72	4.96	5.03
prob.	<.0001	<.0001	<.0001
Min. C.I.	0.47	0.53	0.55

Note: Z-scores reflect the difference between the school and state values in standard-score form. D values are the effect of the change from 2003 to 2006 as a percentage of the standard deviation. T-values test the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Scores for the total group of third graders in the school are compared to the total group of third graders in the state. Scores for Latino third graders are compared to scores for Latino third graders and for all third graders in the state. Degrees of freedom for the comparison for the total sample are 303, those for the Latino children are 152. All probability values are one-tail reflecting the directional hypothesis.

Table B-22

Percentage Fourth Grade Students Meeting or Exceeding State Reading Standards,Evergreen Elementary and Washington State, 1998 – 2005 (Site # 23)

Year	School	State	Z from state	Grades had DI
1998	72	56	0.33	1
1999	83	59	0.49	1, 3
2000	78	66	0.26	1 to 3
2001	86	66	0.42	1 to 3
2002	82	66	0.35	1 to 3
2003	89	67	0.47	1 to 3
2004	94	74	0.45	1 to 3
2005	94	80	0.36	1 to 3
d (1998 to 2005)	0.03			
t-ratio	0.20			
Prob .	0.42			
Min. C.I.	-0.12			

Note: DF = 184. Analyses in the first panel reflect the cohort control group design, those in the second panel reflect the cohort control group with historical comparisons design.

Table B-23

Fourth Grade Students Scoring Proficient of Advanced in Reading, Honey Creek Elementary and State of Wisconsin, 1997-2005 (Site # 24)

Year	School	State	Z from State	Years of DI
1997	45	62	-0.35	0
1999	75	85	-0.28	1
2000	68	78	-0.24	2
2001	71	78	-0.17	3
2002	90	79	0.27	4
2003	88	80	0.20	5
2004	97	81	0.41	5
2005	90	82	0.21	5
Effect size	0.56			
t-ratio	2.87			
prob.	0.003			
Min. C.I.	0.24			

Note the data for the state in 1997 is actually what was given on the state web-site for 1998. I was not able to find 1997 data for the state. Degrees of freedom = 104. Probability is one-tail.

Table B-24

Summary of Results with Analyses Using Cohort Control Group Designs, 13 Sites and 40 Comparisons

Site #	Grade and Comparison	Year began RM	Years DI at School	effect size	z	prob.	Min. for C.I.
1	3 - reading	K	4	0.13	0.73	0.24	-0.17
1	3 - writing	K	4	0.28	1.55	0.06	-0.02
2	3	2	2	0.11	0.65	0.26	-0.17
2 (a)	4	3	2	-0.26	-1.35	0.91	-0.58
2	5	4	2	0.21	1.26	0.10	-0.06
2	6	5	2	0.09	0.52	0.30	-0.2
5	3 - proficient +	2	2	0.63	2.08	0.02	0.13
5	4 - proficient +	3	2	0	-0.02	0.51	-0.5
5	5 - proficient +	4	2	0.35	1.17	0.13	-0.14
5(b)	3 - unsatisfactory	2	2	1.04	3.45	0.001	0.55
5(b)	4 - unsatisfactory	3	2	0.59	1.94	0.03	0.09
5(b)	5 - unsatisfactory	4	2	0.27	0.9	0.19	-0.23
6	3	K	7	0.79	2.36	0.01	0.24
7	3 to 5, AA to AA	k	4	0.29	1.99	0.02	0.05
7	3 to 5, Hisp. to Hisp.	k	4	0.2	1.55	0.06	-0.01
7	3 to 5, Low Inc. to LI	k	4	0.18	1.99	0.02	0.03
7	3 to 5, Excp. to Excp.	k	4	0.27	1.46	0.07	-0.03
7	3 to 5, Cauc. to Cauc.	k	4	0.23	1.74	0.04	0.01
7	3 to 5, total to total	k	4	0.09	1.25	0.11	-0.03
7	3 to 5, AA to Total	k	4	0.34	2.33	0.01	0.1
7	3 to 5, Hisp. to Total	k	4	0.28	2.2	0.01	0.07
7	3 to 5, LI to Total	k	4	0.22	2.36	0.01	0.07
7	3 to 5, Excp. to Total	k	4	0.23	1.25	0.11	-0.07
7	3 to 5, Cauc. to Total	k	4	0.18	1.39	0.08	-0.03
10	4	K	9	0.19	2.18	0.01	0.05
11	4	K	7+	0.22	1.37	0.09	-0.04
13	4	3	2	1.04	4.40	<.0001	0.65
17	3 - sch. A	1	3	0.53	5.57	<.0001	0.37
17	4 - sch. A	2	3	0.35	3.71	0.0001	0.19
17	5 - sch. A	3	3	0.22	2.36	0.01	0.07
17	3 - sch. B	1	3	0.36	2.06	0.02	0.07

17	4 - sch. B	2	3	0.33	1.86	0.03	0.04
17	5 - sch. B	3	3	0.01	0.05	0.67	-0.28
18 (a)	4	k	8	0.63	3.46	0.0005	0.33
18	4 (alternate years)	k	7	0.94	5.17	<.0001	0.64
19	3 - Total to total	1	3	0.65	5.72	<.0001	0.47
19	3 - Hisp. To Hisp.	1	3	0.80	4.96	<.0001	0.53
19	3 - Hisp. To Total	1	3	0.81	5.03	<.0001	0.55
23	4	K	7	0.03	0.20	0.42	-0.12
24	4	K	7	0.56	2.87	0.0025	0.24
Min.				-0.26			-0.58
Max.				1.04			0.65
Aver.				0.36			0.07

Site-Level Summaries

Site #	Average effect size	Aver. Min. for C.I.
1	0.21	-0.09
2	0.04	-0.25
5	0.48	-0.02
6	0.79	0.24
7	0.23	0.01
10	0.19	0.05
11	0.22	-0.04
13	1.04	0.65
17	0.30	0.08
18	0.79	0.49
19	0.75	0.52
23	0.03	-0.12
24	0.56	0.24
Min.	0.03	-0.25
Max.	1.04	0.65
Aver.	0.43	0.13

Note: Comparisons designated with "a" in the site column are those for which the SRA report indicated a non-comparable cohort. Statistics associated with comparisons designated with a b were reversed to preserve the positive sign associated with findings in favor of the DI programs.

Table B-25

Percentage of Second Grade Students Proficient or Advanced in Reading, on California Standards Test, 2001 to 2005, total Group and Economically Disadvantaged, Richfield Elementary and State of California (Site # 4)

Percentage Students Meeting Standards by Group and Year

	Richfield Elementary		State	
Year	Disadvantaged	All	Disadvantaged	All
2001	20	43	17	32
2002	12	28	18	32
2003	25	50	23	36
2004	58	65	22	35
2005	41	46	28	42

Z-Scores Comparing School and State Scores

	Total to Total	Disadvantaged to Disadvantaged	Disadvantaged to Total
2001	0.29	0.06	-0.32
2002	-0.10	-0.13	-0.52
2003	0.33	0.04	-0.26
2004	0.72	0.75	0.56
2005	0.09	0.26	-0.02

Effect Sizes and T-Tests Regarding Changes in District Relative to State

	Total to Total	Disadvantaged to Disadvantaged	Disadvantaged to Total
2001 to 2004			
Effect Size	0.43	0.69	0.87
t-test	1.49	1.62	2.05
prob.	0.07	0.06	0.03
Min.C.I.	-0.04	-0.01	0.17
2001 to 2005			
Effect Size	-0.20	0.20	0.30
t-test	-0.71	0.47	0.70
prob.	0.75	0.32	0.25
Min. C.I.	-0.68	-0.50	-0.40

Note: Z-scores reflect the difference between the school and state values in standard-score form. D values are the effect of the change from 2001 to 2004 and 2001 to 2005 as a percentage of the standard deviation. T-values test the hypothesis that increase in scores over time in the school was greater than in the state as a whole. Scores for the total group of second graders in the school are compared to the total group of third graders in the state. Scores for disadvantaged second graders are compared to scores for disadvantaged second graders and for all second graders in the state. Degrees of freedom for the comparison for the total sample are 46, those for the disadvantaged children are 20. All probability values are one-tail reflecting the directional hypothesis.

Table B-26

Students Scoring at or Above Proficient, Utah Criterion-Referenced Test, Cache County and State, 2004-2005 (Site # 22)

2004					
Grade	County	State	d	z	prob.
1	91	76	0.35	11.14	<.0001
2	90	77	0.32	10.02	<.0001
3	90	75	0.35	10.93	<.0001
4	81	75	0.13	4.17	<.0001
5	87	75	0.28	8.74	<.0001
2005					
Grade	County	State	d	z	prob.
1	90	75	0.35	10.94	<.0001
2	93	78	0.37	11.64	<.0001
3	92	75	0.39	12.39	<.0001
4	82	77	0.12	3.84	0.0001
5	89	76	0.31	9.77	<.0001

Note: An estimated class size of 900 students was used to calculate the standard errors. Effect sizes compare the district value to that for the state. Z-values test the null hypothesis that the district value equals that for the state.

Table B-27

DIBELS Risk Status, 2002-03 K Cohort, Cache County District (Site # 22)

At Risk - Middle of School year					
Grade	Cache County	National Data	d	Z	prob.
K (2002-2003	21	34	0.27	8.67	<.0001
1 (2003-2004)	1	12	0.34	10.70	<.0001
2 (2004-2005)	7	26	0.43	13.69	<.0001
3 (2005-2006)	7	19	0.31	9.67	<.0001
Established or Low Risk - End of School Year					
Grade	Cache County	National Data	d	Z	prob.
K (2002-2003	65	67	-0.04	-1.34	0.91
1 (2003-2004)	95	65	0.63	19.88	<.0001
2 (2004-2005)	79	61	0.37	11.66	<.0001
3 (2005-2006)	73	60	0.27	8.39	<.0001

Note: Phoneme segmentation fluency was used as the measure in kindergarten, and oral reading fluency in later grades. National data taken from Good, et al, 2002. Signs of effect sizes and z-scores have been changed to positive values for the "at risk" analyses to reflect the research hypothesis that students with the DI instruction would have higher achievement.

Table B-28

Changes in DIBELS Scores, Edgewood Academy, 2004-05 (before implementation) and 2006-07 after implementation (Site # 9)

DIBELS Scores, Fall and Mid-Year, by Year and Grade					
	2004-05		2006-07		
Grade	Fall	Mid-year	Fall	Mid-Year	
1	85	58	37	43	
2	58	41	60	64	
3	35	38	39	44	
4	51	42	46	61	
5	64	47	58	52	
Differences, Effect Sizes, and Inferential Tests					
Grade	Dif. 04-05	Dif. 06-07	d	Z	prob.
1	-27	6	0.72	3.58	0.0002
2	-17	4	0.43	2.15	0.02
3	3	5	0.04	0.21	0.42
4	-9	15	0.48	2.42	0.01
5	-17	-6	0.22	1.12	0.13

Note: Estimated sample size for each year is 100 per class.

Table B-29

Iredell Statesville Schools, Percent Attaining Adequate Yearly Progress, 2003-2006, by Special Education Status, District and State (Site # 16)

Percentage Students Attaining Adequate Yearly Progress				
	District		State	
Year	Students with Disabilities	All Students	Students with Disabilities	All Students
2002-03	43	84	47.8	82.6
2003-04	53	86	53.5	84.3
2004-05	64	88	55.3	84.7
2005-06	66	89	56.2	84.9

Effect Sizes and Inferential Tests			
	Dis. Versus State Dis.	Dis. Versus State Total	All students Versus State Total
Effect Size	0.29	0.52	0.08
t-test	6.23	10.96	1.65
prob.	<.0001	<.0001	0.05
Min. C.I.	0.22	0.44	0.00003

Note: Sample size estimated at 9000 for total students and 900 for students with disabilities.

Table B-30

Third Grade Students Scoring Proficient of Advanced in Reading, WRCT, Northside and Parkside elementary, and State of Wisconsin, 2000-2005 (Site # 25)

Year	Northside	Parkside	State	Z from State (Northside)	Z from State (Parkside
2000	85	82	74	0.25	0.18
2001	93	93	77	0.38	0.38
2002	83	83	74	0.21	0.21
2003	87	83	81	0.15	0.05
2004	95	96	85	0.28	0.31
2005	100	100	97	0.18	0.18
Effect size	-0.07	-0.01			
t-ratio	-0.37	-0.03			
prob.	0.64	0.51			
Min. C.I.	-0.41	-0.33			

Note: Degrees of freedom = 94 for Northside and 100 for Parkside. Probabilities are one-tail.

Table B-31

Performance of Special Education Students on State Reading Assessment, Marmaton Valley Elementary School and State of Kansas, 2007 (Site # 12)

Score	School (n)	State (%)	Expected n – School
Warning	0	14.2	2.3
Approached	3	16.0	2.6
Met	7	35.3	5.6
Exceeded	6	20.1	3.2
Exemplary	0	14.3	2.3
	16	100	16

	Met + vs. Others	Exceeded vs. Others
Chi-square	1.00	0.07
prob.	0.32	0.79

Note: Degrees of freedom equaled one for both comparisons.

Table B-32

Growth on Measures of Academic Progress, One Year, Special Education Students and All Students, Clover School District, and National Data, by Grade (Site # 20)

	Average Growth			Ratios	
Grade	SPED Students - Clover SD	All Students - Clover SD	Nation	SPED/Total - Clover SD	SPED - Clover SD/Nation
2	12.3	12.0	9.9	1.02	1.24
3	5.5	7.0	7.4	0.79	0.74
4	5.7	6.0	5.7	0.95	0.99
5	9.6	6.0	4.4	1.60	2.18
6	9.2	4.0	3.2	2.29	2.86
7	7.5	3.0	2.5	2.49	2.99
8	2.4	2.0	2.2	1.22	1.10

Table B-33

Average Gain, Woodcock-Johnson III, Struggling Readers, Grades 9-12, East High School, 2003-04 to 2005-06 (Site # 26)

School Year	Number of Students	Average Gain
2003-2004	13	1 year, 6 months
2004-2005	20	9 months
2005-2006	7	1 year, 3 months
Weighted Average		1 year, 1 month

Promoting Reading Achievement and Countering the "Fourth Grade Slump": The Impact of Direct Instruction on Reading Achievement in Fifth Grade

Methodological Appendix

The data set available for analysis had information on NCE and raw scale CTBS scores for first graders and fifth graders in the BCPSS for 1998 through 2003, the schools that the students attended in each year, and a separate file with school characteristics. No information was provided for grades two through four, nor, as noted in the text, was individual socio-demographic information. It was possible, however, to link identification numbers for students between the years, and that was the basis for forming the data set used for analysis in the paper. From this linking process it was possible to determine which children were in the BCPSS in 1998 or 1999 and again in 2002 or 2003. It was also possible to determine which children made expected progress (in first grade in 1998 and in 5th grade in 2002 or in first grade in 1999 and in 5th grade in 2003) and which children had been in first grade in 1998 but were not in fifth grade until 2003, thus apparently repeating a grade in the interim. Given the limitations of the data set it was not possible to determine which of the students in first grade in 1998 had previously repeated the grade, nor was it possible to determine which of the students in first grade in 1999 repeated a grade that year. Thus, our information on students who were retained was limited.

Table A-1 gives the number of students in the total study population by year and group. The left columns of the table give the raw numbers and the right columns give the percentages. As explained in the text, students in the control group who were in first grade in 1998 (the first three rows of data) had a mixed curriculum in first grade, while those in first grade in 1999 had Open Court. The last two rows of the table give the number of students who could be assumed to have been retained. The figures were slightly smaller in the DI schools. The percentages in Table

A-1 use the total number of students (starting first grade in 1998 or 1999) in each group as a base. When the number of students in 1998 is used as a base, the results indicate that the following percentages: for DI schools, 6.1% in different schools and 1.1 in the same schools and for the control schools, 6.5% in different schools and 1.4% in the same schools. Again, these percentages indicate students who were in first grade in 1998 and in fifth grade in 2003, six years later, suggesting that they had repeated a grade in the interim.

Table A-2 reports the average raw scale and NCE scores for the CTBS reading comprehension and vocabulary tests in first and fifth grade for students in both the DI and control schools. Results are disaggregated by year and by whether or not a child remained in the BCPSS and in the same school in the first and fifth grade samples. Thus, the data in this table allow one to compare scores of students used in our analysis (those who remained in the same school throughout the elementary career in the BCPSS) and other students. As explained in the text, students who remained in the same schools from first to fifth grade had higher achievement scores than other students.

The bottom two lines of each section of Table A-2 compare the scores of students who were determined to have been retained and the average of all other students. Because the sample size was so small if those who had transferred to other schools and those who remained in the same schools were distinguished (see Table A-2), these two groups were collapsed for this analysis. In all comparisons the scores of the students who were retained were lower than those of students who were not. As noted in the text, the students who were retained were not included in the analyses.

Finally, Table A-3 reports the average percentage of students receiving free or reduced lunch by group.

Table A-1

Population and Sample Size by Year and Curriculum

	First Grade Curriculum			
	DI (N)	Control (N)	DI (%)	Control (%)
1st grade 1998, no 5th grade data	283	3641	16.6	23.2
1st grade 1998, in BCPSS in 5th grade, different school	151	1840	8.9	11.7
1st grade 1998, in BCPSS in 5th grade, same school	256	2197	15.0	14.0
1st grade 1999, no 5th grade data	542	4168	31.8	26.6
1st grade 1999, in BCPSS in 5th grade, different school	174	1303	10.2	8.3
1st grade 1999, in BCPSS in 5th grade, same school	244	1875	14.3	12.0
Retained (1st in 1998 and 5th in 2003), different schools	45	544	2.6	3.5
Retained (1st in 1998 and 5th in 2003), same schools	8	116	0.5	0.7
Totals	1703	15684	100.0	100.0

Note: These figures omit students who were in schools that began DI after they were in first grade (n = 341). Control Students in first grade in 1998 had a mixed reading curriculum; control students in first grade in 1999 had Open Court.

Table A-2

Scale Scores and NCE Scores by Grade and Group

First Grade Reading Vocabulary

	Scale Scores						NCE Scores					
	DI			Control			DI			Control		
	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med
1st 1998, left BCPSS	518	69	516	529	71	524	30	21	27	34	23	31
1st 1998, transferred	536	65	538	554	69	552	36	21	36	42	22	41
1st 1998, same school	547	68	545	571	65	572	39	22	38	47	22	48
1st 1999, left BCPSS	533	65	531	542	73	538	34	21	32	37	23	35
1st 1999, transferred	575	60	572	579	68	586	48	20	47	49	22	52
1st 1999, same school	577	61	579	595	61	594	48	21	49	54	21	54
Not Retained	544	68	545	555	72	552	38	22	37	42	24	41
Retained	514	49	516	514	59	516	27	16	27	28	18	27

First Grade Reading Comprehension

	Scale Scores						NCE Scores					
	DI			Control			DI			Control		
	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med
1st 1998, left BCPSS	496	95	504	503	98	515	29	19	24	31	20	28
1st 1998, transferred	515	99	525	535	89	544	34	19	31	38	20	38
1st 1998, same school	541	80	544	557	80	561	39	20	38	43	20	44
1st 1999, left BCPSS	515	90	525	513	100	525	32	19	30	32	21	30
1st 1999, transferred	562	73	569	556	85	569	44	18	46	43	20	46
1st 1999, same school	569	78	578	574	80	578	46	19	48	47	20	48
Not Retained	529	90	544	531	95	544	36	20	36	37	21	36
Retained	485	80	504	488	88	504	25	15	24	26	16	24

Fifth Grade Reading Vocabulary

	Scale Scores						NCE Scores					
	DI			Control			DI			Control		
	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med
1st 1998, transferred	634	36	636	633	35	636	44	18	43	43	17	43
1st 1998, same school	646	40	644	644	35	644	50	21	48	49	18	48

1st 1999, transferred	640	35	644	640	35	644	47	17	48	47	17	48
1st 1999, same school	654	36	652	647	32	652	54	19	53	50	17	53
Not Retained	645	38	644	642	35	644	49	18	48	48	17	48
Retained	608	54	627	632	36	636	33	20	39	42	17	43

Fifth Grade Reading Comprehension

	Scale Scores						NCE Scores					
	DI			Control			DI			Control		
	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med	M	s.d.	Med
1st 1998, transferred	640	38	642	641	38	645	44	19	44	45	18	45
1st 1998, same school	654	38	657	651	34	651	51	19	52	49	17	49
1st 1999, transferred	646	32	648	643	36	645	47	16	47	45	17	45
1st 1999, same school	656	36	657	652	31	654	52	17	52	50	16	50
Not Retained	650	37	654	647	35	651	49	18	50	48	17	49
Retained	632	25	627	626	34	630	39	13	36	37	16	37

Note: The "not retained" group includes all other students within the system but, as noted in the text, may include children who were retained in earlier years. N's for each sub-group are given in Table A-1.

Table A-3

Average Percentage of Children Receiving Free or Reduced Lunch by Group

	DI			Control		
	M.	s.d.	Med.	M.	s.d.	Med.
1st 1998, left BCPSS	78.2	13.6	82.2	74.7	14.0	76.7
1st 1998, transferred	80.7	9.7	84.4	77.5	10.3	78.5
1st 1998, same school	75.5	17.0	82.2	71.0	16.1	74.7
1st 1999, left BCPSS	80.9	10.3	82.8	74.5	14.1	76.7
1st 1999, transferred	81.8	9.7	84.4	76.5	11.1	77.8
1st 1999, same school	74.8	17.9	82.2	71.0	15.4	74.7
Not Retained	77.5	14.3	82.2	73.9	14.0	77.8
Retained	83.7	4.3	83.7	70.6	16.0	73.9
Total	78.8	13.5	82.8	74.1	14.1	76.3

Issue ID 190: Jean Stockard

Dear Dr. Dynarski:

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction (NIFDI). One of my first tasks was to read as much of the background research in the area as I could. Along the way I ran into the "What Works Clearinghouse." In my review I developed several questions that I am hoping you can answer.

First, I noticed that the WWC has listed relatively few of the studies that deal with Direct Instruction. While a few were rejected because they didn't meet some specific standards of design, it appears that many others were rejected because they were published before 1985. I have not been able to find a scientific justification for selecting this date. No studies were cited that suggested children's learning styles had altered after that time or that schools' organizational structures or teachers' actions had changed. In fact, the only research article that I found to justify the methods was actually published in 1974 – 10 years prior to the beginning of your targeted dates. This article dealt with phonological awareness, and had nothing to do with systematic literature reviews and, of course, included nothing to suggest that processes related to phonological awareness have changed over time.

In all the other areas with which I am familiar we use the full range of research articles and, unless there are clear reasons established through the research literature, do not set an arbitrary date for inclusion or exclusion of results. The choice of this date seems to have particularly affected the DI literature because a great deal of the pioneering work was published before the cut-off point. Because I'm new to this particular work, I wondered if you could provide me with any explanation based in the research literature for your choice of this cut-off year. I did find a suggestion that, "if sufficient time and resources remain," studies published before 1985 might be reviewed. Do you know if this step will be taken? From my reading of the literature I believe that your limit on dates of studies may have seriously limited the selection of effective resources that educators will find on your website; and I know that this is a primary aim of your project.

Second, I noticed that a number of studies were rejected because they included children outside the K-3 grade range. I found this somewhat unusual, for the children that are most in need of effective instruction are those who have not learned to read by older ages. Because the effectiveness of DI has been so well demonstrated with regular students at the K-3 grade range, the more recent focus on other ages seems only greater evidence of its worth. This is especially so in the case of special education students and those with behavioral and other issues. Thus, I couldn't understand the rationale of omitting all the studies that targeted these most needy students. Any explanation you might have of this decision would also be greatly appreciated.

Third, and perhaps most important, as I read studies more closely I found that the Clearinghouse's interpretation of some of the studies didn't seem to accurately reflect the

actual content of the reports. I know that these reviews are often done by staff members or graduate students who might not be familiar with all of the intricacies of social research. Thus, I assume that we can send corrections and they will be added.

I look forward to hearing from you, either through e-mail (jeans@uoregon.edu) or by phone (541-346-5005). I'm sure that we share the ultimate goal of helping all students – no matter what their background – develop academic skills and appreciate your appreciate your help.

Sincerely,

Jean Stockard, Professor Emerita

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated December 2, 2007 below.

The first question you raised concerned the cut-off year for studies that are reviewed by the WWC. To maintain a focus on current research, our reviews examine research conducted in the past 20 years. This cut-off date of 20 years is the standard for every topic reviewed by the WWC. The review for Early Childhood Education (ECE), under which Direct Instruction falls, marked its cut-off date at 1985 (the review began in 2005).

The ECE protocol states that "if sufficient time and resources remain after we have completed our review of research on interventions implemented post 1985, the ECE team will consider reviewing older research on curricula that are still in widespread use." This determination has not been made to date.

In terms of the age/grade range for the review of Direct Instruction research, the ECE team determined the parameters for the review. For further details on inclusion criteria please see the ECE protocol at http://ies.ed.gov/ncee/wwc/reports/early_ed/index.asp

In response to your third question about the qualifications of WWC topic review team members, each team is comprised of several trained professionals, each playing an integral part in the review process. The Principal Investigator (PI) for each topic is a well-known expert in his/her field and is responsible for leadership in conceptualizing the specific topic area, identifying and addressing issues during the review, and developing and reviewing topic and intervention reports developed for the topic. Leadership includes overseeing the quality in the production of the reports and making decisions, based on methodological and substantive expertise, that are not otherwise covered in the WWC protocols and procedures for report production.

The Project Coordinator (PC) is an established education researcher with relevant methodological and substantive expertise. The coordinator oversees the work of the WWC Review Team; manages that specific review; reviews research ratings; and writes and revises the work plan, protocol, and draft and final reports in collaboration with the PI.

Individual reviewers, who prepare initial summaries of studies for the WWC, are professional researchers with experience in research design and methodology. These reviewers undergo a rigorous training and certification process before conducting WWC reviews. .For more information about the staff that make up the WWC team, visit http://ies.ed.gov/ncee/wwc/overview/index.asp.

The WWC recently created a Quality Review Team to respond to concerns raised by study authors, curriculum developers or other relevant parties about WWC reviews published on our website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. If a quality review concludes that the original review was flawed, a revision will be published. These quality reviews are one of tools used to ensure that the standards established by the Institute of Educational Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.

If you have concerns about a published WWC review that you think warrant a quality review, please send those concerns to our Help Desk at info@whatworks.ed.gov. Please identify the study in question, the specify issue(s) that you think were handled incorrectly, and where relevant, explain what you think is the correct interpretation of the study.

I hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: Jean Stockard [jeans@uoregon.edu]
Sent: Sunday, March 16, 2008 11:59 AM
To: WhatWorks
Subject: re: Questions on What Works Clearing House Procedures
Thank you for your reply of February 8 to my query of December 2. I am pleased to hear that you have established a Quality Review Team. Since my initial query I have continued to examine the reviews and have found quite a few that appear very problematic. I, and my colleagues, will be submitting our concerns to you within the coming weeks.

Unfortunately, I' m afraid that your reply failed to address two of my concerns. Your responses simply restated WWC' s policies regarding the cut-off year for studies and the age-grade parameters without providing the requested justification. Again, I would be very interested in knowing the scientific and pedagogical justifications for not incorporating the entire corpus of work in your reviews and for limiting your reviews of studies to only those that included children in grades K-3.

My earlier, rather lengthy, explanation of my concerns is included in this e-mail. If you wish me to clarify any part of my query, please let me know. Thank you for your attention to this matter.

Sincerely,

Jean Stockard, Ph.D.
Emerita Professor

At 11:44 PM 2/8/2008, you wrote:

-->

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated December 2, 2007 below.

The first question you raised concerned the cut-off year for studies that are reviewed by the WWC. To maintain a focus on current research, our reviews examine research conducted in the past 20 years. This cut-off date of 20 years is the standard for every topic reviewed by the WWC. The review for Early Childhood Education (ECE), under which Direct Instruction falls, marked its cut-off date at 1985 (the review began in 2005).

The ECE protocol states that " if sufficient time and resources remain after we have completed our review of research on interventions implemented post 1985, the ECE team will consider reviewing older research on curricula that are still in widespread use." This determination has not been made to date.

In terms of the age/grade range for the review of Direct Instruction research, the ECE team determined the parameters for the review. For further details on inclusion criteria please see the ECE protocol at http://ies.ed.gov/ncee/wwc/reports/early_ed/index.asp

In response to your third question about the qualifications of WWC topic review team members, each team is comprised of several trained professionals, each playing an integral part in the review process. The Principal Investigator (PI) for each topic is a well-known expert in his/her field and is responsible for leadership in conceptualizing the specific topic area, identifying and addressing issues during the review, and developing and reviewing topic and intervention reports developed for the topic. Leadership includes overseeing the quality in the production of the reports and making decisions, based on methodological and substantive expertise, that are not otherwise covered in the WWC protocols and procedures for report production.

The Project Coordinator (PC) is an established education researcher with relevant methodological and substantive expertise. The coordinator oversees the work of the WWC Review Team; manages that specific review; reviews research ratings; and writes and revises the work plan, protocol, and draft and final reports in collaboration with the PI.

Individual reviewers, who prepare initial summaries of studies for the WWC, are professional researchers with experience in research design and methodology. These reviewers undergo a rigorous training and certification process before conducting WWC reviews. .For more information about the staff that make up the WWC team, visit http://ies.ed.gov/ncee/wwc/overview/index.asp.

The WWC recently created a Quality Review Team to respond to concerns raised by study authors, curriculum developers or other relevant parties about WWC reviews published on our website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. If a quality review concludes that the original review was flawed, a

revision will be published. These quality reviews are one of tools used to ensure that the standards established by the Institute of Educational Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.

If you have concerns about a published WWC review that you think warrant a quality review, please send those concerns to our Help Desk at info@whatworks.ed.gov. Please identify the study in question, the specific issue(s) that you think were handled incorrectly, and where relevant, explain what you think is the correct interpretation of the study.

I hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: Jean Stockard
[
mailto:jeans@uoregon.edu]

Sent: Sunday, December 02, 2007 12:30 PM

To: Mark Dynarski

Subject: Questions on What Works Clearing House
Procedures

Dear Dr. Dynarski:

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction (NIFDI). One of my first tasks was to read as much of the background research in the area as I could. Along the way I ran into the " What Works Clearinghouse." In my review I developed several questions that I am hoping

you can answer.

First, I noticed that the WWC has listed relatively few of the studies that deal with Direct Instruction. While a few were rejected because they didn't meet some specific standards of design, it appears that many others were rejected because they were published before 1985. I have not been able to find a scientific justification for selecting this date. No studies were cited that suggested children's learning styles had altered after that time or that schools' organizational structures or teachers' actions had changed. In fact, the only research article that I found to justify the methods was actually published in 1974 – 10 years prior to the beginning of your targeted dates. This article dealt with phonological awareness, and had nothing to do with systematic literature reviews and, of course, included nothing to suggest that processes related to phonological awareness have changed over time.

In all the other areas with which I am familiar we use the full range of research articles and, unless there are clear reasons established through the research literature, do not set an arbitrary date for inclusion or exclusion of results. The choice of this date seems to have particularly affected the DI literature because a great deal of the pioneering work was published before the cut-off point. Because I'm new to this particular work, I wondered if you could provide me with any explanation based in the research literature for your choice of this cut-off year. I did find a suggestion that, " if sufficient time and resources remain," studies published before 1985 might be reviewed. Do you know if this step will be taken? From my reading of the literature I believe that your limit on dates of studies may have seriously limited the selection of effective resources that educators will find on your website; and I know that this is a primary aim of your project.

Second, I noticed that a number of studies were rejected because they included children outside the K-3 grade range. I found this somewhat unusual, for the children that are most in need of effective instruction are those who have not learned to read by older ages. Because the effectiveness of DI has been so well demonstrated with regular students at the K-3 grade range, the more recent focus on other ages seems only greater evidence of its worth. This is especially so in the case of special education students and those with behavioral and other issues. Thus, I couldn't understand the rationale of omitting all the studies that targeted these most needy students. Any explanation you might have of this decision would also be greatly appreciated.

Third, and perhaps most important, as I read studies more closely I found that the Clearinghouse's interpretation of some of the studies didn't seem to accurately reflect the actual content of the reports. I know that these reviews are often done by staff members or graduate students who might not be familiar with all of the intricacies of social research. Thus, I assume that we can send corrections and they will be added.

I look forward to hearing from you, either through e-mail (jeans@uoregon.edu) or by phone (541-346-5005). I'm sure that we share the ultimate goal of helping all students – no matter what their background – develop academic skills and appreciate your appreciate your help.

Sincerely,
Jean Stockard, Professor Emerita

NOTICE:

This message is for the designated recipient only and may contain privileged or confidential information. If you have received it in error, please notify the sender immediately and delete the original. Any other use of this e-mail by you is prohibited.

Jean Stockard, Professor Emerita
Department of Planning, Public Policy, and Management
University of Oregon
Eugene, Oregon 97403-1209
phone: 541-346-5005
fax: 541-346-2040

From: Mary Grider
Sent: Friday, January 16, 2009 6:43 PM
To: Mary Grider
Subject: FW: complaint about WWC
Attachments: 090102 Hanushek Stockard.pdf; rmresponse_wwc.pdf

Importance: High

From: Scott Cody
Sent: Friday, January 16, 2009 9:26 AM
To: Mary Grider
Subject: FW: complaint about WWC
Importance: High

From: Sanchez, Susan [mailto:Susan.Sanchez@ed.gov]
Sent: Friday, January 16, 2009 9:22 AM
To: Mark Dynarski; Jill Constantine; Scott Cody; Michael Ponza
Subject: FW: complaint about WWC
Importance: High

-----Original Message-----
From: Cottingham, Phoebe
Sent: Thursday, January 15, 2009 6:42 PM
To: Sanchez, Susan; 'mdynarski@mathematica-mpr.com';
'jconstantine@mathematica-mpr.com'
Subject: Fw: complaint about WWC
Fyi

Sent using BlackBerry

From: (b)(6)
To: Cottingham, Phoebe
Cc: Betka, Sue; Garza, Norma
Sent: Thu Jan 15 18:11:06 2009
Subject: complaint about WWC
Phoebe

Interestingly, while in DC for the discussion of WWC, I received the attached letter and
paper. Can you tell me what you know about this matter? They claim not to have been
properly treated.

Rick

--
Eric Hanushek
Hoover Institution
Stanford University
Stanford, CA 94305-6010

650-736-0942
www.hanushek.net

Promoting Reading Achievement and Countering the "Fourth Grade Slump": The Impact of Direct Instruction on Reading Achievement in Fifth Grade

Jean Stockard

Director of Research

National Institute for Direct Instruction

And

Professor Emerita

University of Oregon

August 26, 2009

Running Head: DI and the fourth grade slump

Reprint requests should be sent to Jean Stockard, National Institute for Direct Instruction, P. O. Box 11248, Eugene, OR 97440; phone: 1 877-485-1973; Fax: 1 541-683-7543; e-mail jstockard@nifdi.org

The author thanks Cristy Coughlin, Kurt Engelmann, Zigfried Engelmann, Daniel Johnston, Jerry Silbert, and Gary Davis for very helpful comments on earlier drafts of this paper. Any errors remaining are the sole responsibility of the author.

Promoting Reading Achievement and Countering the "Fourth Grade Slump": The Impact of Direct Instruction on Reading Achievement in Fifth Grade

Abstract

Previous research has documented a substantial decline of standardized test scores of children from low-income backgrounds relative to more advantaged peers in later elementary grades, the so-called "fourth grade slump." This paper examines changes in reading achievement from first to fifth grade for students in a large urban school system with a high proportion of students from economically deprived backgrounds. Students received first grade reading instruction from Direct Instruction (DI), Open Court, or a mixture of reading curricula. Results indicate that students in schools using DI had significantly greater gains in both reading vocabulary and comprehension than students in the two other settings and that their average levels of achievement in fifth grade were above the national norms, thus countering the "fourth grade slump."

Keywords: reading achievement, fourth grade slump, Direct Instruction, Open Court

Promoting Reading Achievement and Countering the "Fourth Grade Slump": The Impact of Direct Instruction on Reading Achievement in Fifth Grade

An extensive social science literature, spanning at least four decades, has documented the relationship between academic achievement, educational aspirations, eventual educational attainment, and adult occupational and economic success. Young people who do poorly in school are less likely to finish high school or attend college and have lower occupational status and lower incomes as adults (e.g. Blau & Duncan, 1967, Farkas, 1996, Jencks, Crouse, & Mueser, 1983, Sewell, Haller, & Ohlendorf 1970). A slightly smaller, but still convincing, literature links early academic problems to later academic achievement, indicating that children with academic problems early in their school career are much more likely to also exhibit problems at later years. (e.g. Juel, 1988). Reading is generally cited as the most important and central skill in promoting academic achievement, primarily because learning to read is basic to acquiring other skills and proceeding successfully through a schooling career (Murphy, 2004).

These associations carry special weight for children from low-income backgrounds, who often enter school with substantially lower levels of pre-reading skills or "readiness" to achieve. Moreover, even when low income children approach the achievement levels of their more advantaged peers in the early grades, the differences often widen in the later elementary period. This has come to be known as a "fourth-grade slump," a substantial decline of standardized test scores of children from low-income backgrounds relative to their more economically advantaged peers as they progress past third grade and into the later

elementary years (Chall & Jacobs, 2003, Chall, Jacobs, & Baldwin, 1990, Hirsch, 2003, Rosenshine, 2002). In other words, while learning to read is important for later success for all children, those from economically deprived backgrounds appear to face special challenges in maintaining higher levels of academic achievement that can translate into adult occupational and economic success.

Reflecting the centrality of reading to long-term success, a large and growing body of literature has examined the impact of various curricula on reading achievement in the early elementary grades. Additional work has looked at their impact among different populations of older children and young adults. Surprisingly, however, there appears to be far less attention to the longer term impact of various curricula on children's reading achievement, especially the changes in achievement across time periods that span the time of the "fourth grade slump." This paper addresses that gap in the literature by examining changes in reading achievement from first to fifth grade for students in a large urban school system with a high proportion of students from economically deprived backgrounds. We compare the achievement of students who received first grade reading instruction in three different settings: 1) schools that used Direct Instruction (DI), a curriculum that directly addresses areas seen as crucial in counteracting the "fourth grade slump;" 2) schools using Open Court, another highly rated curriculum; and 3) schools using a variety of reading curricula. Our results indicate that students in schools with the Direct Instruction curriculum had significantly greater gains in both reading vocabulary and comprehension than

students in the two other settings and that their average levels of achievement in fifth grade were above the national norms, thus countering the fourth grade slump.

Background

A large body of literature, spanning several decades, has examined curricular and school reform models that can enhance student achievement. This literature has documented the importance of systematic and explicit instruction in promoting reading achievement (Adams, 1990, Anderson et al, 1985, Baker, Kameenui, Simmons, & Stahl, 1994, Bond & Dykstra, 1967, Chall, 1967, Foorman, 1995, Fukkink & deGlopper, 1998, Grossen, 1997, ICHHD, 2000, Juel and Minden-Cupp, 2000, Murphy, 2004, National Reading Panel, 2000, NICHHD, 1996, Pflaum, Walberg, Karigianes, & Rasher, 1980, Smith et al, 2001, Snider, 1990, Snow, Burns, & Griffin, 1998, Stanovich, 1994). Meta-analyses that examine specific curricula support this conclusion, showing that programs that embody these elements consistently result in larger achievement gains (Adams & Engelmann, 1996, AFT, 1998, Beck & McCaslin, 1978, Borman, Hewes, Overman, & Brown, 2003, Herman, et al, 1999).

The Effectiveness of Direct Instruction

One of the most prominent explicit instructional approaches is Direct Instruction (distinguished from other "direct instruction" approaches by the use of capital letters), which was developed by Siegfried Engelmann, Wesley Becker, and their colleagues (Engelmann & Carnine, 1982, Engelmann, 2007). Direct Instruction curricula are specifically designed to accelerate students' learning by teaching more than traditional programs in the same amount of time. Unlike many

curricula, the Direct Instruction programs are extensively field tested before dissemination to ensure that they produce the greatest learning in the most efficient manner. The programs, which are commercially available through SRA/McGraw Hill and Sopris West, involve scripted lessons designed to provide teachers with the most effective wording to allow them to present tasks to students at a relatively high rate. The amount of new material introduced in each lesson is carefully controlled, with applications becoming increasingly complex and designed so that, at the end of each lesson, all children will have mastered all the content in the current lesson. The content of the lessons is also carefully designed to provide the basis for continued academic growth and understanding. Analyses of the DI curriculum suggest that, unlike traditional teaching methods, including those often termed "direct," the DI approach teaches an underlying order of knowledge and provides the basis for accelerated cognitive growth (Carnine, Grossen, & Silbert, 1992).

Numerous studies have documented the superiority of Direct Instruction in promoting reading achievement. These results have appeared with the general student population (e.g. Becker & Carnine, 1980; O'Brien & Ware, 2002; Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977; Vitale & Joseph, 2008) and with students with disabilities. Studies have demonstrated the effectiveness of Direct Instruction reading programs with students with learning disabilities (Benner, 2007; Benner, Kinder, Beaudoin, Stein, & Hirschmann, 2005; Cooke, Gibbs, Campbell, & Shalvis, 2004; Kuder, 1990; Kuder, 1991; Malmgren & Leone, 2000; Scarlato & Asahara, 2004), students with intellectual disabilities

(Flores, Shippen, Alberto, & Crowe, 2004; Haring & Krug, 1975; Maggs & Morath, 1976; Malmgren & Leone, 2000; Riepl, Marchand-Martella, & Martella, 2008), children who demonstrate developmental delays (Flores & Ganz, 2007; Riepl, Marchand-Martella, & Martella, 2008), and students identified with emotional disturbance (Benner, 2007; Cooke, Gibbs, Campbell, & Shalvis, 2004; Malmgren & Leone, 2000; Scarlato & Asahara, 2004; Strong, Wehby, Falk, & Lane, 2004). Additionally, Direct Instruction reading programs have been effectively implemented outside of traditional elementary schools in a variety of settings, including middle schools (Dowdell, 1996; Grossen, 2004; Lewis, 1982; Shippen, Houchins, Steventon, & Sartow, 2006), high schools (Harris, Marchand-Martella, & Martella, 2000; Marchand-Martella, Martella, Orlob, & Ebey, 2000), a residential treatment center (Scarlato & Asahara, 2004), alternative schools (Steventon & Fredrick, 2003), and juvenile corrections facilities (Drakeford, 2002; Houchins, Jolivette, Krezmien, & Baltodano, 2008; Malmgren & Leone, 2000;). Lastly, Direct Instruction programs have been shown to be effective in increasing reading achievement with English Language Learners (Grossen, 2004; Gunn, Smolkowski, Biglan, & Black, 2002; Kamps, Abbott, Greenwood, Arreage-Mayer, Wills, Longstaff, Culpepper, & Walton, 2007). As would be expected, studies have also found that the magnitude of the results is stronger when the program is implemented with greater fidelity (e.g. Ross, et al, 2004; author, 2009).

Longitudinal Studies

Numerous studies of Direct Instruction, such as those noted above, have documented its success in promoting achievement over a relatively short period of time such as one to two years and with children at various grade levels and ages. Yet, many fewer have examined longer time periods. This lack is especially striking given the claim of Carnine and associates (1992) that it should promote long-term cognitive growth, thus potentially countering the well documented decline in achievement for low-income children relative to higher income peers in the later elementary grades. After an extensive review of the literature we found only one study, published over 25 years ago, that examined the impact of receiving Direct Instruction in early elementary school on students' reading achievement in fifth and sixth grade, the later years of elementary school.[1] We also found a few other studies that examined somewhat shorter time spans.

Becker and Gersten's Analysis of Multiple Implementations – Becker and Gersten (1982) examined the achievement of low-income fifth and sixth graders in five diverse sites, all of whom had completed grades 1 through 3 in Direct Instruction as part of the large, federally funded Follow Through program. Students' achievement scores on the Wide Range Achievement Test (WRAT) and the Metropolitan Achievement Test (MAT) were compared to those of students in local, demographically similar, comparison groups who had received traditional educational programs in their early elementary years. Multivariate statistics were used to control for a number of variables that could affect achievement, including ethnicity, family income, home language, mother's education, gender and number of siblings. Analyses of covariance that incorporated the control variables were

[1] We were also unable to find such longitudinal studies of other reading curricula.

conducted separately for each of the communities. Analyses involved data gathered in two different years, and across the sub-dimensions as well as the total scores of the two achievement tests, thus providing multiple replications of the analyses. Three meta-analytic techniques were used to summarize the results, and all of these supported the conclusion that children who had received Direct Instruction in the early grades had significantly higher achievement in fifth and sixth grade than students in the comparison groups. Becker and Gersten noted that the strongest and most consistent findings occurred with reading decoding, as measured by the WRAT, with a median effect size across years and sites of .48 (calculated from Table VI, p. 83). More variable results occurred with the measure of word knowledge and total scores for reading.

The Direct Instruction Follow Through program that Becker and Gersten examined ended when the children were in third grade. Even though the strong start that the DI children received resulted in reading achievement scores in the higher elementary grades that were significantly greater than those of demographically similar students in the comparison groups, their achievement relative to national norms declined after third grade. As Becker and Gersten described it,

> 2 years after the program had ended, all samples made appreciable, significant drops against the national norm group in … reading….Though in many domains Follow Through graduates outperform the control students in Grades 5 and 6…, low-income Follow Through students are losing against the national normal sample. (p. 88)

This loss against the national norms parallels the fourth grade slump noted in the literature as common among students from low income backgrounds (e.g. Chall, Jacobs, & Baldwin, 1990, Hirsch, 2003).

Becker and Gersten concluded their article by noting a need to extend the principles of Direct Instruction, including mastery learning, high levels of feedback, and incremental steps of instruction and learning, to the intermediate grades. In the decades following the publication of this work, DI curriculum developers, including Becker, heeded this call, focusing especially on ways to close the gap in vocabulary knowledge of low income children and their higher income peers. The Direct Instruction corpus has expanded to include instruction for students at higher grade levels and with a variety of higher-level subjects ranging from literary analysis to chemistry to legal reasoning. In addition, the curriculum from grade three to five includes substantial content related to science and literature. Evaluations of these curricula indicate that they can successfully promote student learning at those levels (Carnine 1991; Carnine, Grossen, & Silbert 1992; Carnine & Kameenui 1992). We have not, however, been able to find any longitudinal studies that have examined changes in achievement through the elementary years when exposed to the contemporary DI curriculum.

Studies of Single Districts and Shorter Time Periods – Several studies have examined the impact of receiving Direct Instruction over shorter time periods than that addressed by Becker and Gersten (1982). For instance, Carlson and Francis (2002) compared the third grade achievement of students receiving the Direct Instruction program, Reading Mastery, in earlier grades with the

achievement of students in demographically similar control schools who had used a traditional curriculum. Using multivariate analyses and a variety of achievement measures they found that students with more exposure to Direct Instruction had significantly higher achievement at the end of third grade, but that the largest increases relative to the comparison group were made during first grade. Similarly, Kamps and associates (2003) compared the impact of three different curricula – Reading Mastery, Success for All, and a literature based program – on growth in reading achievement over three years for students in the early primary grades and found that students in the Reading Mastery curriculum had the greatest growth, followed by those having Success for All.

Two previous studies examined the sample used for the analysis in this paper, focusing on a subgroup of the students studied here. Mac Iver and associates (Mac Iver & Kemper 2002; Mac Iver, Kemper, & Stringfield 2003) used a subgroup of the schools studied in this paper to examine changes in achievement for two cohorts of students exposed to Direct Instruction or to a control curriculum. One cohort began DI in kindergarten and was followed through third grade, and the other cohort began DI in second grade and was followed through fifth grade. Results indicated positive impacts of DI on vocabulary and oral reading fluency, but less impact on reading comprehension. Addison and Yakimowski (2003) examined descriptive data regarding reading achievement from first through fifth grade for a portion of the schools in our analysis and data regarding achievement from first through fourth grade for another subset of the schools that we examine. They also used multivariate

analyses to examine changes in achievement from kindergarten to second grade and from third to fifth grade for two subgroups of students. The descriptive results over the longest time period (first through fifth grade) indicated substantial differences in favor of Direct Instruction. The differences were much smaller and often not significant for analyses of the shorter time periods.

Summary and Hypotheses

Even though a large body of literature has documented the relationship of Direct Instruction to higher reading achievement for diverse groups of students, relatively little research has examined the relationship of DI to changes in achievement from first grade to the end of elementary school, a time period that is especially important in predicting later academic success. The major exception is work by Becker and Gersten (1982), published more than a quarter of a century ago. They found that students who had received Direct Instruction through grade 3 had significantly higher reading achievement in grades 5 and 6 than demographically similar students in traditional curricula. They also noted, however, that the achievement of students in their study, most of whom were from lower income families, was declining relative to national norms as they progressed to the higher grades, a finding that parallels the well documented fourth grade slump.

Our analysis returns to the question posed by Becker and Gersten, looking at the impact of Direct Instruction on achievement gains from first to fifth grade. We approach this issue by looking at the relationship of exposure to the Direct Instruction curriculum in early elementary school to reading achievement at the

end of elementary school (grade 5) in a low income, urban school system. Specifically we compare the changes in reading achievement from first to fifth grade of students exposed to three different curricular programs in first grade: 1) the highly rated Direct Instruction program; 2) Open Court, another highly rated program; and 3) varied traditional instructional reading programs. If DI is more effective than other programs in promoting achievement, as Becker and Gersten found, we would expect that the changes from first to fifth grade would be larger for the students exposed to Direct Instruction than for students in the two other groups. In addition, as noted above, the DI curriculum has expanded markedly since the publication of Becker and Gersten's (1982) work to incorporate elements that should promote the development of higher order cognitive skills and enhance continuing achievement. To the extent that these modifications and extensions have been successful, we would expect that the students exposed to DI would have achievement scores that were closer to the national norms than students in the control groups.[2]

Methodology

The data for this analysis come from the Baltimore City Public School System (BCPSS) and regard the reading achievement of students who were fifth graders in either 2001-02 or 2002-03 and had been in the same schools five years earlier (in 1997-98 or 1998-99). The sections below describe the procedures, sample, measures, and analysis techniques that were used.

[2] Even though we examine data from the same school system as Mac Iver and Kemper (2002), and Addison and Yakimowski (2003), our work differs from theirs by using multivariate statistics to examine changes from first to fifth grade for all of the students exposed to Direct Instruction in the system, rather than a subsample, and by comparing the impact of exposure to three different curricular programs in first grade.

Procedures

The Baltimore City Public School System (BCPSS) is similar to many other large city school districts that serve students with high levels of poverty and struggle with low levels of achievement. During the period from which our data were obtained, the average school within the system was 84% African American, 1% Hispanic, 14% non-Hispanic white, and had 75% of its students receiving free or reduced lunch.

In the late 1990s, curricular reforms were implemented in the BCPSS elementary schools to address this low achievement. Sixteen BCPSS schools chose to use Direct Instruction (DI) for reading instruction. The curriculum was introduced as part of a whole school reform effort called the Baltimore Curriculum Project and sponsored by the Abell Foundation. The other schools followed the BCPSS curricular guidelines to increase student achievement (Berkeley, 2002, Mac Iver, 2004, Mac Iver and Kemper, 2002, Stringfield & Yakimowski-Srebnick 2005).

Direct Instruction – The major DI reading curriculum is Reading Mastery Classic, a scripted, mastery-based core reading program that focuses on decoding and comprehension. Students are placed in homogeneous groups according to skill level. Ideally, the teacher ensures that all members of the group achieve mastery on all material the program introduces. Students who master content substantially faster or slower than others in their group are placed into other groups in which students have skill profiles similar to those of the incoming student. Student skill is continuously monitored, and problems of mastery are

addressed to ensure all students are at a level commensurate with their current level of skill. Instruction is designed to elicit frequent oral student responses, which increase engagement and create a high rate of active responding. Student responses during independent seat work are also closely monitored and immediately remediated (Marchand-Martella, Slocum, & Martella, 2004).

The students in the Direct Instruction schools also received language instruction for 30 minutes per day. The curricula used for this instruction were Language for Learning, Language for Thinking, and Reasoning and Writing. These are general knowledge programs that focus on oral language development. The Direct Instruction curriculum extended through fifth grade includes instruction in U.S. History, writing, reasoning, and spelling. The reading programs also include extensive general knowledge material related to the sciences and social studies. In general, the DI curriculum is designed to provide students with background knowledge and underlying schemata that can help them address and understand content throughout the academic curriculum. As noted above, many of these extensions of the curriculum were, in part, a response to the findings of Becker and Gersten (1982), who expressed concern regarding falling achievement in the later elementary years and called for the development of additional instructional modules addressing this issue.

At the beginning of implementation the Baltimore Curriculum Project contracted with the National Institute for Direct Instruction (NIFDI) to provide implementation support through preservice and inservice training, coaching, and technical assistance to all the Direct Instruction Schools. The developer of Direct

Instruction is affiliated with NIFDI, and the organization prides itself on strict fidelity to the DI model as it was originally validated through extensive field testing.

When a school implements the NIFDI model, instructional programs are phased in over several years. In the first year, language and reading programs are introduced in kindergarten and first grade. In subsequent years the curriculum is added to the higher grades and other parts of the program, including mathematics, are implemented. To ensure appropriate placement of students the NIFDI model requires homogeneous instructional groupings, with regrouping as often as four times a year; monitoring student progress through direct observations, in-program tests, records of lessons completed, and at least weekly check-ins.

A NIFDI Implementation Manager (IM) trains teachers, assistants, and coaches. The IM is typically on site a total of approximately 35 person days per year, working in classrooms with the teachers and presenting in-service sessions that address problems teachers are experiencing. There also are weekly conference calls that address current problems each classroom is experiencing in meeting projected performance gains. All teachers receive pre-service training and coaching until they teach each program to a minimum adequate level of fidelity. Teachers continue to receive in-service coaching to improve implementation fidelity. During the second year, teachers who perform well are identified as coaches and are deployed to work with other teachers in the school. NIFDI-support is usually phased out after three years, although the Baltimore

Curriculum Project continued to operate the DI schools as Charter Schools and to provide on-going consultation and coaching for teachers. [3]

Eight of the schools began implementing DI in 1996-97, three in 1997-98 and the remaining five in 1998-99. In the analysis reported below students were designated as belonging to the intervention group (DI) only if they had received this instruction in first grade. If they received DI in grades 2-5 but not in grade 1, as occurred for students who were in first grade in 1997 in the five schools that began implementation in 1998-99, they were omitted from the sample

Control Schools – Before 1998 the other schools in the BCPSS were free to use any curriculum program they desired. There was no district-wide structured reading program, and schools reportedly used a variety of instructional programs. Yet, schools in the BCPSS had substantially lower achievement than other schools in the state and there was extensive pressure for reform (Stringfield and Yakimowski-Srebnick, 2005).

In response to these pressures the BCPSS embarked on a concentrated and extensive reform process. As part of this process the system adopted new citywide reading curricula in the fall of 1998, using Open Court Reading in kindergarten through second grade. Open Court is a phonics-based, highly structured program that has been favorably reviewed by the Florida Center for Reading Research (FCRR 2004) and the Oregon Reading First Center (ORFC 2004) as a core instructional program for Reading First. Students in third grade and beyond

[3] Five of the sixteen schools that introduced DI ceased their association with NIFDI soon after implementation and began to work with another provider of implementation support. To maintain an adequate sample size the results for all sixteen schools are combined for the present analysis. Separate results for these two subsets of the DI schools are available upon request from the author and do not alter the substantive findings reported here.

received instruction from a Houghton Mifflin series. The BCPSS provided extensive professional development support in 1998-99 for teachers as they began to use this new curriculum. At the same time the system introduced other reforms such as lowering student/teacher ratios, expanding kindergarten programs to full-day, and expanding before-school, after-school, and summer school programs (Berkely 2002, Mac Iver, Kemper, & Stringfield 2003; Stringfield and Yakimowski-Srebnick 2005). Note that these changes also affected students in the DI schools.

In this paper we differentiate students in the Control Schools into two groups: 1) those that had diverse first grade reading instruction (the 1997-8 first grade cohort) and 2) those that had first grade reading instruction in Open Court (the 1998-99 cohort). Note, however, that students with the diverse instruction in first grade had Open Court in second grade and that all of the students in both sets of Control Schools had a Houghton Mifflin series beginning in third grade.

Participants

Data for this study included student cohorts who began first grade in the BCPSS in 1997-98 or 1998-99 and were in the same schools in fifth grade in either 2001-02 (for the first graders in 1997-98) or 2002-03 (for the first graders in 1998-99).[4] Students who were held back were not included in the analyses.[5]

[4] To preserve degrees of freedom and keep the sample size as large as possible, the two cohorts of DI students were grouped together for the analysis presented here. Results are virtually identical when they are separated. An appendix to this article provides descriptive results separating the two cohorts and details of the multivariate analyses separating the cohorts are available upon request from the author.

[5] Mac Iver and Kemper (2002) analyzed data from some of the schools included in this analysis and found that the DI schools had a lower rate of retention ("failing" a grade) than other schools in the system. This could, conceivably, result in a larger proportion of potentially "low achieving" or "higher risk" students in the fifth grade in the DI schools and thus provides a conservative test of

There was substantial student turnover during the study period. Of the over 17,000 students in first grade in the BCPSS in 1997-98 and 1998-99 about half (49.7 percent) had left the BCPSS five years later, while almost one-quarter (23.3%) had transferred to another school within the system. Slightly more than one-fourth of the students who were in first grade in the system in 1997-98 and 1998-99 were in fifth grade in the same school five years later. Students in the DI schools were slightly, but significantly, more likely to remain in the same schools in fifth grade than those in the two control groups. Almost thirty percent (29.3 percent) of the students in the DI schools persisted in the same schools, compared to 26 percent of those in the other two groups (chi-square = 7.63, df = 2, p = .02).

The panel sample (those that were in the same schools in 5th grade that they attended in 1st grade five years earlier) analyzed in this report includes a total of 4,572 students. Of these, 500 were in DI Schools, 2,197 were in control schools with variable curricula in first grade, and the remaining 1,875 were in first grade in schools that implemented Open Court.[6] Six schools in the BCPSS, all within the Control Groups, had no students who persisted from first to fifth grade. As a result, the sample included students from 113 schools. An appendix to this article provides additional details on the number of students in these various groups in each year.

---

the impact of DI on changes in achievement from first to fifth grade. However, when students who were in first grade in 1997 and were in fifth grade in 2003 (6 years after first grade, rather than 5) were included in the multivariate analyses, substantive results were virtually identical to those reported in this article. These results are also available upon request from the author.

[6] There was no way to control for the possibility that a student had attended multiple other schools between first and fifth grade. It is possible that some of the students in the panel sample were in the targeted schools in both first and fifth grade, but had attended other schools in the interim. Assuming that such children would have less of the "full treatment," this would bias results in a conservative direction.

Instrumentation

The Comprehensive Test of Basic Skills/TerraNova (CTBS), a widely used standardized achievement test, was administered to all students in the spring of each school year from 1997-1998 through 2002-2003 as part of a system-wide testing program (CTB/McGraw-Hill 2001). The 4th edition was administered in the spring of 1998 and 1999, and the 5th edition in 2002 and 2003. Two subtest scores, Reading Comprehension and Reading Vocabulary, were analyzed. Normal Curve Equivalent (NCE) scores were used to help ensure comparability from one year to another and to allow the use of statistical calculations. Because the meaning of NCE scores is not intuitively obvious, the scores were also converted, when appropriate for descriptive purposes, into percentiles using a standard conversion table.

To adjust for the impact of socio-economic status on student achievement and achievement gains over time, we used, as a control variable, the proportion of students within a school who received free or reduced lunch. Because we did not have access to individual level demographic data, a school level measure was used as a proxy. In preliminary analyses we also used a factor score incorporating measures of race-ethnicity and school poverty as a control variable. Results, available from the author, were identical to those reported here.

Most schools in the BCPSS included high proportions of students receiving FRL. The average student in the sample attended a school where 71 percent of students were in this category. Students in the panel group differed from other first grade students (those who did not remain in the same schools

from first to fifth grade) in the average socio-demographic characteristics of their schools, with those who persisted being significantly more likely to attend higher status schools (72% FRL versus 76% FRL, $F = 226.275$, $p < .001$). At the same time, students in the DI schools were more likely than other students to be in high poverty schools (75% for students in DI schools versus 71% for other students; $F = 68.983$, $p < .001$). There was, however, no significant interaction; that is, the differences in school poverty context between students in the panel group and the other students was the same across the three groups of schools (DI schools and the two control schools) (F interaction of group and panel status = 0.81, $p = .44$).

In addition, as could be expected, the students who persisted in the same schools had higher first grade achievement than their more mobile counterparts. These differences were similar across the three groups for the measure of comprehension (F interaction = 1.89, $p = .15$), but varied across the groups for the measure of vocabulary. With this measure, the difference between the persisters (those in the panel group) and the mobile students was smaller for students in the Direct Instruction schools than for students in the other schools (F, interaction, = 7.72, $p < .001$). (See the appendix to the article for full descriptive data.)

Note that these differences resulted in the students in the DI schools being somewhat more at risk in first grade than the other students. They attended schools where more students received free and reduced lunch and they also had lower average vocabulary scores in first grade. If anything, this could produce a conservative impact on the results (a smaller chance of significant findings in favor of Direct Instruction).

Even though students who remained in the panel sample and were included in our analysis were more likely than those who left the school system to be in schools with more advantaged characteristics, it is important to stress that the average student in the sample attended a very high poverty school with substantial proportions of racial-ethnic minorities in attendance. In fact, the average (mean) values given above understate the degree of segregation and poverty in the BCPSS. A somewhat more accurate picture may come from positional measures, such as quartiles. These indicate that 50 percent of the students attended schools where at least 97 percent of the students were African American, and 75 percent attended schools where at least 78 percent of the students were African American. Similarly, 75 percent of the students attended schools where two-thirds of the students were on free or reduced lunch.

Analysis

The impact of Direct Instruction on changes in students' reading achievement from first to fifth grade was examined in two ways. First, we used descriptive statistics, paired t-tests, and repeated measures analysis of variance to examine the average change in students' scores over the years in each of the three groups of schools (DI, varied first grade curricula, and Open Court in first grade). The percentage change in NCE scores and effect sizes were also used to assess the substantive magnitude of these changes. Effect sizes were calculated using the method developed by Dunlap and associates for calculating Cohen's d with

correlated samples (1996).[7] If having Direct Instruction as a core curriculum enhances growth in students' achievement from first to fifth grade more than either the varied curricula or Open Court, we would expect a significant interaction effect in the analysis of variance and that the pair-wise comparisons would indicate that students in DI schools had stronger growth in achievement over time.

Second, mixed model regressions were conducted regressing fifth grade achievement on first grade achievement, the percentage of students in the school who received free or reduced lunch, and dummy variables for group. Students who received Open Court were the omitted category in the dummy variable coding. To the extent that having Direct Instruction in first grade was more effective at promoting higher student achievement in later years, it would be expected that students in the DI Schools would have higher achievement scores than those in the Control Schools, even when school characteristics and first grade achievement were controlled.

Mixed models are particularly appropriate for analyzing multi-level data, such as that regarding students and the schools that they attend. In these models a "random variable" is used to control for differences between schools (often termed the level 2 entity) while calculating regression coefficients regarding the impact of variables from both students and schools on achievement. The random variable is equivalent to having a separate intercept in the regression equation for each school. The coefficients associated with the various individual and school

[7] McLean, O'Neal, & Barnette (2000) caution that effect sizes calculated with NCE scores are inherently smaller than those with other metrics. Thus, the effect sizes presented might be a conservative estimate.

related variables are then calculated while this between school variance is controlled. The analysis also allows one to calculate the amount of variance in the dependent variable that occurs between schools (Raudenbush and Bryk 2002, Singer 1998). We used the Mixed procedure in SPSS for these calculations.

Results

Descriptive Statistics and Analysis of Variance

Descriptive results regarding the first and fifth grade reading achievement of students in the three groups of schools are summarized in Table 1. The mean and standard deviation of NCE scores for each measure and grade and the associated percentage change in these scores are in Panel A, the results of paired t-tests and associated effect sizes are in Panel B, the percentiles that correspond to the means in Panel A are in Panel C, and the results of the repeated measures analyses of variance are in Panel D. The percentiles can be interpreted as the score that an average student in each group would have in a given grade. All calculations were done with NCE scores.

As hypothesized, the students in the DI schools had significantly greater gains in achievement from first to fifth grade than students in the other two groups. This is indicated by the significant interaction effects in the analyses of variance (Panel D), the larger average differences and effect sizes for the DI group (Panel B), and the larger percentage changes for the DI group (Panel A). In first grade, students in the DI schools had lower average vocabulary and comprehension scores than students in either of the other two groups of schools. The average student in the DI schools scored at the 38th percentile in vocabulary

and the 36th percentile in comprehension, while the average student in the Control Schools scored at the 45th (variable curricula schools) or 58th percentile (Open Court schools) in vocabulary and the 38th (varied curricula) and 45th percentile (Open Court) in comprehension. By fifth grade these same DI students had higher scores than students in the other schools and scored above the national mean on both measures.

The differences in average change from first to fifth grade were not small in magnitude. On average, students in the DI schools had vocabulary scores in fifth grade that were 19 percent higher than their scores in first grade, compared to an increase of four percent for those in the varied curricula first grades and a decline of seven percent for the students receiving Open Court. Students in all the groups had higher comprehension scores in fifth grade than in first grade, but again the changes were substantially larger for those in the DI schools – an average increase of 22 percent – compared to changes of 14 percent for those with the varied first grade curriculum and six percent for those with Open Court. These conclusions are supported by the effect sizes, which surpass the .25 criterion for substantively significant effects (Wolf 1986) for both measures for the students in the DI schools, but for only one of the other four comparisons (for comprehension for students with the varied first grade curriculum).

[Table 1 About Here]

Mixed-Model Results

Table 2 gives the results of the mixed model analyses, regressing fifth grade achievement on first grade achievement and controlling for differences

between the schools in poverty context. The first panel gives results regarding vocabulary, and the second gives results regarding comprehension. Four models were tested. Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This model tests the null hypothesis that the schools are equal in average achievement. The associated statistics are reported in the lower lines of each panel of the table. The correlation ratio is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that from seven to nine percent of the variance in achievement is between schools rather than simply between students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model and school differences are controlled. The null hypotheses can be easily rejected with all the models. There is significant variation between schools and also between students in all the models. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those available for this analysis.

[Table 2 About Here]

Model 2 adds the percentage of students receiving FRL within the school as a predictor, and results with this model indicate the extent to which the poverty context of students' schools is related to their achievement in fifth grade. As would be expected, the coefficients, with both dependent measures, indicate that

students had significantly higher achievement in fifth grade when they attended schools with fewer low income students. Model 3 adds the dummy variables indicating the type of school that students attended, testing the hypothesis that students' achievement in fifth grade varied between the three groups of schools once school poverty context was controlled; and Model 4 adds first grade achievement, controlling for students' skills at the end of first grade.

The -2 log likelihood measures and the BIC values can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the comparisons between these values are in the final rows of each part of Table 2. The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. Both the BIC values and the changes in the -2 log likelihoods indicate that Model 4, which includes all the variables, provides the best fit to the data.[8]

The coefficients associated with the dummy variables for group support the conclusions obtained with the descriptive results in Table 1. They indicate that students who attended DI schools in first grade had significantly higher reading achievement scores in fifth grade than students in other types of schools, even when the poverty level of their school and their levels of first grade achievement were controlled. The students in the varied curriculum first grades also had significantly higher scores than the students in the Open Court schools, but the differences were not as large. The mixed model results indicate that, on average,

[8] We also tested interaction effects, specifically the interaction of group and school characteristics and the interaction of group and first grade achievement. Neither of these sets of interaction effects was significant.

once characteristics of their schools and their first grade achievement were controlled, students in the DI schools would have vocabulary achievement scores in fifth grade that were more than five NCE points higher than those in the Open Court schools and comprehension achievement scores that were more than three points higher. The students in the schools with the varied curriculum would have scores from 1.3 to 1.4 NCE points higher than those in the Open Court schools.

Summary and Discussion

Our results from a low income, large, urban school district indicate that it is possible to promote higher reading achievement over time. The data presented above indicate that the efforts of the BCPSS to promote extensive system wide change and higher student achievement have met with success. With one exception (changes in vocabulary scores for students who had Open Court in first grade), students in all three groups of schools had significantly higher reading achievement in fifth grade than in first grade.

However, as hypothesized, our results also indicate that students who had the Direct Instruction curriculum through their elementary grades had significantly greater gains than students in the other curricula. These differences were both statistically and substantively significant and appeared with both descriptive and inferential analyses, with strong controls for the socio-demographic context of the schools they attended, and controlling for levels of first grade achievement.

In addition, the results suggest that the DI curriculum may help counteract the so-called fourth grade slump, when students from low-income backgrounds

begin to fall progressively farther behind their more advantaged peers. Even though the students in the DI schools had first grade achievement scores that were lower, on average, than those of students in the other schools, by fifth grade they had significantly higher average achievement scores than students in the other schools and these scores were above the national average. In contrast, only one of the four fifth grade achievement measures for the control groups averaged above the 50th percentile (the national average), and this represented a decline from the first grade score. Thus, our results suggest that the modifications and expansions of the DI curriculum that occurred after Becker and Gersten's (1982) work have been successful in promoting higher order cognitive skills and continuing achievement, at least through the end of fifth grade.

Hirsch (2003) summarizes three basic principles that influence students' reading comprehension and are related to the decline in the later grades: fluency, allowing the student to focus on comprehension; vocabulary, which increases the probability of comprehending more complex material; and domain knowledge, a threshold level of understanding that makes it easier to understand ever more complex and demanding material. The students in the DI schools had significantly greater achievement gains on both the measure of vocabulary and the measure of comprehension than students in the other schools, reflecting at least two of the three areas that Hirsch has highlighted. In addition, the achievement gains over time for the DI students were stronger for the measure of comprehension than vocabulary. Given the centrality of comprehension to learning in the upper

grades, this difference can also be seen as supporting the efficacy of DI in counteracting declines in achievement in the upper elementary grades.

It is, of course, very important to replicate the work reported here, especially given the apparent paucity of longitudinal research. It would be important to examine changes in achievement over the elementary years, and beyond, with both the DI curriculum and other curricula. Our results suggest that students in the DI curriculum had higher levels of achievement than those with a variety of approaches or Open Court in first grade, a finding reported elsewhere (e.g. Crowe, Connor & Petscher, 2009; O'Brien and Ware, 2002).
It is possible, however, that once school reforms became more fully implemented and stabilized, the achievement of students in these other curricula would be higher; and this possibility should be investigated.

It would also be important to have studies that include measures of achievement prior to beginning school. Our study examined changes from the primary to intermediate grade-school years (from the end of first grade to the end of fifth grade), as have other examinations of the "fourth grade slump," but we had no measure of entry level skills prior to beginning school. Having such data could provide better estimates of the long-term impact of curricula on growth. In addition, a very worthwhile area of further research could be to investigate the way in which extending academic DI into the preschool years might increase first grade performance and result in even higher achievement at fifth grade for especially vulnerable children.

Finally, future analyses would be improved by including individual measures of socio-economic characteristics. Given limitations of our data set, we did not have individual level measures of socio-demographic characteristics. While substantial literature has documented the influence of school contexts of disadvantage on achievement, the most precise estimates of the effects found here would occur if we had both individual and school level measures of socio-demographic characteristics.

Earlier analyses of the impact of Direct Instruction on student achievement in the BCPSS, using subsets of the data examined here, recommended that the system continue using the program (Mac Iver, Kemper, and Stringfield 2003; Addison and Yakimowski 2003). Our results would confirm that recommendation and suggest, as have many other studies cited above, that students are well served by the DI curriculum. Even more important, our results suggest that the curriculum has long-term impacts and, at least for students in this high poverty school system, can help counter the well-documented tendency for declining achievement over time. The importance of this result for the well-being of low income students, as well as for society as a whole, cannot be overstated (see Belfield & Levin 2007).

References

Adams, M. J. (1990). Beginning to read: Thinking and learning about print. Cambridge, Mass.: The MIT Press.

Adams, G. L. & Engelmann, S. (1996). Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.

Addison, K.L. &Yakimowski, M.E. (2003). An Evaluation of the Direct Instruction Program. Baltimore City Public School System: Division of Research, Evaluation, Assessment, and Accountability:

American Federation of Teachers. (1998). Building on the best, learning from what works: Seven promising reading and language arts programs. Washington, D.C.: AFT.

Anderson, R. C., Hiebert, E.H., Scott, J.A., & Wilkinson, I.A.G. (1985). Becoming a nation of readers: The report of the Commission on Reading. Washington, D.C.: National Institute of Education.

Baker, S. K. Kameenui, E.J., Simmons, D.C., & Stahl, S.A. (1994). Beginning reading: Educational tools for diverse learners. School Psychology Review, 23, 372-391.

Beck, I.L. & McCaslin, E.S. (1978). An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. LRDC Report No. 1978/6 Pittsburgh.

Becker, W.C. & Carnine, D.W. (1980). Direct Instruction: An effective approach to educational intervention with the disadvantaged and low performers. In B.B. Lahey & E. E. Kazdin (Eds.) Advances in clinical child psychology (Vol. 3, pp. 429-473). New York: Plenum.

Becker, W.C. & Gersten, R. (1982). A follow-up of Follow Through: The later effects of the Direct Instruction model on children in fifth and sixth grades. American Educational Research Journal, 19, 75-92.

Belfield, C.R. & Levin, H.M. (2007).The Price We Pay: Economic and Social Consequences of Inadequate Education. Washington, D.C. Brookings.

Benner, G. J. (2007). The relative impact of remedial reading instruction on the basic reading skills of students with emotional disturbance and learning disabilities. Journal of Direct Instruction, 7(1), 1-15.

Benner, G. J., Kinder, D., Beaudoin, K.M., Stein, M., & Hirschmann, K. (2005).The effects of the "Corrective Reading Decoding" program on the basic reading skills and social adjustment of students with high-incidence disabilities. Journal of Direct Instruction, 5(1), 67-80.

Berkeley, M. (2002). The importance and difficulty of disciplined adherence to the educational reform model. Journal of Education for Students Placed at Risk, 7, 221-239.

Blau, P.M. & Duncan, O.D. (1967). The American Occupational Structure. New York: Wiley.

Bond, G. & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. Reading Research Quarterly, 2, 5-142.

Borman, G. D., Hewes, G.M., Overman, L.T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research, 73, 125-230.

Carlson, C. D. & Francis, D.J. (2002). Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE). Journal of Education for Students Placed at Risk, 7 (2), 141-166.

Carnine, D. (1991) Curricular interventions for teaching higher-order thinking for all students. Introduction to the special series. Journal of Learning Disabilities, 24 (5),261-269.

Carnine, D., Grossen, B., & Silbert, J. (1992). Direct instruction to accelerate cognitive growth," ADI News, Fall, 1992, p. 33.

Carnine, D. & Kameenui, E. (1992). Teaching higher order thinking to all students. Austin Texas: Pro Ed.

Chall, J.S. (1967). Learning to read: The great debate. New York: McGraw Hill.

Chall, J. S. & Jacobs, V.A. (2003). Poor Children's Fourth-Grade Slump. American Educator Spring 2003.

Chall, J. S., Jacobs, V.A., & Baldwin, L.E. (1990). The Reading Crisis: Why Poor Children Fall Behind. Cambridge, Mass.: Harvard University Press.

Cooke, N. L., Gibbs, S. L., Campbell, M. L., & Shalvis, S. L. (2004). A comparison of Reading Mastery Fast Cycle and Horizons Fast Track A-B on the reading achievement of students with mild disabilities. Journal of Direct Instruction, 4(2), 139-151.

Crowe, E.C., Connor, C.M., & Petsher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. Journal of School Psychology 47, 187-214.

CTB/McGraw-Hill. (2001). TerraNova technical manual report. Monterey, CA: Author.

Dowdell, T. (1996). The effectiveness of Direct Instruction on the reading achievement of sixth graders. ERIC # 396 268.

Drakeford, W. (2002). The impact of an intensive program to increase the literacy skills of incarcerated youth. Journal of Correctional Education, 53(4), 139-144.

Dunlap, William P., Jose M. Cortina, Joel B. Vaslow, and Michael J. Burke. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. Psychological Methods, 1, 170-177.

Engelmann, S. & Carnine, D. (1982). Theory of instruction: Principles and applications. New York: Irvington Publishers.

Engelmann, S. (2007). Teaching needy kids in our backward system: 42 Years of trying. Eugene, Oregon: ADI Press.

Farkas, G. (1996). Human Capital or Cultural Capital: Ethnicity and poverty groups in an urban school district. New York: Aldine.

Flores, M. M., Shippen, M. E., Alberto, P., & Crowe, L. (2004). Teaching letter-sound correspondence to students with moderate intellectual disabilities. Journal of Direct Instruction, 4(2), 173-188.

Flores, M. M., & Ganz, J. B. (2007). Effectiveness of Direct Instruction for teaching statement inference, use of facts, and analogies to students with developmental disabilities and reading delays. Focus on Autism and Other Developmental Disabilities, 22(4), 244-251.

Florida Center for Reading Research (2004). Open Court Reading Pre-K. Florida Center for Reading Research. http://www.fcrr.org/index.htm, accessed January 29, 2009.

Foorman, B. R. (1995) Research on "the great debate": Code-oriented versus whole language approaches to reading instruction. School Psychology Review, 24, 376-392.

Fukkink, R.G. & deGlopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. Review of Educational Research, 68, 450-469.

Grossen, B. (2004). Success of a Direct Instruction model at a secondary level school with high-risk students. Reading & Writing Quarterly, 20, 161-178.

Grossen, B. (1997). A synthesis of research on reading from the National Institute of Child Health and Human Development. Eugene, Oregon: University of Oregon.

Gunn, B., Smolkowski, K., Biglan, A., & Black, C. (2002). Supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. The Journal of Special Education, 36(2), 69-79.

Haring, N. G., & Krug, D. A. (1975). Evaluation of a program of systematic instructional procedures for extremely poor retarded children. American Journal on Mental Retardation, 79, 627-631.

Harris, R. E., Marchand-Martella, N. E., & Martella, R. C. (2000). Effects of a peer-delivered Corrective Reading program. Journal of Behavioral Education, 10, 21-36.

Herman, R., Aladjam, D., McMahon, P., Masem, E., Mulligan, I., Smith, O., O'Malley, A., Quinones, S., Reeve, A., & Woodruff, D. (1999). An *educator's guide to schoolwide reform*. Washington, D.C.: American Institutes for Research.

Hirsch, E.D., Jr. (2003). Reading comprehension requires knowledge – of words and the world: Scientific insights into the fourth-grade slump and the nation's stagnant comprehension scores. American Educator, 27 (1), 10-29.

Houchins, D. E., Jolivette, K., Krezmien, M. P., & Baltodano, H. M. (2008). A multi-state study examining the impact of explicit reading instruction with incarcerated students. The Journal of Correctional Education, 5(1), 65-85.

Institute of Child Health and Human Development (ICHHD). (2000). Report of the National Reading Panel. Teaching Children to Read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). Washington, D. C.: U. S. Government Printing Office.

Jencks, C., Crouse, J. & Mueser, P. (1983). The Wisconsin model of status attainment: A national replication with improved measures of ability and aspiration. Sociology of Education 56 (1): 3-19.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. Journal of Educational Psychology, 80 (4), 437-447.

Juel, C. & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly, 35, 458-492.

Kamps, D., Abbott, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., et al. (2007). Use of evidence-based, small group reading instruction for English language learners in elementary grades: Secondary-tier intervention. Learning Disability Quarterly, 30, 153-168.

Kamps, D.M., Will, H.P., Greenwood, C.R., Thorne, S., Lazo, J.F., Crockett, J.L., Akers, J.M., & Swaggart, B.L. (2003). Curriculum influences on growth in early reading fluency for students with academic and behavioral risks: A descriptive study. Journal of Emotional and Behavioral Disorders 11 (4): 211-224.

Kuder, S. J. (1991). Language abilities and progress in a Direct Instruction reading program for students with learning disabilities. Journal of Learning Disabilities, 24(2), 124-127.

Kuder, S. J. (1990). Effectiveness of the DISTAR reading program for children with learning disabilities. Journal of Learning Disabilities, 23(1), 69-71.

Lewis, A. (1982). An experimental evaluation of a direct instruction programme (Corrective Reading) with remedial readers in a comprehensive school. Educational Psychology, 2(2), 121-135.

Mac Iver, M. A. & Kemper, E. (2002). The impact of direct instruction on elementary students' reading achievement in an urban school district. Journal of Education for Students Placed at Risk, 7, (2) 197-220.

Mac Iver, M.A. (2004) Systemic supports for comprehensive school reform: The institutionalization of Direct Instruction in an urban school system. Journal of Education for Students Placed at Risk 9 (3), 303-321.

Mac Iver, M.A., Kemper, E. & Stringfield, S. (2003). The Baltimore Curriculum Project: Final Report of the Four-Year Evaluation Study. Center for Research on the Education of Students Placed at Risk, Johns Hopkins University.

Maggs, A., & Morath, P. (1976) Effects of direct verbal instruction on intellectual development of institutionalized moderately retarded children: A 2-year study. Journal of Special Education, 10(4), 357-364.

Malmgren, K. W., & Leone, P. E. (2000). Effects of a short-term auxiliary reading program on the reading skills of incarcerated youth. Education and Treatment of Children, 23(3), 239-247.

Marchand-Martella, N., Martella, R., Orlob, M., & Ebey, T. (2000). Conducting action research in a rural high school setting using peers as Corrective Reading. Rural Special Education Quarterly, 19(2).

Marchand-Martella, N.E., Slocum, T.A., & Martella, R.C. (2004). Introduction to Direct Instruction. Boston: Pearson.

McLean, J.E., O'Neal, M.R., Barnette, J.J. (2000). Are all effect sizes created equal? Paper presented at the annual meeting of the mid-south educational

research association (28th, Bowling Green, KY, November 15-17, 2000). ED 488 188

Murphy, Joseph (2004). Leadership for Literacy: Research-Based Practice, PreK-3. Sage: Corwin.

National Institute of Child Health and Human Development (NICHHD). (1996). Thirty years of NICHD research: What we now know about how children learn to read. Effective School Practices, 15, 33-46.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, D.C.: National Institute of Child Health and Human Development.

O'Brien, D.M. & Ware, A. M. (2002). Implementing research-based reading programs in the Fort Worth Independent school district. Journal of Education for Students Placed at Risk, 7, 167-195.

Oregon Reading First Center. (2004). Review of Open Court. Eugene, Oregon: Oregon Reading First. http://oregonreadingfirst.uoregon.edu/downloads/instruction/curriculum_review/core_review/oc2002_core_review.pdf (last viewed January 2009).

Pflaum, S., Walberg, H.J., Karigianes, M.L., & Rasher, S.P. (1980). Reading instruction: A quantitative analysis. Educational Researcher, 9, 12-18.

Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, California: Sage.

Riepl, J. H., Marchand-Martella, N. E., & Martella, R. C. (2008). The effects of Reading Mastery Plus on the beginning reading skills of students with intellectual and developmental disabilities. Journal of Direct Instruction, 8(1), 29-39.

Rosenshine, B. (2002). Helping students from low-income homes read at grade level. Journal of Education for Students Placed at Risk 7 (2), 273-283.

Ross, S.M., Nunner, J.A., Goldfeder, E., McDonald, A., Racho, R. & Hornbeck, M., et al. (2004). Using school reform models to improve reading achievement: A longitudinal study of Direct Instruction and Success for All in an urban district. Journal of Education for Students Placed at Risk 9 (4), 357-388.

Scarlato, M. C., & Ashara, E. (2004). Effects of Corrective Reading in a residential treatment facility for adjudicated youth. Journal of Direct Instruction, 4(2), 211-217.

Sewell, W. H., Haller, A.O. & Ohlendorf, G.W. (1970). The Educational and early occupational status attainment process: Replication and Revision. American Sociological Review 35 (6): 1014-1027.

Shippen, M. E., Houchins, D. E., Steventon, C., & Sartor, D. (2005). A comparison of two Direct Instruction reading programs for urban middle school students. Remedial and Special Education, 26(3), 175-182.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics 23, 323-355.

Smith, S., Simmons, D., Gleason, M., Kameenui, E., Baker, S., Sprick, M., Gunn, B., & Thomas, C. (2001). An analysis of phonological awareness instruction in four kindergarten basal reading programs. Reading and Writing Quarterly, 17, 25-50.

Snider, V.E. (1990). Direct instruction reading with average first graders. Reading Improvement, 27, 143-148.

Snow, C. E., Burns, M.S., & Griffin, P. (eds.). (1998). Preventing reading difficulties in young children. Washington, D. C.: National Academy Press.

Stanovich, K.E. (1994). Romance and reality. The Reading Teacher, 47, 280-291.

Stebbins, L.B., St. Pierre, R.G., Proper, E.C., Anderson, R.B., & Cerva, T.R. (1977). Education as experimentation: A planned variation model (Vol. 4-A). Cambridge, MA: Abt Associates.

Steventon, C. E., & Fredrick, L. D. (2003). The effects of repeated readings on student performance in the Corrective Reading program. Journal of Direct Instruction, 3(1), 17-27.

Stringfield, S.C. & Yakimowski-Srebnick (2005). Promise, progress, problems, and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore city public schools. American Educational Research Journal 42, 43-75.

Strong, A. C., Wehby, J. H., Falk, K. B., & Lane, K. L. (2004). The impact of a structured reading curriculum and repeated reading on the performance of junior high students with emotional and behavioral disorders. School Psychology Review, 33(4), 561-581.

Vitale, M. R., & Joseph, B. L. (2008). Broadening the institutional value of Direct Instruction implemented in a low-SES elementary school:Implications for scale-up and reform. Journal of Direct Instruction, 8(1),1-18.

Wolf, F.M. (1986). Meta-Analysis: Quantitative methods for research synthesis. Newbury Park: Sage.

Descriptive Statistics and Paired t-test Results, 1st and 5th Grade Reading Achievement Scores, NIFDI Supported Schools, Other DI Schools, and Control Schools, BCPSS, 1998-2003

A. Descriptive Statistics

		Direct Instruction		Variable Curricula		Open Court	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Vocabulary	1st	43.6	22.0	47.3	21.9	54.4	20.7
	5th	51.8	19.7	49.0	17.8	50.5	16.8
	% ch.	18.8		3.7		-7.0	
Comprehension	1st	42.5	19.8	43.4	20.1	47.4	20.0
	5th	51.6	18.1	49.5	17.1	50.0	16.3
	% ch.	21.6		14.1		5.6	

B: Average Differences, Correlations, paired (dependent sample) t-scores, and effect sizes

	Ave. Dif.	r	t	p	N	effect size
Direct Instruction						
Vocabulary	8.2	0.55	9.09	<.001	490	0.39
Comprehension	9.2	0.53	11.05	<.001	494	0.48
Variable Curricula						
Vocabulary	1.7	0.47	3.84	<.001	2143	0.09
Comprehension	6.1	0.53	15.74	<.001	2175	0.33
Open Court						
Vocabulary	-3.8	0.47	-8.38	<.001	1841	-0.20
Comprehension	2.6	0.49	6.12	<.001	1843	0.14

C: Percentile Score of Average Student by School Type, First and Fifth Grade

		Direct Instruction	Variable Curricula	Open Court
Vocabulary	1st	38	45	58
	5th	53	48	51
Comprehension	1st	36	38	45
	5th	53	49	50

D: Results of Repeated Measures Analysis of Variance

Vocabulary	F	df	p
Achievement (repeated)	29.73	1, 4471	<.001
Group	37.33	2, 4471	<.001

Group by Achievement	81.06	2, 4471	<.001
Comprehension	F	df	p
Achievement (repeated)	315.77	1, 4509	<.001
Group	9.76	2, 4509	<.001
Group by Achievement	32.04	2, 4509	<.001

Note: The percentiles in Panel C correspond to the average (mean) NCE values given in Panel A. They may be interpreted as the percentile at which an average student in a given group and grade would score. Median values for all scores are also given in the Appendix to this article. Simple one-way analysis of variance comparing scores between the three groups for each measure and grade also indicate significant differences: for 1st grade vocabulary, $F = 75.991$, $p<.001$; for 1st grade comprehension, $F = 24.528$, $p < .001$; for 5th grade vocabulary, $F = 6.69$, $p = .001$; for 5th grade comprehension, $F = 2.98$, $p = .05$.

Table 2

Mixed Model Regressions, Fifth Grade Reading (norm equivalent scores) on School Advantage/Disadvantage, First Grade Achievement, and School Type (DI, Varied First Grade Curriculum, and Open Court)

Vocabulary

	Model 1			Model 2			Model 3			Model 4		
	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	49.45	0.59	<.0001	63.21	2.50	<.0001	64.17	2.49	<.0001	37.47	2.65	<.0001
School % FRL	-----	-----	-----	-18.70	3.33	<.0001	-19.20	3.32	<.0001	-12.10	3.41	0.0004
DI	-----	-----	-----	-----	-----	-----	0.98	1.54	0.52	5.16	1.56	0.001
Varied 1st grade	-----	-----	-----	-----	-----	-----	-1.55	0.54	0.004	1.30	0.48	0.007
1st grade ach.	-----	-----	-----	-----	-----	-----	-----	-----	-----	0.40	0.01	<.0001
Random Effect (schs.)	28.32	4.96	<.0001	19.89	3.80	<.0001	19.35	3.73	<.0001	22.11	4.05	<.0001
Residual	283.85	6.03	<.0001	283.89	6.0352	<.0001	283.41	6.03	<.0001	215.97	4.63	<.0001
- 2 Log Likelihood	38602.1			38574.8			38565.2			36913.6		
BIC	38616.1			38593.5			38593.2			36946.2		
				change	df	p.	change	df	p.	change	df	p.
Change in LL				27.30	1	<.001	9.60	2	<.001	1651.60	1	<.001
Correlation ratio	0.09											

Comprehension

	Model 1			Model 2			Model 3			Model 4		
	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	49.37	0.51	<.0001	62.10	2.11	<.0001	62.46	2.10	<.0001	36.65	2.23	<.0001
School % FRL	-----	-----	-----	-17.30	2.82	<.0001	-17.74	2.80	<.0001	-10.89	2.86	0.0001
DI	-----	-----	-----	-----	-----	-----	1.37	1.32	0.30	3.27	1.32	0.01
Varied 1st grade	-----	-----	-----	-----	-----	-----	-0.45	0.52	0.38	1.36	0.45	0.002
1st grade ach.	-----	-----	-----	-----	-----	-----	-----	-----	-----	0.44	0.01	<.0001
Random Ef. (schs.)	20.16	3.68	<.0001	12.90	2.70	<.0001	12.41	2.65	<.0001	14.69	2.83	<.0001
Residual	262.91	5.56	<.0001	262.96	5.57	<.0001	262.98	5.57	<.0001	191.19	4.08	<.0001

- 2 Log Likelihood	38546	38514			38512			36654		
BIC	38560	38533			38540			36687		
		change	df	p.	change	df	p.	change	df	p.
Change in LL		31.70	1	<.001	2.30	2		1858.10	1	<.001
Correlation ratio	0.07									

a. No coefficient is included for Open Court because it is the omitted category for the dummy variables.

From: Sanchez, Susan [Susan.Sanchez@ed.gov]
Sent: Friday, September 05, 2008 10:40 AM
To: Mark Dynarski; Jill Constantine; Scott Cody; Michael Ponza

Dear Dr. Decker:
This note is a follow-up to my letter of June 25 regarding the efforts of the What Works Clearinghouse and your note of August 6 indicating that you would respond to my concerns. As you may know, in the interim, the What Works Clearinghouse issued a report on Reading Mastery, one of the Direct Instruction programs, claiming that they could find "no studies…that meet WWC evidence standards." I am in the process of preparing a complete response to their analysis, but was dismayed to find that errors that I pointed out to you in my original letter remain in the new posting. To continue to display such erroneous conclusions so prominently on the WWC website does nothing to help, and potentially much to hurt, the achievement of our nation's children. As I suggested in my letter of June 25, until the procedures of the WWC can be altered to conform to minimal standards of scientific inquiry, I believe that all reviews on the website should be removed.

In addition, your note of August 6 mentioned that contractual relationships between Mathematica and the WWC may limit Mathematica's ability to prompt changes in the work of WWC. We have received a copy of the contract. Examination of the contract indicates that, rather than prohibiting WWC from taking some of the steps I outlined, it, in fact, supports taking several of these steps. For instance, on page 5 of the statement of work, section B (WWC Systematic Review Process), the document refers to WWC reviews "of the full body of evidence" on a particular topic area or intervention: "The contractor shall examine the efficiency of carrying out each of the steps in the systematic review of the full body of evidence." Yet the WWC has eliminated a large proportion of the research studies on Reading Mastery and other Direct Instruction interventions by examining only the last 20 years of research. The section on the WWC Systematic Review Process indicates that Mathematica has the responsibility to develop a review protocol "that specifies the…explicit inclusion and exclusion criteria to assess each potential study" and to identify the topic area boundaries, including "recency." The client, IES, already explicitly granted to the contractor, Mathematica, flexibility in making the recommended changes.

Moreover, section A of the statement of work (WWC Operating Assumptions and Guidelines), indicates that the WWC was designed to assure transparency and accuracy in reporting research reviews. One assumption/guideline listed on page 4 (Quick and Accurate Resolution of Complex, Technical Issues) warns that "WWC's systematic reviews surface many complex, technical issues in current research in education that, if left unresolved, would lead to inaccurate statements about the impact of an intervention." The report on Reading Mastery violates this guideline by conveying a false impression about the extent and quality of the research base on Reading Mastery. In addition, the very slow reaction to the issues I raised months ago in my first letter violates the

quick-and-accurate provision.
Another assumption/guideline listed on page 4 is that "The WWC review process, outcomes, and the products that result from the review process are clearly 'transparent' to assure accountability and easy access to the review process and its rationale." The wildly inconsistent treatment of research studies on different interventions, which I discussed in my previous letters, indicates a puzzling lack of transparency in the WWC review process.
In short, the concerns regarding the WWC that I expressed in my June 25 letter remain and, in fact, have intensified in the ensuing months. My review of the contract language reinforces my belief that Mathematica is responsible for dealing with these very serious problems. I am copying Dr.
Summers, as well as Dr. Grover Whitehurst, IES Director, for their information. I presume that I will hear from you soon in response to my concerns.
Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

Fourth Graders' Growth in Reading Comprehension:

A Pretest-Posttest Randomized Control Study Comparing Reading Mastery and Scott Foresman Basal Reading Program

December 31, 2010

Jean Stockard, Ph.D.
Director of Research,
National Institute for Direct Instruction
Eugene, Oregon
jstockard@nifdi.org

Fourth Graders' Growth in Reading Comprehension:
A Pretest-Posttest Randomized Control Study Comparing Reading Mastery and Scott Foresman

Purpose:
The purpose of this study was to compare growth in reading comprehension of fourth grade general education students in a Midwestern elementary school.

Sample and Design:
In the spring of 2009, general education third graders were randomly assigned to a control or treatment condition, using an alternative assignment with random start technique. Students were predominantly non-Hispanic whites from middle income families. Beginning in the fall of 2009, students in the treatment condition received instruction in the SRA McGraw Hill program, Reading Mastery (RM) Signature Edition. Students in the control condition received instruction in the Scott Foresman basal reading program, which the school had been using in prior years. Both groups received instruction for ninety minutes a day. Each group had two teachers (four teachers in all). Teachers were randomly assigned to teach the different groups. The analysis reported here is limited to 54 students (28 in control and 26 in RM) with data available for all testing periods. Identical results appear when those with missing data at one point are included.

Measures and Analysis:
The curriculum-based AIMSweb measurement system, which has well-established validity and reliability, was used to gather data on reading comprehension in the form of number of correct answers on a maze test. Data were gathered in the spring of 2009 before instruction began as a baseline measure, in the fall of 2009 shortly after the start of the school year and again in spring of 2010. Mean scores of students in the control and treatment group at the spring baseline were compared using a Student t-test. Differences in fall and winter scores were analyzed using t-tests, effect sizes, and repeated measures analysis of variance.

Results:
Table 1 gives descriptive statistics for each group and the associated t-test for the spring 2009 maze measure. The results indicate that there were no significant differences between the groups at the time of assignment to RM or control.

Table 2 gives descriptive statistics for fall and spring, t-tests comparing the groups at each time period, associated effect sizes (Cohen's d) and the results of a repeated measures analysis of variance with fall and spring scores as a repeated measure and group as factor. The interaction between group assignment and change over time was statistically significant, indicating that the pattern of change was different between the two groups. The effect size in the fall was close to zero (d = .002), but well above the level generally considered educationally important by spring (d = .42). Figure 1 illustrates the different patterns of change and how the increase in comprehension, as measured by the maze, was stronger for the RM students than for the control students.

Conclusion
The results of this randomized trial replicate numerous other studies that have found that students receiving instruction in Reading Mastery have significantly greater growth in reading comprehension than students studying in other curricula.

Table 1

AIMSWeb Maze Scores, Spring 2009, by Treatment Group

Control Group			Reading Mastery Group				
Mean	S.D.	N	Mean	S.D.	N	t	prob.
19.04	5.04	28	19.38	5.15	26	-.25	.80

Note: Probability is two-tailed. Sample is limited to students who also had data for fall, 2009 and spring, 2010.

Table 2

AIMS Web CBM Words Read Correct, Fall 2009 and Winter 2010 by Treatment Group and Repeated Measures Analysis of Variance

	Control		Reading Mastery		t-value	prob.	Effect size
	Mean	S.D.	Mean	S.D.			
Fall 2009	16.9	3.8	16.9	3.4	-.01	0.50	-0.002
Spring, 2010	25.4	7.3	28.5	7.2	1.53	0.07	0.42
	Change Over Time		Group Differences		Interaction		
	F	prob.	F	prob.	F	prob.	
Anova Results	158.66	<.001	1.27	.27	3.63	0.06	

Note: Only students with spring, 2009 pretest scores were included. N = 28 for control group and 26 for RM group. T-test probabilities are one-tail, reflecting the hypothesis of interest. Effect sizes are Cohen's d.

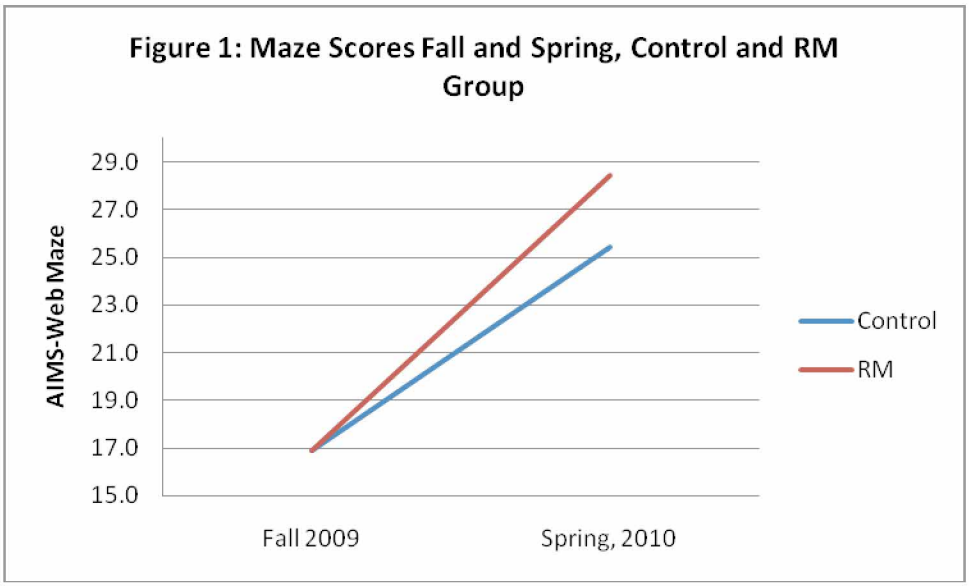


Figure 1: Maze Scores Fall and Spring, Control and RM Group
AIMS-Web Maze
29.0
27.0
25.0
23.0
21.0
19.0
17.0
15.0
Fall 2009
Spring, 2010
Control
RM

From: Sanchez, Susan [Susan.Sanchez@ed.gov]
Sent: Friday, January 16, 2009 9:22 AM
To: Mark Dynarski; Jill Constantine; Scott Cody; Michael Ponza
Subject: FW: complaint about WWC
Attachments: 090102 Hanushek Stockard.pdf; rmresponse_wwc.pdf

Importance: High

-----Original Message-----
From: Cottingham, Phoebe
Sent: Thursday, January 15, 2009 6:42 PM
To: Sanchez, Susan; 'mdynarski@mathematica-mpr.com'; 'jconstantine@mathematica-mpr.com'
Subject: Fw: complaint about WWC

Fyi

Sent using BlackBerry

From: (b)(6)
To: Cottingham, Phoebe
Cc: Betka, Sue; Garza, Norma
Sent: Thu Jan 15 18:11:06 2009
Subject: complaint about WWC
Phoebe

Interestingly, while in DC for the discussion of WWC, I received the attached letter and paper. Can you tell me what you know about this matter? They claim not to have been properly treated.

Rick

--
Eric Hanushek
Hoover Institution
Stanford University
Stanford, CA 94305-6010
650-736-0942
www.hanushek.net

+/v8APAAh-DOCTYPE HTML PUBLIC +ACI-//W3C//DTD HTML 4.0 Transitional//EN+ACIAPg- +ADw-HTML+AD4APA-HEAD+AD4- +ADw-META http-equiv+AD0-Content-Type content+AD0AIg-text/html+ADs- charset+AD0-utf-7+ACIAPg- +ADw-META content+AD0AIg-MSHTML 6.00.6000.16544+ACI- name+AD0-GENERATOR+AD4APA-/HEAD+AD4- +ADw-BODY+AD4APA-B+AD4-From:+ADw-/B+AD4- Mark Dynarski+ADw-BR+AD4APA-B+AD4-Sent:+ADw-/B+AD4- Sunday, December 02, 2007 4:25 PM+ADw-BR+AD4APA-B+AD4-To:+ADw-/B+AD4-whatworks+AEA-ed.gov+ADw-BR+AD4APA-B+AD4-Subject:+ADw-/B+AD4- FW: Questions on What Works Clearing House Procedures+ADw-BR+AD4- +ADw-DIV dir+AD0-ltr align+AD0-left+AD4APA-FONT face+AD0-Arial color+AD0AIw-0000ff size+AD0-2+AD4APA-SPAN class+AD0-980102121-02122007+AD4-This came to me but it's not a practice I want to encourage.+ADw-/SPAN+AD4APA-/FONT+AD4APA-/DIV+AD4- +ADw-DIV dir+AD0-ltr align+AD0-left+AD4APA-FONT face+AD0-Arial color+AD0AIw-0000ff size+AD0-2+AD4APA-SPAN class+AD0-980102121-02122007+AD4APA-/SPAN+AD4APA-/FONT+AD4AJg-nbsp+ADsAPA-/DIV+AD4- +ADw-DIV dir+AD0-ltr align+AD0-left+AD4APA-FONT face+AD0-Arial color+AD0AIw-0000ff size+AD0-2+AD4APA-SPAN class+AD0-980102121-02122007+AD4-Mark+ADw-/SPAN+AD4APA-/FONT+AD4APA-/DIV+AD4- +ADw-DIV dir+AD0-ltr align+AD0-left+AD4APA-FONT face+AD0-Arial color+AD0AIw-0000ff size+AD0-2+AD4APA-SPAN class+AD0-980102121-02122007+AD4APA-/SPAN+AD4APA-/FONT+AD4AJg-nbsp+ADsAPA-/DIV+AD4APA-BR+AD4- +ADw-DIV class+AD0-OutlookMessageHeader lang+AD0-en-us dir+AD0-ltr align+AD0-left+AD4- +ADw-HR tabIndex+AD0--1+AD4- +ADw-FONT face+AD0-Tahoma size+AD0-2+AD4APA-B+AD4-From:+ADw-/B+AD4- Jean Stockard +AFs-mailto:jeans+AEA-uoregon.edu+AF0- +ADw-BR+AD4APA-B+AD4-Sent:+ADw-/B+AD4-Sunday, December 02, 2007 12:30 PM+ADw-BR+AD4APA-B+AD4-To:+ADw-/B+AD4- Mark Dynarski+ADw-BR+AD4APA-B+AD4-Subject:+ADw-/B+AD4- Questions on What Works Clearing House Procedures+ADw-BR+AD4APA-/FONT+AD4APA-BR+AD4APA-/DIV+AD4-+ADw-DIV+AD4APA-/DIV+AD4-Dear Dr. Dynarski:+ADw-BR+AD4AJg-nbsp+ADsAPA-BR+AD4-I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction (NIFDI). One of my first tasks was to read as much of the background research in the area as I could. Along the way I ran into the +IBw-What Works Clearinghouse.+IB0- In my review I developed several questions that I am hoping you can answer.+ADw-BR+AD4AJg-nbsp+ADsAPA-BR+AD4-First, I noticed that the WWC has listed relatively few of the studies that deal with Direct Instruction. While a few were rejected because they didn+IBk-t meet some specific standards of design, it appears that many others were rejected because they were published before 1985. I have not been able to find a scientific justification for selecting this date. No studies were cited that suggested children+IBk-s learning styles had altered after that time or that schools+IBk- organizational structures or teachers+IBk- actions had changed. In fact, the only research article that I found to justify the methods was actually published in 1974 +IBM- 10 years prior to the beginning of your targeted dates. This article dealt with phonological awareness, and had nothing to do with systematic literature reviews and, of course, included nothing to suggest that processes related to phonological awareness have changed over time. +ADw-BR+AD4AJg-nbsp+ADsAPA-BR+AD4-In all the other areas with which I am familiar we use the full range of research articles and, unless there are clear reasons established through the research literature, do not set an arbitrary date for inclusion or exclusion of results.+ACY-nbsp+ADs- The choice of this date

seems to have particularly affected the DI literature because a great deal of the pioneering work was published before the cut-off point. Because I+IBk-m new to this particular work, I wondered if you could provide me with any explanation based in the research literature for your choice of this cut-off year. I did find a suggestion that, +IBw-if sufficient time and resources remain,+IB0-studies published before 1985 might be reviewed. Do you know if this step will be taken? From my reading of the literature I believe that your limit on dates of studies may have seriously limited the selection of effective resources that educators will find on your website+ADs- and I know that this is a primary aim of your project.+ADw-BR+AD4AJg-nbsp+ADsAPA-BR+AD4-Second, I noticed that a number of studies were rejected because they included children outside the K-3 grade range. I found this somewhat unusual, for the children that are most in need of effective instruction are those who have not learned to read by older ages. Because the effectiveness of DI has been so well demonstrated with regular students at the K-3 grade range, the more recent focus on other ages seems only greater evidence of its worth. This is especially so in the case of special education students and those with behavioral and other issues. Thus, I couldn+IBk-t understand the rationale of omitting all the studies that targeted these most needy students. Any explanation you might have of this decision would also be greatly appreciated.+ADw-BR+AD4AJg-nbsp+ADsAPA-BR+AD4-Third, and perhaps most important, as I read studies more closely I found that the Clearinghouse+IBk-s interpretation of some of the studies didn+IBk-t seem to accurately reflect the actual content of the reports. I know that these reviews are often done by staff members or graduate students who might not be familiar with all of the intricacies of social research. Thus, I assume that we can send corrections and they will be added. +ADw-BR+AD4AJg-nbsp+ADsAPA-BR+AD4-I look forward to hearing from you, either through e-mail (+ADw-A href+AD0AIg-mailto:jeans+AEA-uoregon.edu+ACIAPg-jeans+AEA-uoregon.edu+ADw-/A+AD4-) or by phone (541-346-5005). I+IBk-m sure that we share the ultimate goal of helping all students +IBM- no matter what their background +IBM-develop academic skills and appreciate your appreciate your help.+ADw-BR+AD4AJg-nbsp+ADsAPA-BR+AD4-Sincerely, +ADw-BR+AD4APA-BR+AD4APA-X-SIGSEP+AD4-+ADw-P+AD4APA-/X-SIGSEP+AD4-Jean Stockard, Professor Emerita+ADw-BR+AD4-Department of Planning, Public Policy, and Management+ADw-BR+AD4-University of Oregon+ADw-BR+AD4-Eugene, Oregon 97403-1209+ADw-BR+AD4-phone: 541-346-5005+ADw-BR+AD4-fax: 541-346-2040 +ADw-/P+AD4APA-/BODY+AD4APA-/HTML+AD4-

The Long-Term Impact of NIFDI-Supported Implementation of Direct Instruction on Reading Achievement: An Analysis of Fifth Graders in the Baltimore City Public School System

Jean Stockard
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

September 2008

Technical Report 2008-02



© 2008 National Institute for Direct Instruction
PO Box 11248 • Eugene, OR 97440
1-877-485-1973

Table of Contents

	Page
Executive Summary	iii
List of Tables	iv
List of Figures	v
Report	1
Methodology	1
Results	3
Summary	7
Appendix I: The Sample and Measures	8
Measuring Student Achievement	8
Composite Measures of School Context	8
The Panel Sample	11
Appendix II: Detailed Statistical Results	17
Analysis	17
Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Reading Achievement from First to Fifth Grade than Students in Other Schools?	18
Research Question Two: Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?	26
Research Question Three. What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?	35
References	36

The Long-Term Impact of Direct Instruction on Reading Achievement: An Analysis of Fifth Graders in the Baltimore City Public School System

Executive Summary

In the mid 1990s, in response to very low reading achievement scores, the Baltimore City Public School System (BCPSS) implemented curricular reforms. Sixteen schools used Direct Instruction. All the Direct Instruction schools used the *Reading Mastery Classic, Language for Learning, Language for Thinking,* and *Reasoning and Writing* curricula. While all of these schools initially received technical support from the National Institute for Direct Instruction, five eventually ceased this support and used an alternative provider. Beginning in 1998 schools in the district that were not using Direct Instruction adopted Open Court as a standard reading curriculum. Scores were available from the Comprehensive Test of Basic Skills (CTBS) for all first graders and all fifth graders in BCPSS in the spring of 6 school years: 1997-1998 through 2002-2003.

This report examines the achievement of two BCPSS cohorts of students who were enrolled in the Baltimore City Public School System in both first grade and fifth grade during this time: in first grade in 1997-98 and fifth grade in 2001-2003 or in first grade in 1998-99 and fifth grade in 2002-2003. Almost 5,000 students who had data for both years were included in the analysis. The study compared students' achievement in 1) schools with NIFDI support, 2) schools that implemented Direct Instruction without continual NIFDI support (other DI schools), and 3) schools without Direct Instruction (the Control schools). Schools in the three groups had similar demographic characteristics, although one of the Other DI schools had a markedly lower rate of poverty and a substantially lower rate of student turnover. Data for this low poverty DI school were also analyzed separately.

Key findings are listed below.

- Students who received Direct Instruction in first grade had significantly higher achievement scores in fifth grade than other students. The results were strongest with the measures of vocabulary and the composite score.
- While students in all schools had higher average achievement in fifth grade than in first grade, the increases were strongest for students in the NIFDI-supported schools and in the low-poverty Other DI school. On average, students in NIFDI-supported schools had a 25 percent gain in their composite reading achievement scores from first grade to fifth grade. Students in the control schools had a gain of only 5 percent.
- The magnitude of changes from first to fifth grade far surpassed the traditional criterion of educationally important effects for both the NIFDI-supported school and the low poverty school in the Other DI group.

List of Tables

Title	Page
A-1. Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels, Total and by Treatment Group, BCPSS	9
A-2. Factor Analysis of School Characteristics, Elementary Schools, BCPSS, 1998-2003	10
A-3. Schools in Treatment Groups and Start Dates of Reading Direct Instruction	12
A-4. Turnover in the BCPSS, 1997-98, 1998-99, to 2001-02, 2002-03	12
A-5. Percentage of Students Who Persisted in the Same Schools from First Grade to Fifth Grade by Group, BCPSS, 1998-2003	13
A-6. Comparison of Panel Group to Other First Graders, First Grade Reading Achievement	14
A-7. School Characteristics by Presence in Panel Sample and Group	15
A-8: Descriptive Statistics and Paired t-test Results, 1st and 5th Grade Reading Achievement, BCPSS, 1998-2003	19
A-9: Independent t-tests, Average Reading Achievement, DI and Control Schools, 1st and 5th Grade, BCPSS, 1998-2003	20
A-10. Repeated Measures Analysis of Variance, Norm Equivalent Reading Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools	21
A-11. Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction	22
A-12. Coefficients Associated with Mixed Model Results, Fifth Grade Reading Regressed on First Grade Reading, School Advantage/Disadvantage, and Attending a DI School	24
A-13. Simulated Fifth Grade Achievement Scores (Norm Equivalent and Percentile of Hypothetical Average Student), DI and Control Schools, Controlling for School Context and First Grade Achievement, BCPSS, 2001-03	25
A-14. Descriptive Statistics and Paired t-test Results, 1st and 5th Grade Reading Achievement Scores, NIFDI-Supported Schools, Other DI Schools, and Control Schools, BCPSS, 1998-2003	27

A-15. One-Way Analysis of Variance, Average Reading Achievement, Norm Equivalent Scores, NIFDI, Other DI, and Control Schools, 1st and 5th Grade, BCPSS, 1998-2003 29

A-16. Repeated Measures Analysis of Variance, Normal Curve Equivalent Scores, Reading 30

A-17. Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction, Comparing NIFDI Supported Schools, Other DI Schools (with and without low poverty school) and Control Schools 31

A-18. Mixed Model Regressions, Fifth Grade Reading (Norm Equivalent Scores) on School Advantage/Disadvantage and School Type 32

A-19. Effect Sizes of Change, First to Fifth Grade, by Group Reading Achievement, Norm Equivalent Scores 35

List of Figures

Page

1. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average Student, First Grade and Fifth Grade, DI and Control Schools, BCPSS 3

2. Percentage Change in Average Achievement Scores, First to Fifth Grade, DI and Control Schools 4

3. First and Fifth Grade Composite Achievement, NIFDI-Supported, Other DI and Control Schools 5

4. Percent Change in Average Achievement, First to Fifth Grade, NIFDI-Supported, High Poverty Other DI Schools, Low Poverty Other DI Schools, and Control Schools 6

5. Effect Sizes of Change from First to Fifth Grade in Vocabulary, Comprehension and Composite Reading Achievement Scores, NIFDI-Supported, High Poverty Other DI Schools, Low Poverty Other DI School, and Control Schools 7

The Long-Term Impact of Direct Instruction on Reading Achievement: An Analysis of Fifth Graders in the Baltimore City Public School System

Low reading achievement of elementary school students has been a major concern of school districts throughout the country. The Baltimore City Public School System (BCPSS) is similar to many other large city school districts that serve students with high levels of poverty and struggle with low levels of achievement. In the mid 1990s, the BCPSS implemented curricular reforms throughout its elementary schools to address this low achievement.

The reforms introduced in Baltimore provided the opportunity for a natural experiment that could compare the effects of different curricula on student achievement. Some of the schools used Direct Instruction for reading instruction, specifically the *Reading Mastery Classic, Language for Learning, Language for Thinking,* and *Reasoning and Writing* curricula. Other schools in the district adopted the Open Court Reading curriculum. While all of the schools using Direct Instruction started implementation with the National Institute for Direct Instruction (NIFDI), six of these schools ceased involvement with NIFDI after implementation and began to work with another provider. These schools are termed "Other DI schools" in this report.

A very large body of literature, including well designed meta-analyses, has demonstrated the effectiveness of Direct Instruction in promoting high reading achievement. A somewhat smaller literature has addressed the impact of DI instruction over the span of the elementary years. This paper addresses the question of longer-term impact by examining the change in reading achievement for two cohorts of BCPSS students: those who were in first grade in 1997-98 and fifth grade in 2001-02 and those who were in first grade in 1998-99 and fifth grade in 2002-2003. Three questions were addressed: 1) Did students who received Direct Instruction in first grade have greater gains in reading achievement from first to fifth grade than students in other schools? 2) Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools? and 3) What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?

Results are summarized in graphical form in this report, while appendices provide full methodological details and statistical results. A brief overview of the methodology is below followed by the results with each research question.

Methodology

The analysis uses data from almost 5,000 students who were first graders in the Baltimore City Public School System in 1997-98 or 1998-99 and fifth graders in 2001-2 or 2002-2003. Measures of reading achievement were obtained each spring, using the Comprehensive Test of Basic Skills (CTBS), a widely used standardized test. Subtest measures of Vocabulary and Comprehension, as well as a Composite measure were available. Results in this part of the report are presented as percentiles. Percentiles can be simply interpreted as the percentage of students that would have scores lower than a given student. For instance, if a student has a score at the 60th percentile, 60 percent of all students had scores that were lower.

The average school in the BCPSS had large proportions of low income and minority students. On average, 75 percent of the students in a school were on free or reduced lunch. This varied however, from a low of 22 percent to a high of 93 percent. On average 84 percent of the students in average elementary school were African American, although the percentage varied from a minimum of 3 to almost 100 percent. Reflecting this high level of segregation, schools also varied in their representation of non-Hispanic white students, from having virtually no white students to being 94 percent white, with the average school having 14 percent of its students reported as being non-Hispanic white. Asian, Hispanic and Native American students were quite rare in the district, with an average representation of 1 percent in the student bodies.

The NIFDI-supported schools were similar to the Other DI schools and the other BCPSS schools in both racial-ethnic composition and rates of poverty. There were no significant differences between the three groups in the average proportion of students receiving free and reduced lunch or the representation of African American or non-Hispanic white students. The NIFDI schools, however, had significantly higher proportions of Hispanic and Native American children. In addition, one of the schools in the Other DI group had a much lower rate of poverty. On average, over the years in the study, only about one-third of the students in this school received free or reduced lunch. In all of the other schools that received Direct Instruction, three quarters to ninety percent of the students had free or reduced lunch.

There was substantial student turnover in the BCPSS in the targeted years. Of students who were in first grade in the years under study, only 27 percent remained in the same school in fifth grade. This percentage was not significantly different between the three treatment groups. However, the low-poverty school in the Other DI group had a substantially higher rate of persistence, with 57 percent of its first graders still in the school at the end of fifth grade. In addition, although there was no difference across the three groups in the percentage of students who remained in the same schools for the five years, the characteristics of those who remained differed significantly between the three groups. Students who remained in the NIFDI- supported schools were more likely than students in the other schools to have lower achievement levels in first grade. Note that these differences would produce conservative estimates of the impact of NIFDI support on changes in achievement, for students in the NIFDI-supported schools began first grade under arguably more difficult socio-demographic circumstances than students in the other schools. (Appendix I provide additional details on the sample and measures.)

The analysis focused on comparing students' achievement in first grade with their achievement in fifth grade and the ways in which these differences varied across the three groups of schools: those that had NIFDI support throughout the elementary careers of the students, those that only had NIFDI support in the early years, and the Control schools that did not implement Direct Instruction. Full details on the statistical techniques and results are given in Appendix II.

Three research questions were addressed:

- Did students who received Direct Instruction in first grade have greater gains in reading achievement from first to fifth grade than students in other schools?
- Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?

- What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?

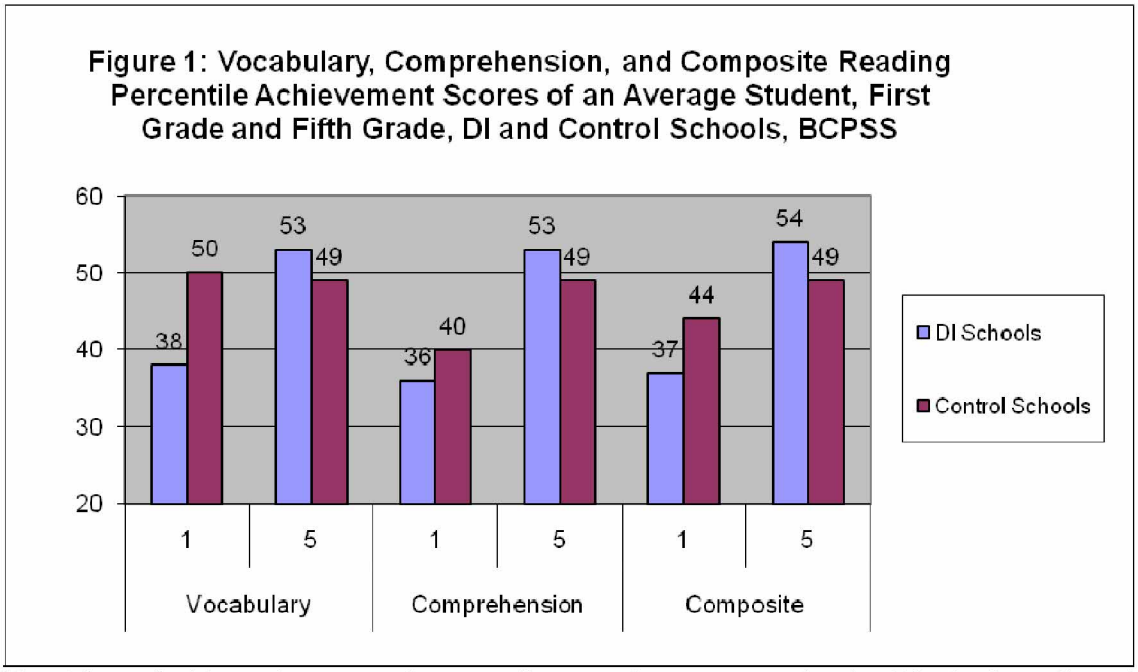
Results

A general summary of the results related to each of the research questions is given in this section. Each research question is addressed in turn.

1. Did students who received Direct Instruction in first grade have greater gains in reading achievement from first to fifth grade than students in other schools?

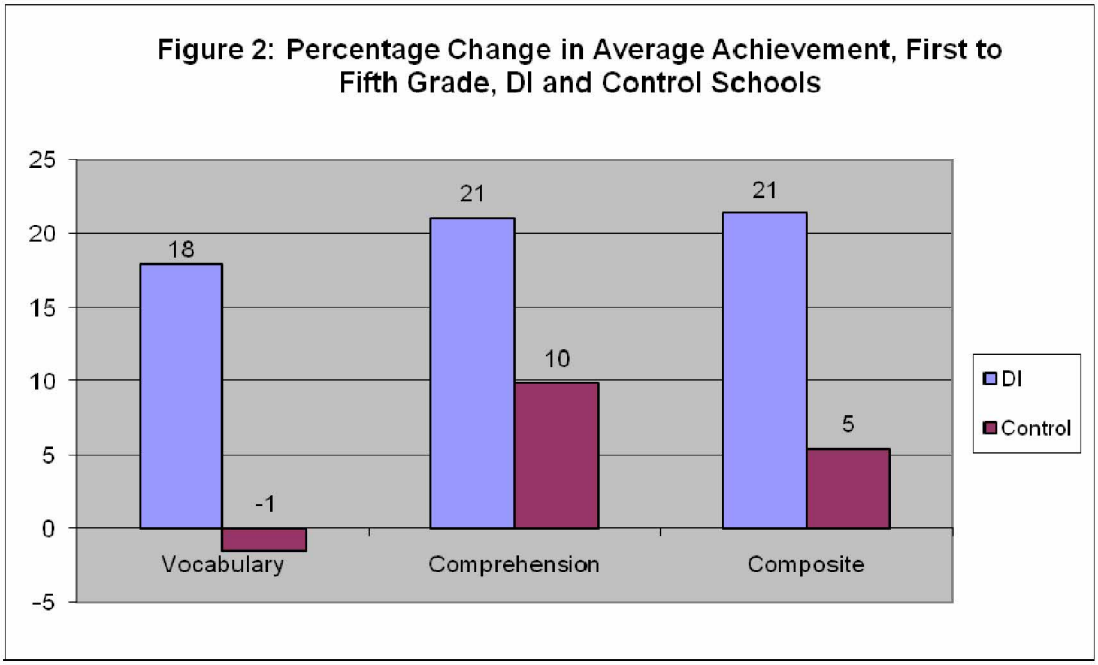
Changes in achievement from first to fifth grade were much stronger for students in the DI schools than for students in the Control Schools. Several statistical tests were used to examine this question, and the results confirmed that the changes from first to fifth grade were significantly stronger for students in the DI schools.

Figure 1 illustrates the magnitude of these results. It displays the percentile scores of the average student in both the Direct Instruction and the Control Schools for both first and fifth grade. In first grade, the average student in a DI school had achievement scores that were significantly lower than those of students in the Control Schools. By fifth grade, this pattern had reversed, and the students in the DI schools had scores that were significantly higher than those of students in the Control school. For instance, an average student in a Direct Instruction school had a composite reading achievement score at the 37th percentile in first grade, but a score at the 54th percentile in fifth grade. In contrast, the average student in the control schools moved from the 44th percentile in first grade to the 49th percentile in fifth grade.



Note: Figures in this graph represent the percentile score of an average student in a DI school or a Control school on each component of the CTBS in first and fifth grade. The group of DI Schools includes those that were NIFDI-supported as well as both the low poverty and high poverty Other DI schools.

Figure 2 summarizes the extent of these differences by showing the percentage of change in the average score of student from first grade to fifth grade. On average, for students in the DI schools, achievement scores rose from 18 to 21 percent from first to fifth grade. In contrast, although scores for students in the control schools rose on two of the three measures, the magnitude of this increase was less than half that of students in the DI schools.



Note: The data in this figure represent the percentage change in average scores from first to fifth grade, calculated as the difference between first and fifth grade divided by the first grade score. Calculations were made with norm equivalent scores. Appendix II provides further details.

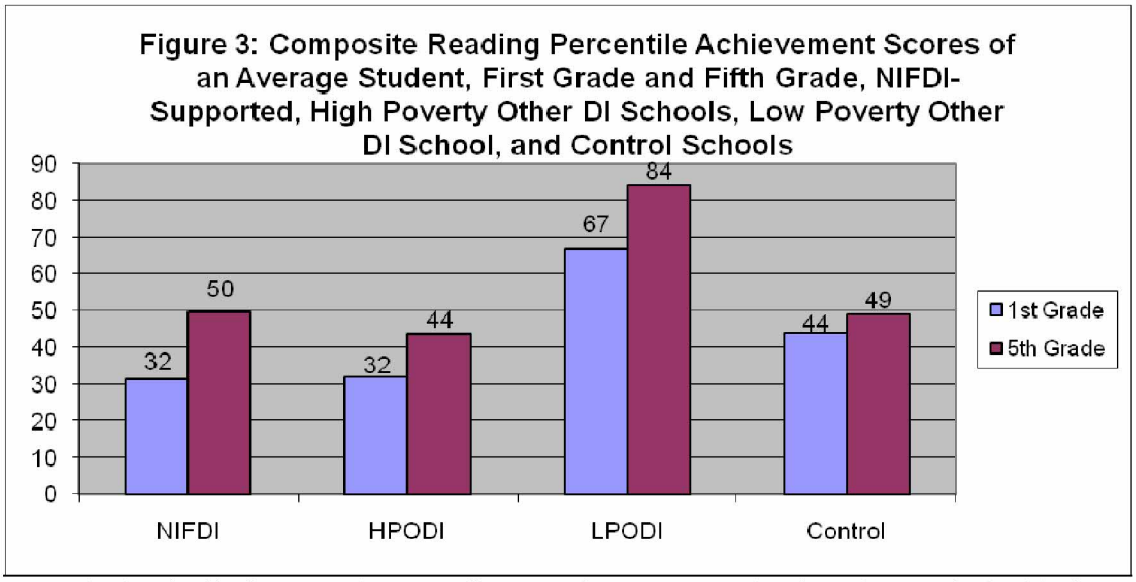
Data in Figures 1 and 2 are unadjusted for students' reading levels in first grade or the schools' contexts of disadvantage. We also used multivariate statistics to predict students' fifth grade achievement based on their achievement in first grade, the poverty context of their schools, and whether or not they had Direct Instruction in first grade. The results confirmed the finding that students who had Direct Instruction in first grade had higher achievement scores in fifth grade than in first grade and that these higher scores occurred independently of the poverty context of the school and their level of first grade achievement. These results were strongly significant for the measures of both vocabulary and the composite score. With the measure of comprehension the results were in the same direction, but just shy of traditional levels of significance.

2: Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?

All of the students in the sample who received DI in first grade were in schools that were receiving technical support from the National Institute of Direct Instruction at that time. However, only students in the NIFDI-supported schools were in schools with this support throughout their elementary career. Various statistical analyses were used to examine differences in gains from first to fifth grade for students in the NIFDI supported schools, the Other DI schools, and the Control Schools.

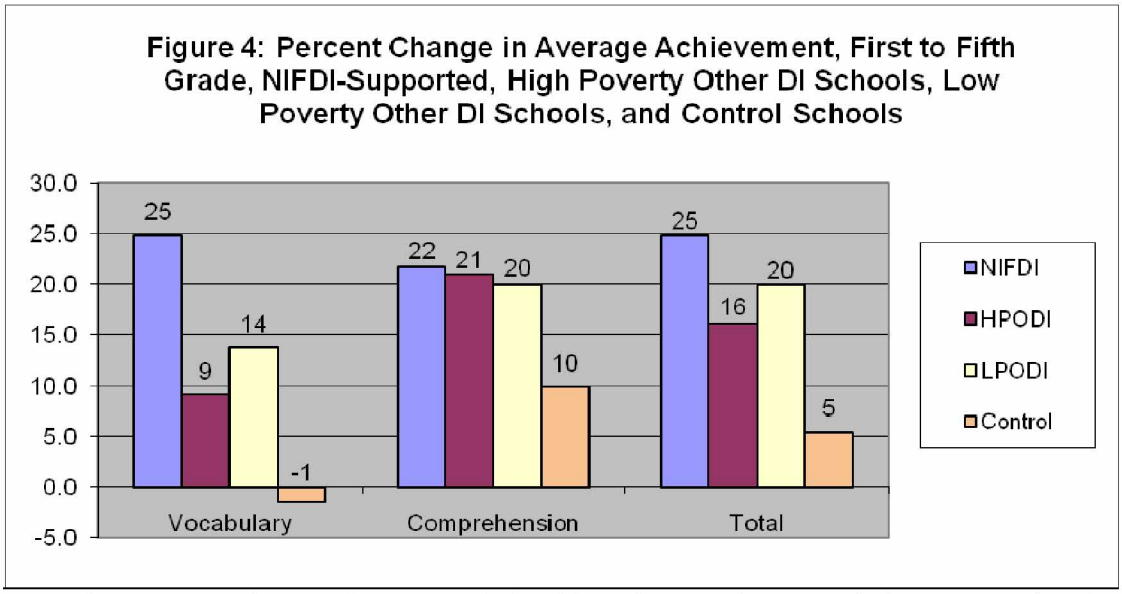
Figure 3 gives the percentile scores on the composite measure of reading achievement in first and fifth grade for an average student in 1) NIFDI supported schools, 2) the high poverty Other DI (HPODI) schools, 3) the low poverty school in the Other DI (LPODI) group, and 4) the control schools. Students in the LPODI had the highest achievement scores in first grade, followed by the students in the Control Schools. An average student in the LPODI school had a composite score at the 67th percentile in first grade, while an average student in the control schools had a composite score at the 44th percentile. Students in the demographically similar NIFDI-supported and HPODI schools had much lower first grade scores, with the average student scoring at the 32nd percentile.

By fifth grade the situation had changed. Average fifth grade achievement scores were higher than average first grade scores in all four groups of schools, and the LPODI school again had the highest achievement score, with an average student scoring at the 84th percentile. In contrast to results in first grade, the next highest average scores occurred with students in the NIFDI-supported schools, where an average student scored at the 50th percentile. These were followed closely by the Control Schools, where an average student scored at the 49th percentile. Students in the HPODI schools had the lowest average scores in fifth grade.



Note: The data in this figure are the percentile score of an average student in each type of school in first grade and fifth grade. HPODI refers to the high poverty Other DI schools, LPODI refers to the low poverty Other DI school.

Figure 4 shows the percentage change in scores from first to fifth grade for each of the four groups. The changes were largest for students in the NIFDI-supported schools for each of the measures of achievement. On the comprehension and composite measures the changes for the both the high poverty and low poverty Other DI schools were close to the values for the NIFDI-supported schools. In contrast, the average percentage change for students in the Control schools was substantially smaller, and even negative for the measure of vocabulary.

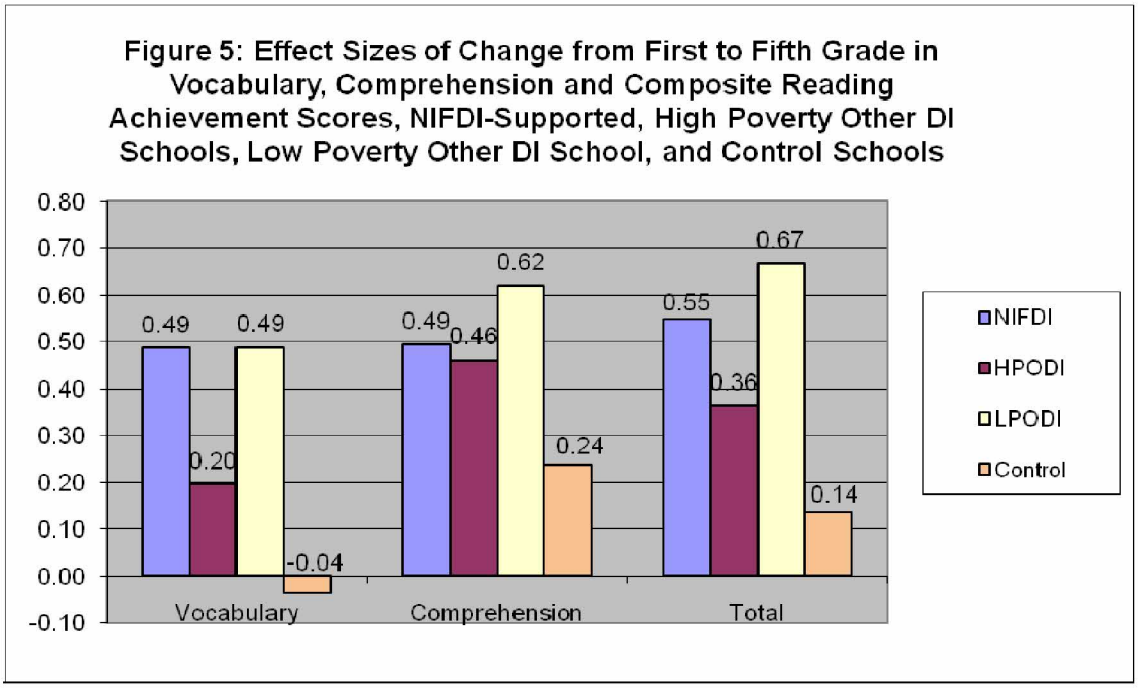


Note: The percentage change values were calculated from the normal curve equivalent scores rather than the percentile scores such as those shown in Figure 3. They are the difference in average first and fifth grade scores divided by first grade scores. Details are in Appendix II. HPODI refers to the high poverty Other DI schools, LPODI refers to the low poverty Other DI school.

3. What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?

Standard tests of statistical significance are influenced by the size of a sample in the calculations. To avoid this bias researchers sometimes use measures of effect sizes. With data involving change over time these calculations consider both the size of the scores at the two time points as well as the correlation between these two measures. (Details are in Appendix II.) Traditionally, effect sizes of .25 or larger are considered educationally important.

Figure 5 gives the effect sizes associated with the changes from first to fifth grade for each of the measures of achievement and for the four groups considered in the analysis of research question two. All of the effect sizes associated with the NIFDI-supported schools and the LPODI school far surpass the .25 criterion, as do two of the scores for the HPODI school. Effect sizes for change for the NIFDI supported schools range from .49 (vocabulary and comprehension) to .55 (composite score). Those for the LPODI school, which had very low turnover, range from .49 to .67, and those for the HPODI schools range from .20 to .46. The effect sizes for changes for students in the Control Schools are substantially smaller, although the value for Comprehension approaches the .25 criterion (d = .24).



Note: HPODI refers to the high poverty Other DI schools, LPODI refers to the low poverty Other DI school.

Summary

Compared to other students in the BCPSS, those who were in schools with Direct Instruction in first grade had significantly higher achievement in 5th grade. These differences were both statistically significant and educationally important. On average, achievement test scores for students in DI schools rose from 18 to 21 percent over the 5 grades. In contrast, test scores for students in the Control schools rose by only 10 percent or less, and declined in one area.

While all of the students receiving DI were in schools supported by the National Association for Direct Instruction in first grade, several of those schools later ceased this involvement. Students in schools with NIFDI support throughout their elementary career had the highest percentage change in achievement scores from first to fifth grade. One of the schools that ceased involvement with NIFDI but continued using DI had a relatively low poverty rate and a very low rate of turnover of students throughout the study period. Students in this school had significantly higher achievement scores in both first and fifth grade than students in the other schools. They also had percentage gains in achievement that were second only to students in the NIFDI-supported schools and effect sizes that surpassed those of the NIFDI-supported schools.

The results presented in this analysis are important in showing the long-term impact of Direct Instruction on students' achievement as well as the role of continuous, high quality technical support for schools. Even though the students in the NIFDI-supported schools were in high poverty school environments and had lower first grade achievement than the other students, they had gains in achievement from first grade to fifth grade that were statistically significant and educationally important. The presence of NIFDI-supported Direct Instruction throughout their elementary career appears to have promoted significantly higher rates of learning. This pattern of results appeared in all the measures of reading achievement that were considered.

Appendix I
The Measures and Sample

This appendix provides details on the measures of achievement and school context used in the analysis and on the panel sample, the students who were in the same schools in both first and fifth grade.

Measuring Student Achievement

The BCPSS provided data on students' scores on the Comprehensive Test of Basic Skills for all first and fifth graders from the spring of 1998 to the spring of 2003. The Comprehensive Test of Basic Skills is a widely used standardized measure of academic achievement. The 4th edition was administered in the spring of 1998 and 1999 and the 5th edition was administered in the spring of 2000 through 2003. Two subtest scores, Reading Comprehension and Reading Vocabulary, and a Composite Reading Achievement score were available and analyzed.

Both raw scores and Normal Curve Equivalent (NCE) scores were provided. NCE scores are calculated by translating the percentile scores to a distribution that is normally distributed. Like percentile scores, NCE scores range from 1 to 99 with a mean of 50. However, while percentile scores are evenly distributed (the graph of percentile scores would look like a rectangle), NCE scores comprise a normal distribution. The transformation results in scores that can be meaningfully added and subtracted, so that the difference of an NCE score of 50 and 55 (=5) is the same as the difference between 30 and 35. This interval scale allows computations of most of the statistics used in analysis. This is not characteristic of percentiles or raw scores, and thus statisticians urge researchers to analyze data based on NCE scores. All statistical calculations for this report were done using NCE scores.

The meaning of NCE scores is, however, not intuitively obvious, and, for this reason, many of the results in the body of the text and the executive summary have been translated into percentile scores after the statistical calculations were completed. The resulting descriptive information (mean NCE score) was converted into percentiles for display in the graphs in the body of the report and display in the tables in the appendices. These conversions were done using a standard conversion table and are included in the tables of results in Appendix II.[1]

Composite Measures of School Context

Given the strong influence of school context on student achievement, it was important to develop an efficient, yet strong, measure of the demographic context of the schools in the sample. Preliminary analysis indicated that the demographic characteristics of schools were very highly correlated from one year to another. Thus, one summary measure was developed for each school that would be valid for all the years included in the data set.

There were a few cases in which data were not available for a school for all years. A regression-based method was used to predict values of missing cases from other years (e.g.

[1] The conversion can also be accomplished manually through using a normal curve table. The NCE scores can be converted to z-scores (z = ((nce-50)/21.06))). The percentile that corresponds to the z score can then be found in the normal curve table.

predicting 1998 levels of proportion black from levels in 1999 through 2003).[2] Values for each variable were averaged across the years to produce aggregate measures of the demographic characteristics of the school.

Table A-1: Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels, Total and By Treatment Group

A: Descriptive Statistics

	NIFDI		Other DI		Control		Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Asian	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01
African-American	0.75	0.34	0.93	0.13	0.85	0.27	0.84	0.27
Hispanic	0.05	0.12	0.002	0.003	0.01	0.01	0.01	0.04
Native American	0.01	0.02	0.001	0.001	0	0	0.003	0.01
Non-Hispanic White	0.17	0.25	0.06	0.12	0.14	0.25	0.14	0.25
Free and Reduced Lunch	0.83	0.06	0.72	0.21	0.74	0.15	0.75	0.15
Factor 1	-0.12	0.96	-0.07	1.06	0.02	1.01	0.01	1
Factor 2	1.18	2.82	-0.34	0.25	-0.11	0.47	0	1
N	11		5		103		119	

B: Analysis of Variance and Tukey Post-Hoc Tests

	Analysis of Variance		Post-Hoc Comparisons (significance)		
	F	p	NIFDI v. Other DI	NIFDI v. Control	Other DI v. Control
Asian	2.01	0.14	0.61	0.12	0.96
African-American	0.85	0.43	0.46	0.54	0.78
Hispanic	5.51	0.01	0.08	0.004	0.95
Native American	6.99	0.001	0.04	0.001	0.92
Non-Hispanic White	0.35	0.7	0.68	0.9	0.78
Free and Reduced Lunch	2.06	0.13	0.35	0.12	0.97
Factor 1	0.12	0.89	1	0.89	0.98
Factor 2	9.76	<.001	0.009	<.001	0.85

Table A-1 gives means and standard deviations on each of the demographic characteristics for each group of schools. Analysis of variance indicated no significant differences between the schools in their proportion of African American, non-Hispanic

[2] There were two cases where a predicted value fell outside of the theoretical range. Both of these involved the predictions for proportion white, where the predicted values were less than zero. For these cases (Malcolm X and Mildred Monroe schools), an average of the other years in the data file was used as the predictor. The average values for both cases were .01 or less.

White, or Asian students, as well as no difference in the proportion receiving free or reduced lunch. The NIFDI schools, however, had significantly more Hispanic and Native American students.

A factor analysis was conducted to develop summary measures of school context. Six variables were included in this analysis: the average proportion of Asian, Hispanic, African American, Native American, and white students as well as the average proportion of students receiving free and reduced lunch. A principle components extraction method was used, with varimax rotation. Two significant factors (with eigenvalues greater than 1.0) were found.[3] Factor scores were computed for each school and saved.

Results for the factor analysis are given in Table A-2. The commonalities indicate that, except for the measure of proportion of Asian students, there is a great deal of shared variance among the indicators, ranging from .66 for the proportion of students receiving free and reduced lunch to .94 for the proportion of African American students. Eigenvalues are also relatively high, with 73.5 percent of the total variance between schools explained by the two factors.

Table A-2: Factor Analysis of School Characteristics, Elementary Schools, BCPSS, 1998-2003

	Rotated Factor Loadings		Communalities
	Factor 1	Factor 2	
Proportion Asian	0.64	0.09	0.42
Proportion Black	-0.88	-0.40	0.94
Proportion Hispanic	0.13	0.85	0.74
Proportion Native American	0.08	0.87	0.77
Proportion White	0.90	0.27	0.89
Proportion Free/Reduced Lunch Eligible	-0.80	0.14	0.66
Eigenvalues	2.66	1.75	
% of total variance	44.4	29.1	

N = 120 schools

The first factor, which accounts for 44 percent of the total variance, has a strong positive loading of the proportion of white students, a slightly smaller positive loading for the proportion of Asian students, and strong negative loadings for the proportion of African American students and the proportion receiving free and reduced lunch. Thus, schools with positive scores on this factor would have proportionately more white and Asian students and many fewer Black and low income students. As can be seen in Table A-1, there are no significant differences between the three treatment groups on this factor.

The second factor, which accounts for 29 percent of the total variance, has positive loadings on the proportion of Hispanics and the proportion of Native Americans in a school, and a negative loading, of a somewhat smaller magnitude, of the proportion of African

[3] The factor analyses were also conducted with data for each year separately (i.e. the non-aggregated data), and the results were virtually identical to those obtained with the aggregated data. N = 120 in the factor analysis for Charles Carroll was included. The results did not differ when it was excluded.

Americans. Thus, schools with a positive score on this factor would have proportionately more Hispanic and Native American students and somewhat fewer African American students. The results in Table A-1 indicate that the NIFDI schools had significantly higher scores on this factor.

In addition, one of the schools in the Other DI group, Roland Park, was strikingly different from the other schools in its rate of children receiving free and reduced lunch. On average only .34 of the students in this school had free or reduced lunch, the third lowest rate in the BCPSS. The other schools in the Other DI group had free lunch proportions ranging from .76 to .84, while schools in the NIFDI group had proportions ranging from .74 to .90. While the presence of Roland Park in one of the treatment groups allows the examination of the effects of DI in more advantaged schools, it was also important to control for the presence of this school in the various analyses. In the analysis Roland Park is referred to as a Low Poverty Other DI school (LPODI) and the other schools within this group that received Direct Instruction but without NIFDI support is referred to as the High Poverty Other DI Schools (HPODI).

The Panel Sample

This section provides details on the sample used in the analysis of changes in achievement from first to fifth grade. The potential pool of students included all who were in the BCPSS in 1997-98 or 1998-99 and thus could still be in the system in fifth grade in 2001-02 or 2002-03, the last years for which data were available. Thus, data for this study included student cohorts who began first grade in the BCPSS in 1997-98 or 1998-99 and were in the same schools in fifth grade in either 2001-02 (for the first graders in 1997-98) or 2002-03 (for the first graders in 1998-99). Students who were held back were not included.

Table A-3 lists schools in each of the treatment groups (those that implemented Direct Instruction with NIFDI-support throughout the study period and those that implemented DI but did not retain support from NIFDI throughout the study period) and the date at which the DI Reading program was implemented in each school. [4] In the analysis reported below schools were designated as belonging to a given condition only in the years in which their school was receiving an implementation. For instance, Collington Square had no intervention in the first year of data collection (1997-98), but began implementing Direct Instruction reading programs in 98-99. Thus, students in Collington Square who were in first grade in 1998-99 were determined to be in the NIFDI reading intervention group, but those who were in first grade in 1997-98 were in the control group.

There was substantial student turnover during the study period, as summarized in Table A-4. There were almost 18,000 students in first grade in the BCPSS in 1997-98 and 1998-99. Five years later over half of these students (53 percent) had left the BCPSS, and an additional 20 percent had transferred to another school within the system. Only 27 percent of the students who were in first grade in the system in 1997-98 and 1998-99 were still in the same school in 5th grade in 2001-2 or 2002-3. The panel sample analyzed in this report included a total of 4,771 students. Of these, 355 were in NIFDI-supported Schools, 228 were

[4] There were two additional schools in the data set, but they, unfortunately, had the same name: Highland Town. The number assigned to the schools was not available for all years, and alternative spellings of the schools' name across years made it impossible to clearly differentiate them. Thus these two schools were eliminated from the analysis. Both were among the control group of schools.

in Other DI schools, and the remaining 4,188 were in the Control Schools.[5] Six schools in the BCPSS, all within the Control Group, had no students that persisted from first to six grade. As a result, the final panel sample included students from 113 schools.

Table A-3: Schools in Treatment Groups and Start Dates of Reading Direct Instruction

NIFDI Schools	
Arundel	96 – 97
CC Barrister	97 – 98
City Springs	96 – 97
Collington Square	98 – 99
Dickey Hill	98 – 99
Federal Hill	97 – 98
General Wolfe	96 – 97
Hampstead Hill	96 – 97
Langston Hughes	98 – 99
Margaret Brent	98 – 99
Dr. Rayner Browne	98 – 99
Other DI Schools	
William Pinderhughes	96 – 97
George Kelson	96 – 97
Robert Coleman	96 – 97
Roland Park	96 – 97
Westport	97 - 98

Table A-4: Turnover in the BCPSS, 1997-98, 1998-99 to 2001-02, 2002-03

	N	Percent
Left BCPSS (In first grade, but not in fifth)	9,362	52.8%
Transferred within BCPSS (In BCPSS in fifth grade, but in a different school)	3,595	20.3%
Panel Sample (Same School, first and fifth grade)	4,771	26.9%
Total Number of Students In First Grade 1997-98 or 1998-99	17,728	100.0%

[5] Students who were retained were omitted from the analysis. In addition, there was no way to control for the possibility that a student had attended multiple other schools between first and fifth grade. It is possible that some of the students in the panel sample were in the targeted schools in both first and fifth grade, but had attended other schools in the interim. Assuming that such children would have less of the "full treatment," this would bias results in a conservative direction.

Table A-5 gives the percentage of students who persisted within the same school, that is, who were in the same school for both first and fifth grade, for each of the groups. It can be seen that the persistence was much higher for students within the Low Poverty Other DI school than within the other schools: 57 percent compared to 26 to 28 percent for students in the other schools. When the LPODI school was removed from the analysis (the third column of data in Table A-5) there were no differences in the persistence of students in the groups of schools.

Table A-5: Percentage of Students Who Persisted in the Same Schools from First Grade to Fifth Grade by Group, BCPSS, 1998-2003

	% Persisting			N persisted	Total N
NIFDI	28	28	28	280	985
Other DI	32	-----	-----	228	718
HPODI	-----	26	26	156	591
LPODI	-----	57	-----	72	127
Control	27	27	27	4263	16025
chi-square	10.49	59.3	1.61		
Df	2	2	2		
P	0.005	<.001	0.445		

Note: The percentage of students who did not persist is omitted from the table. This can be simply calculated by subtracting the percentage given from 100.

As would be expected the students in the panel group (those that were in the same schools in 1st and 5th grade) differed from other students in both achievement and the nature of the schools that they attended. In addition, however, these differences varied from one treatment group to another. Table A-6 compares the first grade reading achievement scores of students in the panel group and the other students (combining those who moved elsewhere in the system and those for whom data for only first grade were available). Analyses of variance were calculated with treatment group (NIFDI, Other DI, and Control schools) and data availability (presence or absence of panel data) as factors. In all cases, the first grade achievement scores were higher for those in the panel sample than for those who moved. However, the differences between the panel sample and those who moved were significantly smaller for students in the NIFDI supported schools than for students in the other schools. (See the significant interaction term for all three achievement measures.) For instance, translating the norm equivalent scores to percentiles, the average student in a NIFDI supported school who was in the panel sample had a composite reading score at the 32nd percentile in first grade, compared to a score at the 42rd percentile for an average student in the panel who attended an Other DI school and a score at the 44th percentile for an average student in the panel who attended a control school. Differences are substantially smaller for the group that is not in the panel sample. (See Section B of Table A-6 for the percentiles that correspond to the norm equivalent scores.) Note that these differences would bias results in a

conservative direction, for the initial (first grade) differences between the treatment groups favor the non-NIFDI-supported groups.

Table A-6: Comparison of Panel Group to Other First Graders, First Grade Reading Achievement

A: Reading Achievement (Norm Equivalent Scores)

	Vocabulary		Comprehension		Composite	
NIFDI Supported	Mean	s.d.	Mean	s.d.	Mean	s.d.
Non-Panel Sample	34	21.5	33.1	19.3	33.3	19.3
Panel Sample	40	21.4	40.4	18.4	40	18.3
Other DI						
Non-Panel Sample	34.7	20.5	31.1	19.1	32.8	18.7
Panel Sample	48	21.9	45	20.9	46.3	20.4
Control Schools						
Non-Panel Sample	37.4	32.2	33.5	20.6	35.2	21.1
Panel Sample	50.1	21.6	44.8	20.2	47.1	19.9
ANOVA Results	F	P	F	p	F	P
Panel vs. Non-Panel	172.9	<.001	219.25	<.001	205.3	<.001
Treatment Group	35.5	<.001	5.98	0.003	19	<.001
Interaction	8.1	<.001	4.99	0.007	6.7	0.001

B: Corresponding Percentile Scores

	Vocabulary	Comprehension	Composite
NIFDI Supported			
Non-Panel Sample	22	21	21
Panel Sample	32	32	32
Other DI			
Non-Panel Sample	24	18	21
Panel Sample	46	41	42
Control Schools			
Non-Panel Sample	27	22	24
Panel Sample	50	41	44

Table A-7 compares the measures of school context between the panel and non-panel groups of students and the groups of schools. As explained more fully in Appendix I, higher scores on Factor 1 indicate a school with more non-Hispanic white students and fewer students who receive free and reduced lunch. A higher score on Factor 2 indicates a school with more Hispanic and Native American students. Note that data are included for the Other DI schools as a group and differentiating the HPODI and LPODI schools. The results of three different analyses of variance are included for both dependent variables: one comparing the three groups (NIFDI-supported, Other DI Schools, and the Control Schools), one comparing the four groups (NIFDI-supported, HPODI Schools, the LPODI School, and the Control Schools), and one comparing three groups (omitting the LPODI School).

The F-test associated with the interaction term is significant in all comparisons, indicating that differences between those in the panel sample and not in the panel sample varies from one group to another. Students in the panel sample were more likely to attend schools with higher scores on Factor 1, indicating a school with more non-Hispanic white children and fewer low income and African American children. Yet, when the data for the students in the Other DI group were disaggregated by the poverty status of the schools it can be seen that the differences in this group are highly influenced by the heterogeneity of the category. The results with the analysis of Factor 2 indicate that the tendency for the students in the NIFDI-supported schools to be in schools with more Hispanic and Native American students is strengthened within the panel sample. These results indicate, again, the importance of controlling for school characteristics in the analysis.

Table A-7: School Characteristics by Presence in Panel Sample and Group

	Factor 1(White and Higher Income)		Factor 2 (Hispanic and Native American)		
NIFDI Supported	Mean	s.d.	Mean	s.d.	N
Non-Panel Sample	-0.06	0.91	1.34	2.71	705
Panel Sample	0.17	1.07	1.76	2.89	280
Total	0.01	0.96	1.46	2.77	985
Other DI					
Non-Panel Sample	-0.28	0.75	-0.29	0.18	490
Panel Sample	0.21	1.10	-0.41	0.25	228
Total	-0.12	0.91	-0.33	0.21	718
HPODI					
Non-Panel Sample	-0.55	0.08	-0.23	0.07	435
Panel Sample	-0.53	0.09	-0.24	0.07	156
Total	-0.54	0.08	-0.23	0.07	591
LPODI					
Non-Panel Sample	1.82	0.00	-0.77	0.00	55
Panel Sample	1.82	0.00	-0.77	0.00	72
Total	1.82	0.00	-0.77	0.00	127
Control Schools					
Non-Panel Sample	-0.13	0.85	-0.10	0.48	11762
Panel Sample	0.14	1.09	-0.15	0.46	4263
Total	-0.06	0.93	-0.12	0.47	16025
ANOVA Results	F	p	F	p	df
Comparison 1 (NIFDI, Other DI, and Control)					
Panel vs. Non-Panel	100.44	<.001	8.40	0.004	1, 17722
Treatment Group	1.80	0.16	1743.22	<.001	2, 17722
Interaction	7.52	0.01	33.30	<.001	2, 17722
Comparison 2 (4 groups)					
Panel vs. Non-Panel	7.21	0.01	4.37	0.04	1, 17720
Treatment Group	225.11	<.001	1177.85	<.001	3, 17720
Interaction	3.98	0.01	21.54	<.001	3, 17720

Comparison 3 (3 groups, omitting LPODI)					
Panel vs. Non-Panel	23.55	<.001	14.25	<.001	1, 17595
Treatment Group	82.53	<.001	1701.09	<.001	2, 17595
Interaction	4.63	0.01	32.07	<.001	2, 17595

Appendix II: Detailed Statistical Results

This appendix expands upon the material presented in the text, including a more extended discussion of the analysis techniques employed and the results related to each of the three research questions.

Analysis

The analysis of data focused on changes in students' achievement from first to fifth grade and the extent to which students in schools with Direct Instruction had greater gains over time. Three research questions were addressed. The first was, "Do students who received Direct Instruction in first grade have greater gains in reading achievement from first to fifth grade than students in other schools?" Several comparisons were used to examine this question. First, the average change from first grade to fifth grade was computed for each of the three measured areas and in both groups (DI and Control Schools). Simple paired t-tests were used to examine changes from first to fifth grade. If students in DI schools had greater gains than those in the Control schools, the t-values associated with their average change would be expected to be larger. Second, to provide more insight into the results, independent sample t-tests were conducted to compare the average achievement scores of students in the DI and Control schools in both first and fifth grade. It was expected that results would favor the students in the DI schools much more strongly in fifth grade than in first grade. Third, a repeated measures analysis of variance was conducted with the first and fifth grade scores as repeated measures and the school as a factor. If gains were greater in the DI schools, the interaction between the repeated measures would be expected to be significant. Finally, the first research question was addressed by employing a mixed model regression, having schools as a random variable and regressing students' fifth grade scores on first grade scores, the measures of a school context of advantage/ disadvantage, and school type (DI or Control). This analysis is important to control for possible influences of school context and initial differences in achievement on the results. Successively more complex models were used and differences between these models were examined. It was expected that students in the DI schools would continue to have an advantage in fifth grade achievement when these controls were introduced.

Mixed models are particularly appropriate for analyzing multi-level data, such as data regarding students and the schools that they attend. In these models a "random variable" is used to control for differences between schools (often termed the level 2 entity) while calculating regression coefficients regarding the impact of variables from both students and schools on achievement. The random variable is equivalent to having a separate intercept in the regression equation for each school. The coefficients associated with the various individual and school related variables are then calculated while this between school variance is controlled. The analysis also allows one to calculate the amount of variance in the dependent variable that occurs between schools and the extent to which various independent variables can account for this between school variance.

The second research question was, "Do students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?" It should be recalled that all of the students who received DI instruction were in schools supported by NIFDI in first grade. The Other DI schools, however, ceased this involvement soon after the first grade year of students in the panel sample. To analyze this question three treatment groups were used in the

comparisons: NIFDI supported schools, Other DI schools and the Control schools. In addition, within the Other DI group results for the higher income school were differentiated from those for the Other DI schools. The various techniques used to analyze the first research question were also used with this question, except that, because of the number of comparisons involved, one-way analysis of variance was used instead of independent t-tests.

The third research question asked, "What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?" To answer this question, the results were translated into effect sizes. Unlike tests of significance, effect sizes are unaffected by the size of a sample. Because the issue of interest in this report is the change in scores from one year to another the standard method of calculating effect sizes (dividing the difference between two means by the common standard deviation) is not appropriate. Instead, the correction suggested by Dunlap and associates (1996) is used. The calculations involve an adjustment for the correlation between the two scores. Results may be interpreted in the same manner as Cohen's d. Effect sizes greater than .25 have traditionally been considered educationally significant (Fashola and Slavin 1990). However, McLean and Associates (2000) caution that effect sizes calculated with norm equivalent scores, as with this analysis, are inherently smaller than with other metrics. Thus the effect sizes presented are a conservative estimate of the results.

Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Reading Achievement from First to Fifth Grade than Students in Other Schools?

Four different approaches were used to address this question: paired t-tests, independent sample t-tests, repeated measures analysis of variance, and mixed-model regressions.

Difference Scores and Paired t-tests

Table A-8 summarizes the data used to examine differences in achievement gains from first to fifth grade for students in the Direct Instruction (whether or not they were in NIFDI-supported schools) and Control Schools. Panel A gives descriptive statistics for each measure of reading achievement using norm equivalent scores. Panel B gives the average change from first to fifth grade, the correlations of first and fifth grade scores, and results of paired t-tests that examine the null hypothesis that there was no change, on average, in achievement from one grade to the other. Panel C translates the averages in Panel A into percentile scores. The data in Panel C can be interpreted as the percentile score that the average student obtained in the two groups of schools in first grade and fifth grade. The data in Panel C were used to construct Figure 1 in the text.

Results indicate that the changes in achievement were much larger for students in the DI schools than in the Control Schools. Students in the DI schools had norm equivalent achievement scores in fifth grade that were significantly higher than their scores in first grade on all three dimensions. Students in the Control schools had comprehension and composite scores that were significantly higher in fifth grade, but vocabulary scores that were significantly lower. Examination of the average difference (Panel B) shows that the changes were much larger for students in the DI schools than in the Control Schools.

Table A-8: Descriptive Statistics and Paired t-test Results, 1st and 5th Grade Reading Achievement, BCPSS, 1998-2003

Panel A: Descriptive Statistics (Norm Equivalent Scores)

		DI Schools		Control Schools	
		Mean	S.D.	Mean	S.D.
Vocabulary	1st	43.7	22	50.1	21.6
	5th	51.5	19.9	49.4	17.5
	% gain	18		-1.5	
Comprehension	1st	42.5	19.7	44.9	20.2
	5th	51.4	18.1	49.3	16.8
	% gain	21.1		9.9	
Composite	1st	42.8	19.5	47.1	19.9
	5th	51.9	19.1	49.7	16.9
	% gain	21.4		5.4	

B: Average Differences, Correlations, paired (dependent sample) t-scores

	Ave. Dif.	R	T	p	N
DI Schools					
Vocabulary	7.8	0.54	8.70	<.001	497
Comprehension	9	0.53	10.89	<.001	501
Composite	9.1	0.62	12.03	<.001	495
Control Schools					
Vocabulary	-0.7	0.47	-2.31	0.021	4172
Comprehension	4.5	0.52	15.67	<.001	4207
Composite	2.6	0.56	9.48	<.001	4136

C: Percentile Score of Average Student by School Type, First and Fifth Grade

		DI Schools	Control Schools
Vocabulary	1st	38	50
	5th	53	49
Comprehension	1st	36	40
	5th	53	49
Composite	1st	37	44
	5th	54	49

The numbers in Panel C illustrate the magnitude of these changes. An average student in a DI school scored from the 36th to 38th percentile on all three measures in first grade, but by 5th grade, the average student scored above the national norms, at the 53rd or 54th percentile. In contrast, although doing better on two of the three measures in 5th grade than in 1st grade, the average student in the Control school still had scores just below the national norm (at the 49th percentile) on all measures. The lines in panel A labeled percent gain also illustrate the magnitude of the changes over time. These were calculated from the norm

equivalent scores and are simply the percentage change from first to fifth grade. It can be seen that the values are much larger, on average, for students in the DI schools than students in the Control schools. These data were used to create Figure 2 in the text.

Independent t-tests

Table A-9 provides an alternative way to examine the data by looking at the difference in average achievement scores of students in the DI schools and Control schools in first grade and in fifth grade. In first grade students in the Control schools had norm equivalent scores that were significantly higher than those of students in the DI schools. By fifth grade, however, this had reversed, with students in the Control schools having scores that were significantly lower than those of students in the DI schools. (See the mean scores in Table A-8.)

Table A-9: Independent t-tests, Average Reading Achievement, DI and Control Schools, 1st and 5th Grade, BCPSS, 1998-2003

First Grade	t-value	df	P
Vocabulary	6.34	4711	<.001
Comprehension	2.51	4713	0.01
Composite	4.62	4676	<.001
Fifth Grade			
Vocabulary	-2.3	600	0.02
Comprehension	-2.45	614	0.02
Composite	-2.57	602	0.01

Note: F tests indicated that the t-value assuming unequal variances between the groups should be used for all comparisons with fifth grade scores. The degrees of freedom reported are adjusted for that result. Negative t values indicate that students in the DI schools had higher average scores than students in the Control schools. Positive values indicate that the Control school students had higher scores.

Repeated Measures ANOVA

Table A-10 gives the results of a repeated measures analysis of variance with the first and fifth grade scores as a repeated measure and the type of school (DI or Control) as a factor. This approach is more parsimonious than the separate t-tests reported in Tables A-8 and A-9. If exposure to Direct Instruction in the first grade is related to higher achievement in fifth grade, a significant interaction effect would be expected. The results are as expected. The interaction effect is significant for each of the measures of reading achievement, indicating that the change from grade 1 to grade 5 is significantly different for students in the DI schools than in the Control schools. As noted above, the changes were greater for students in the DI schools than in the Control schools.

Table A-10: Repeated Measures Analysis of Variance, Norm Equivalent Reading Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools

	Vocabulary	Comprehension	Composite
F – Reading	54.24	237.71	202.02
P	<.001	<.001	<.001
F- Group	7.1	0.038	1.79
P	0.008	0.84	0.18
Interaction	78.76	26.91	63.86
P	<.001	<.001	<.001
df	1, 4668	1, 4707	1, 4630

Note: The F test associated with reading tests the null hypothesis that the change from first grade to fifth grade equals zero. The F test associated with group tests the null hypothesis that the average reading score of students in the DI schools and that of students in the Control schools are equal. The F associated with the Interaction tests the hypothesis that the differences between groups of schools are the same for first and fifth grade or, alternatively, that the differences between first and fifth grade are the same for the two groups of schools.

Mixed Model Regressions

Tables A-11 and A-12 report the results of our final method of testing differences in achievement gains between the two groups of schools: the mixed model regressions. Panel A of Table A-11 summarizes the models that were used and Panel B gives summary statistics associated with each model. The models become increasingly more complex. Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average reading achievement in fifth grade. The correlation ratio attached to Model 1 is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that, depending on the dependent measure, between 7 and 9 percent of the variance in fifth grade achievement is between schools rather than simply between students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model and school differences are controlled. These null hypotheses can be easily rejected with all three models. There is significant variation between schools and also between students even when the explanatory variables are included. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those in this analysis.

The models become incrementally more complex, with each subsequent model adding more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the norm equivalent reading achievement score for year one and the two factor scores measuring the advantage/disadvantage of the schools the students attended. It was expected that students' first grade achievement would be positively and significantly associated with their achievement in fifth grade. Including the two factor

scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African American students and more non-Hispanic white students,

Table A-11, Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3
Random intercept for Schools	x	x	x
First Grade Achievement		x	x
Factors 1 and 2		x	x
DI in First Grade			x

Panel B: Model Fit Statistics

		Model 1			Model 2			Model 3	
Vocabulary	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects Est.	28.6	4.9	<.0001	23.76	4.2	<.0001	21.4	3.8	<.0001
Residual Estimate	286.6	6	<.0001	218.8	4.6	<.0001	218.7	4.6	<.0001
BIC	40347			38621			38615		
- 2 Log Likelihood	40333			38593			38583		
Change in LL				1740			10		
Df				3			1		
P				<.001			<.01		
Correlation ratio	0.09								
Comprehension									
Random Effects Est.	19.9	3.6	<.0001	17.0	3.0	<.0001	16.0	2.9	<.0001
Residual Estimate	266.4	5.5	<.0001	193.3	4.0	<.0001	193.4	4.0	<.0001
BIC	40299			38343			38345		
- 2 Log Likelihood	40285			38315			38312		
Change in LL				1970			2.5		
Df				3			1		
P				<.001			>.10		
Correlation ratio	0.07								
Composite									
Random Effects Est.	25.6	4.4	<.0001	22.2	3.8	<.0001	20.1	3.5	<.0001
Residual Estimate	269.1	5.6	<.0001	177.9	3.7	<.0001	178.0	3.7	<.0001
BIC	40012			37363			37360		
- 2 Log Likelihood	39998			37335			37327		
Change in LL				2663			7.5		
Df				3			1		
P				<.001			<.01		
Correlation ratio	0.09								

would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient).

Model 3 adds the treatment group, testing the hypothesis that students in DI schools have higher achievement than students in the other schools when first grade achievement and school advantage/disadvantage are controlled. If having Direct Instruction in first grade influences students' fifth grade achievement, independent of their first grade achievement and school context, we would expect positive associations with this variable. It should be recalled that students were only coded as having Direct Instruction if they were in a school in a year when the school was in a treatment condition.

Two other hypotheses were tested. The first was that the impact of first grade achievement on fifth grade achievement would vary significantly between the two groups of schools. This was tested by adding the interaction of school group (DI versus control schools) with first grade scores to the variables in Model 3. The second hypothesis was that the impact of school advantage/disadvantage would differ between the two groups. This was tested by adding the interaction of school group with factor 1 to Model 3. Neither of these interaction effects was significant either when added singly or when added jointly with any of the measures of reading achievement. Thus, those results are not included in Tables A-11 or A-12.

The -2 log likelihood measures and the BIC values in Panel B of Table A-11 can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the comparisons between these values are also included in Table A-8. For example, Model 2 provides a significantly better fit to the data than Model 1 for all measures. For comprehension, the change in the -2 Log Likelihood = 1970 (40,285-38,315). With three degrees of freedom (because three new variables were added to Model 2 compared to Model 1), this result is highly significant. The comparison of Model 2 with Model 1 indicates that adding the variables in the more complex model significantly improves the fit with all three measures. (See the significance associated with the change in the -2 log likelihood.) However the significance of the change from Model 2 to Model 3 does not reach statistical significance for the analysis of reading comprehension. The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. The lowest BIC values appear for Model 3 for the analyses of vocabulary and the composite scores, but for Model 2 for the analysis of comprehension scores.

Table A-12 gives the coefficients associated with variables in Models 2 and Model 3 for each of the measures of achievement. The coefficients associated with first grade achievement are positive and highly significant in all models for all three dependent variables, indicating that, as expected, students with higher achievement scores in first grade also have higher achievement scores in fifth grade. The coefficients associated with the two factor scores are also as expected, although only those associated with Factor 1 are statistically significant. Students in schools with higher scores on Factor 1 (indicating higher proportions of white students and fewer low income and African American students) had significantly higher achievement scores. Students in schools with higher scores on Factor 2 (indicating more Hispanic and Native American students) had lower achievement scores, although this did not reach traditional levels of significance.

Table A-12: Coefficients Associated with Mixed Model Results, Fifth Grade Reading Regressed on First Grade Reading, School Advantage/Disadvantage, and Attending a DI School

	Model 2			Model 3		
Vocabulary	b	s.e.	p.	b	s.e.	p.
Intercept	29.6	0.7	<.0001	29	0.7	<.0001
First grade score	0.4	0.01	<.0001	0.4	0.01	<.0001
Factor 1	1.3	0.5	0.0081	1.3	0.5	0.005
Factor 2	-0.5	0.5	0.3442	-1	0.5	0.061
Had DI in First Grade	------	------	------	4.9	1.5	0.0009
Comprehension						
Intercept	29.2	0.7	<.0001	29	0.7	<.0001
First grade score	0.4	0.01	<.0001	0.4	0.01	<.0001
Factor 1	1	0.4	0.02	1.1	0.4	0.01
Factor 2	-0.1	0.4	0.83	-0.3	0.5	0.5
Had DI in First Grade	------	------	------	2.1	1.3	0.1
Composite						
Intercept	25.9	0.7	<.0001	25.4	0.7	<.0001
First grade score	0.5	0.01	<.0001	0.5	0.01	<.0001
Factor 1	1.2	0.5	0.01	1.2	0.5	0.01
Factor 2	-0.3	0.5	0.56	-0.7	0.5	0.18
Had DI in First Grade	------	------	------	3.8	1.4	0.01

As expected, the coefficient associated with attending a DI school in first grade is positive in all cases. The results with both the measure of vocabulary and the composite score are statistically significant. Independent of students' first grade reading score and characteristics of their school context, students who had Direct Instruction in first grade had higher vocabulary and composite achievement scores in fifth grade. However, in the analysis of comprehension scores, the result does not reach traditional levels of statistical significance ($p = .10$).

Table A-13 summarizes the impact of these effects assuming that students had the same first grade reading score and that all schools were equivalent in their socio-economic characteristics. Average scores for the total group were substituted into the regression equations for Model 3 in Table A-9 for these calculations. Results are given both in the norm equivalent scores and as the corresponding percentile score for a student scoring at this average. It is clear that, after equalizing for first grade scores and school context, BCPSS students who had Direct Instruction in first grade would be expected to score at or above the national average in fifth grade, while students in the control schools would have scores that were still substantially below the national norm.

Table A-13: Simulated Fifth Grade Achievement Scores (Norm Equivalent and Percentile of Hypothetical Average Student), DI and Control Schools, Controlling for School Context and First Grade Achievement, BCPSS, 2001-03

Norm Equivalent Scores	Direct Instruction	Control Schools
Vocabulary	51	46
Comprehension	50	48
Composite	51	47
Corresponding Percentiles		
Vocabulary	51	42
Comprehension	50	46
Composite	51	44

Note: Scores were calculated by substituting the average values of first grade achievement and factor scores in the equation for Model 3 in Table A-12. For Vocabulary the value was 43.6, for comprehension the value was 42.5, for the composite score the value was 42.8. The value for Factor 1 was .19 and the value for Factor 2 was .79. The values for the Factors were not equal to zero both because the panel sample had a slightly different sample of schools than the group for which the factor scores were calculated and, more importantly, because the unit of analysis for calculating the means was students rather than schools.

Summary

To summarize, four different methods were used to examine the first research question, which asked if students who were in DI schools in first grade had higher achievement scores in fifth grade than students in the control schools. First, paired (dependent sample) t-tests were used to compare students' scores in first and fifth grade. As expected, changes from first to fifth grade were much larger for students in the DI schools than in the Control Schools. Second, independent t-tests compared scores of students in the DI schools and students in the Control schools in both first and fifth grade. While students in the Control Schools had significantly higher scores in first grade, students in the DI schools had significantly higher scores in fifth grade. Third, a repeated measures analysis of variance, with achievement as a repeated measure and group as factor, resulted in significant interaction effects, again reflecting the larger gains from first to fifth grade for the students in the DI schools. Finally, mixed model regressions predicted fifth grade scores, controlling for students' first grade scores and characteristics of the schools. As expected, the coefficient associated with attending a DI school in first grade was positive in all cases and statistically significant for the measures of vocabulary and the composite measure.

Research Question Two: Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?

Four analysis techniques were used to address the second research question: paired t-tests within each group, one-way analysis of variance, repeated measures analysis of variance, and mixed model regressions. Note that this analysis differs from that with Research Question 1 by differentiating students in NIFDI-supported schools from students in the Other DI schools.

Descriptive Statistics and Paired t-tests

Table A-14 summarizes the data used to examine differences in achievement gains from first to fifth grade for students in the NIFDI supported schools, the Other DI schools, and the Control Schools. Scores are also presented separately for students in the low poverty Other DI (LPODI) school. This school had an extraordinarily higher rate of persistence from first to fifth grade and a much lower poverty rate than the other schools. It should also be recalled that all of the students studied in schools with NIFDI support during first grade, but only students in the NIFDI-supported schools had this support throughout their elementary career.

Panel A of Table A-14 gives descriptive statistics for each measure of reading achievement using norm equivalent scores. It also includes the percentage change from first to fifth grade. Panel B gives the average change from first to fifth grade, the correlations of first and fifth grade scores, and results of paired t-tests that examine the null hypothesis that there was no change, on average, in achievement from one grade to the other for each group. Panel C translates the average norm equivalent scores in Panel A into percentile scores. The data in Panel C can be interpreted as the percentile score that the average student obtained in each of the groups of schools in first grade and fifth grade. The data in Panel C were used to construct Figure 3 in the text, and the data on percentage change in Panel A were used to construct Figure 4.

Results indicate, as in Table A-8, that all students had higher norm equivalent scores in fifth grade than in first grade. These changes were significant for students in all groups. Yet, the percentage change in scores was greatest for students in the NIFDI-supported schools for all measures.[6] Values were lowest for students in the control schools for both the measures of absolute change (Panel B) and percentage change (Panel A).

The percentile scores for students scoring at the average (panel C) indicate that, by 5th grade, the average student in schools that received NIFDI support throughout their elementary careers scored at the national norm (49th or 50th percentile) on all three measures, substantially above the level that the average student had scored in 1st grade (the 32nd and 33rd percentile). The average student in the Control Schools also scored at around the national norm at 5th grade, but this represented very little change from first grade scores. The average score for the students in the Other DI group translated to percentiles above the mean (56 to 59), but the overall results for this group mask differences between the high poverty schools and the low poverty school. The average student in the low poverty DI school scored well

[6] The absolute change in scores was greater for students in the LPODI school for the measure of comprehension and the composite score, but the percentage change was actually lower because these students began at a higher level of achievement.

within the 8th decile, while the average student in the remaining Other DI schools scored well below the national norm.

Table A-14: Descriptive Statistics and Paired t-test Results, 1st and 5th Grade Reading Achievement Scores, NIFDI Supported Schools, Other DI Schools, and Control Schools, BCPSS, 1998-2003

A: Descriptive Statistics

		NIFDI		**Other DI**		**Control**		**HPODI**		**LPODI**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Vocabulary	1st	40	22	48	22	50	22	42	21	62	19
	5th	50	19	53	21	49	18	46	18	70	16
	% ch.	25		11		-2		9		14	
Comprehension	1st	40	17	45	21	45	20	39	19	57	20
	5th	49	18	54	19	49	17	48	17	69	15
	% ch.	22		21		10		21		20	
Composite	1st	40	18	46	21	47	20	40	18	59	19
	5th	50	18	55	20	50	17	47	17	71	16
	% ch.	25		18		5		16		20	

B: Average Differences, Correlations, paired (dependent sample) t-scores

	Ave. Dif.	r	t	p	N
NIFDI Schools					
Vocabulary	9.9	0.44	7.61	<.001	275
Comprehension	8.7	0.42	7.62	<.001	277
Composite	9.9	0.5	9.04	<.001	273
Other DI Schools					
Vocabulary	5.3	0.65	4.43	<.001	223
Comprehension	9.2	0.62	7.86	<.001	225
Composite	8.2	0.72	7.99	<.001	223
Control Schools					
Vocabulary	-0.7	0.47	-2.31	0.021	4172
Comprehension	4.5	0.52	15.67	<.001	4207
Composite	2.6	0.56	9.48	<.001	4136
HPODI					
Vocabulary	3.8	0.55	2.52	0.013	151
Comprehension	8.2	0.54	5.92	<.001	153
Composite	6.5	0.64	5.29	<.001	151
LPODI					
Vocabulary	8.4	0.56	4.43	<.001	72
Comprehension	11.4	0.49	5.22	<.001	72
Composite	11.8	0.62	6.51	<.001	72

C: Percentile Score of Average Student by School Type, First and Fifth Grade

		NIFDI	Other DI	Control	HPODI	LPODI
Vocabulary	1st	32	46	50	35	71
	5th	50	56	49	42	83
Comprehension	1st	33	41	40	30	63
	5th	49	58	49	45	81
Composite	1st	32	43	44	32	67
	5th	50	59	49	44	84

One-Way Analyses of Variance

Table A-15 summarizes the results of one-way analyses of variance with each of the measures of reading achievement. This analysis differs from that above by focusing on comparing scores between the groups of schools in both first grade and then in fifth grade. Two sets of comparisons are included. The first, in Panel A, involves a comparison between the three groups (NIFDI supported schools, Other DI schools, and Control Schools). The second, in Panel B, includes four categories, separating the students in the Other DI group into those in the low poverty school from those in other schools. Results of Tukey post-hoc tests for pair-wise comparisons are also included.

The analysis of variance results indicate that the groups of schools had significantly different average achievement scores in both first grade and in fifth grade. These results occur with both the comparison of three groups (in Panel A) and of four groups (in Panel B). However, the post-hoc tests show that these differences only occur between some pairs. In addition, paralleling the results reported above, the direction of the difference varies markedly from first grade to fifth grade.

In first grade, the students in the NIFDI-supported schools had significantly lower achievement scores than students in either the Other DI or the Control Schools on all three measures (see the first two columns in the post-hoc section of Panel A). The results in Panel B, however, show that this difference occurs only because of the presence of low poverty school in the Other DI group. First grade achievement scores of students in schools supported by NIFDI throughout the span of the study are virtually identical, on average, to those of students in the high poverty Other DI schools.

Results for Fifth Grade achievement indicate that there is no longer any difference between average scores of students in the NIFDI-supported schools and students in the control schools on any of the measures. Students in the NIFDI supported schools have significantly lower scores than students in the Other DI group, but this disappears when data for the LPODI school are analyzed separately. Students in the LPODI school have significantly higher scores than students in any of the other three groups, both in first grade and in fifth grade. Inspection of the means in Table A-14 indicates that students in the NIFDI-supported schools have higher scores than those in the HPODI schools on all three measures in fifth grade, but the results of the pair-wise post-hoc tests, given in Table A-15, indicate that these differences are significant only with the measure of vocabulary.

Table A-15: One-Way Analysis of Variance, Average Reading Achievement, NIFDI, Other DI, and Control Schools, 1st and 5th Grade, BCPSS, 1998-2003

A. Comparison of NIFDI, Other DI, and Control Schools

	Analysis of Variance Results			Tukey Post-Hoc Tests (prob.)		
First Grade	F	df	p	NIFDI v. Other DI	NIFDI v. Control	Other DI v. Control
Vocabulary	28.66	2, 4170	<.001	<.001	<.001	0.34
Comprehension	6.3	2, 4712	0.002	0.03	0.001	1.00
Composite	16.98	2, 4675	<.001	0.001	<.001	0.81
Fifth Grade						
Vocabulary	4.91	2, 4724	0.007	0.16	0.73	0.01
Comprehension	8.53	2, 4762	<.001	0.004	1.00	<.001
Composite	7.59	2, 4720	0.001	0.02	0.91	<.001

B. Comparison of NIFDI, HPODI, LPODI, and Control Schools

	Analysis of Variance Results			Tukey Post-Hoc Tests (prob.)					
First Grade	F	df	p	NIFDI v. HPODI	NIFDI v. Control	NIFDI v. LPODI	HPODI v. Control	HPODI v. LPODI	LPODI v. Control
Vocabulary	33	3, 4709	<.001	0.87	<.001	<.001	<.001	<.001	<.001
Comprehension	17.24	3, 4711	<.001	0.94	0.002	<.001	0.004	<.001	<.001
Composite	26.65	3, 4674	<.001	1.00	<.001	<.001	<.001	<.001	<.001
Fifth Grade									
Vocabulary	35.67	3, 4723	<.001	0.03	0.87	<.001	0.02	<.001	<.001
Comprehension	31.93	3, 4761	<.001	0.68	1.00	<.001	0.47	<.001	<.001
Composite	39.84	3, 4719	<.001	0.13	0.98	<.001	0.08	<.001	<.001

Repeated Measures Analysis of Variance

Table A-16 provides additional controls by using a repeated measures analysis of variance with the first and fifth grade scores as a repeated measure and the type of school (NIFDI, DI and Control for panel A and NIFDI, HPODI, LPODI, and control for panel B) as a factor. As noted in the analysis of the first research question, this approach is more parsimonious than the separate one-way analyses of variance or groups of paired t-tests. If exposure to Direct Instruction in first grade is related to higher achievement in fifth grade, a significant interaction effect would be expected. As expected, the interaction effects are significant in all cases. The changes in scores from first to fifth grade were not the same in the various groups. As demonstrated in Table A-14, these changes were substantially stronger in the NIFDI-supported schools than the other schools.

Table A-16: Repeated Measures Analysis of Variance, Normed Scores, Reading

A: Comparison of NIFDI, Other DI, and Control Schools

	Vocabulary	Comprehension	Composite
F - reading	60.45	178.7	169.64
Sig	<.001	<.001	<.001
F- Group	10.77	5.48	7.98
Sig	<.001	0.004	<.001
Interaction	42.59	13.5	32.54
Sig.	<.001	<.001	<.001

B: Comparison of NIFDI, HPODI, LPODI, and Control Schools

	Vocabulary	Comprehension	Composite
F - reading	45.37	130.65	129.21
Sig	<.001	<.001	<.001
F- Group	35.16	27.82	34.57
Sig	<.001	<.001	<.001
Interaction	29.25	9.48	23.26
Sig.	<.001	<.001	<.001

Mixed Model Analyses

Tables A-17 and A-18 give the results of the mixed model analysis. Panel A in Table A-17 summarizes the models that were used and Panel B gives summary statistics associated with each model. Four models are tested. The first is the simple, random effects model. The second adds first grade achievement scores and Factor 1 as a predictor.[7] The third model adds dummy variables for being in a NIFDI-supported school throughout one's elementary career and for being in an Other DI school during the latter portions of the career. The fourth model is like Model 3, but separates the Other DI category into the LPODI and HPODI schools. For both Model 3 and Model 4 the omitted category of schools is the Control Schools.
Using the procedures described more fully in the discussion of results with the first research question, it can be seen that the model that best fits the data is Model 4, which includes first

[7] Factor 2 was omitted from this analysis because it had no significant effect in the analyses of Hypothesis One and to help conserve degrees of freedom. Results do not differ when it is included.

Table A-17: Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction, Comparing NIFDI Supported Schools, Other DI Schools (with and without low poverty school) and Control Schools

A: Variables in the Models

Variables	Model 1	Model 2	Model 3	Model 4
Rand. Int. - Schools	x	x	x	x
1st Gr. Achievement		x	x	x
Factor 1		x	x	x
NIFDI support			x	x
Other DI			x	
LPODI				x
HPODI				x

B: Model Fit Statistics

	Model 1			Model 2			Model 3			Model 4		
Vocabulary	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects	28.6	4.9	<.0001	24	4.2	<.0001	22.4	4	<.0001	20.7	3.8	<.0001
Residual	286.6	6	<.0001	218.8	4.6	<.0001	218.7	4.6	<.0001	218.7	4.6	<.0001
BIC	40347			38618			38619			38617.8		
- 2 Log Likelihood	40333			38594			38586			38580.5		
Change in LL				1739			8.1			14		
df				2			2			3		
p				<.0001			<.05			<.01		
Correlation ratio	0.09											
Comprehension												
Random Effects	19.9	3.6	<.0001	17	3	<.0001	16	2.9	<.0001	14.9	2.8	<.0001
Residual	266.4	5.5	<.0001	193.3	4	<.0001	193.3	4	<.0001	193.4	4	<.0001
BIC	40299			38338			38343			38343		
- 2 Log Likelihood	40285			38315			38311			38306		
Change in LL				1970			4			8.7		
df				2			2			3		
p				<.0001			>.05			<.05		

Correlation ratio	0.07											
Composite												
Random Effects	25.6	4.4	<.0001	22.2	3.8	<.0001	20.6	3.6	<.0001	19.3	3.4	<.0001
Residual	269.1	5.6	<.0001	177.9	3.7	<.0001	177.9	47.6	<.0001	178	3.7	<.0001
BIC	40012			37358			37361			37361		
- 2 Log Likelihood	39998			37335			37329			37324		
Change in LL				2663			6.4			11.2		
df				2			2			3		
p				<.0001			<.05			<.02		
Correlation ratio	0.09											

Note: The comparison of the log-likelihood values for Model 4 is with Model 2.

Table A-18: Mixed Model Regressions, Fifth Grade Reading (norm equivalent scores) on School Advantage/Disadvantage and School Type (NIFDI Supported, Other DI, and Control Schools)

	Model 2			Model 3			Model 4		
Vocabulary	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	29.6	0.7	<.0001	29.1	0.8	<.0001	29.2	0.7	<.0001
First grade score	0.4	0	<.0001	0.4	0	<.0001	0.4	0	<.0001
Factor 1	1.3	0.5	0.008	1.3	0.5	0.01	1.1	0.5	0.02
NIFDI	-----	-----	-----	4.3	1.7	0.01	4.2	1.7	0.01
Other DI	-----	-----	-----	3.8	2.4	0.12	-----	-----	-----
HPODI	-----	-----	-----	-----	-----	-----	0.8	2.6	0.77
LPODI	-----	-----	-----	-----	-----	-----	14.3	5	0.004
Comprehension									
Intercept	29.2	0.7	<.0001	29	0.7	<.0001	29	0.7	<.0001
First grade score	0.4	0	<.0001	0.4	0	<.0001	0.4	0	<.0001
Factor 1	1	0.4	0.02	1.1	0.4	0.01	0.9	0.4	0.04

NIFDI	-----	-----	-----	0.7	1.5	0.63	0.8	1.5	0.6
Other DI	-----	-----	-----	4.17	2.1	0.05	-----	-----	-----
HPODI	-----	-----	-----	-----	-----	-----	1.8	2.3	0.44
LPODI	-----	-----	-----	-----	-----	-----	12.4	4.3	0.004
Composite									
Intercept	25.9	0.7	<.0001	25.4	0.7	<.0001	25.5	0.7	<.0001
First grade score	0.5	0	<.0001	0.5	0.01	<.0001	0.5	0	<.0001
Factor 1	1.2	0.5	0.013	1.2	0.5	0.009	1	0.5	0.03
NIFDI	-----	-----	-----	2.8	1.6	0.07	2.9	1.5	0.06
Other DI	-----	-----	-----	4.3	2.3	0.05	-----	-----	-----
HPODI	-----	-----	-----	-----	-----	-----	1.7	2.5	0.5
LPODI	-----	-----	-----	-----	-----	-----	13.7	4.8	0

year reading achievement, factor 1 (the measure of school advantage/disadvantage), and the three dummy variables of treatment group (NIFDI supported, HPODI, and LPODI).

The coefficients in Table A-18 confirm the results found in conjunction with Hypothesis 1 regarding the impact of first grade reading scores and Factor 1. Students with higher reading achievement scores in first grade have significantly higher reading scores in fifth grade. Students in schools with more non-Hispanic white children and fewer children receiving free and reduced lunch also have higher achievement scores. These results appear with all three measures of achievement.

Results with model three, comparing the three groups of schools, indicate that students in schools supported by NIFDI throughout their elementary careers had higher achievement scores than other students, even after controlling for first grade scores and school context, but this difference is significant only for vocabulary scores ($p = .01$), although near significance ($p = .07$) for the composite score. Students in the Other DI schools also had higher scores than other students on all three measures, and these differences were significant ($p=.05$) for both comprehension and the composite score.

Results with Model 4, the best fitting model, which separates the low poverty and high poverty Other DI schools, show that the positive effect for the Other DI schools in Model 3 was due to the presence of the LPODI school. In Model 4, students in the LPODI school had significantly higher achievement scores than other students, even after controlling for their first grade scores and school context, on all three measures. Compared to the other coefficients associated with group membership, the magnitude of the coefficient associated with the LPODI school is quite large. Students in this school, once first grade scores and the socioeconomic characteristics of the school are controlled, are predicted to have fifth grade scores that range from 12 to 14 NCE points higher than students in the Control Schools. The advantage for students in the NIFDI-supported schools ranges from one to four points, while that for students in the Other DI schools ranges from one to two points. Even with such a low magnitude, however, the advantage for the NIFDI students in vocabulary scores remains statistically significant (.01), and the advantage in composite scores is close to traditional significance levels ($p =.06$).

Summary

As with research question number 1, the results indicate that students who had Direct Instruction in first grade had higher achievement scores in 5th grade than students in the Control Schools. However, the impact of receiving Direct Instruction was strongest for students in the low poverty school in the Other DI group. It should be recalled, however, that students in this school had significantly higher first grade achievement scores than students in other schools. Perhaps even more important, the school also had a much lower rate of student turnover throughout the years of the study.

Research Question Three. What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?

Because the measures of first and fifth grade achievement are statistically dependent, Cohen's d, the most often used measure of effect size, is not appropriate. To control for the dependence of these measures, effect sizes were calculated adjusting for the correlation between scores in the two years (Dunlap et al 1996). It was anticipated that effect sizes would be larger for the students who received Direct Instruction in the first grade than for other students. The effect sizes were calculated using the descriptive statistics from the norm equivalent scores. McLean and associates, 2000, have demonstrated that effect sizes for NCE scores are lower than those that would be obtained for equivalent raw and scaled scores. Thus, the calculations presented here should be seen as conservative estimates.

Table A-19 summarizes effect scores for the change in achievement from first grade to fifth grade on each of the measures. Using the criterion of .25 as an educationally significant effect size, it may be seen that all but one of the effect sizes associated with a group that received Direct Instruction in first grade meet this threshold. The one exception involves vocabulary scores for schools in the HPODI group, where the effect size is .20. The effect size for comprehension for the Control schools is almost at the .25 threshold. However, this effect is half or less than half that found for the other schools. The effect sizes associated with the NIFDI-supported schools all far surpass the .25 criterion and are second in size only to those for the LPODI school, which had very low levels of turnover.

Table A-19: Effect Sizes of Change, First to Fifth Grade, by Group, Reading Achievement Norm Equivalent Scores

	Vocabulary	Comprehension	Composite
Control Schools	-0.04	0.24	0.14
All DI Schools	0.37	0.47	0.47
NIFDI Supported	0.49	0.49	0.55
Other DI Schools	0.25	0.46	0.4
HPODI Schools	0.20	0.46	0.36
LPODI Schools	0.49	0.62	0.67

Note: All effect sizes are based on norm equivalent scores. The computations for the effect sizes for "all DI schools" and the "Control Schools" are based on the data in Table A-8 associated with Research Question 1. The computations for the other groups are based on the descriptive statistics given in Table A-14.

References

Dunlap, William P., Jose M. Cortina, Joel B. Vaslow, and Michael J. Burke. 1996. "Meta-analysis of experiments with matched groups or repeated measures designs." *Psychological Methods* 1: 170-177.

Fashola, Olatokunbo S. and Robert E. Slavin. 1997. Promising programs for elementary and middle schools: Evidence of effectiveness and replicability." *Journal of Education for Students Placed at Risk* 2: 251-307.

McLean, James E., Marcia R. O'Neal, and J. Jackson Barnette. 2000. Are All Effect Sizes Created Equal? Paper presented at the 28th Annual Meeting of the Mid-South Educational Research Association, November, 2000, ED 448 188.

From: jstockard@nifdi.org
Sent: Thursday, September 11, 2008 6:32 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number:
894997019

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: I wish to submit this for the beginning reading category.
For some reason the category button above was not working.

From: WhatWorks
Sent: Friday, September 12, 2008 7:30 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse
Hello,

Thank you for contacting the What Works Clearinghouse (WWC). We apologize that the document upload feature on our website did not work properly. We were unable to access the study you provided. Please reply to this email with the study attached.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Thursday, September 11, 2008 6:32 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID
Number: 894997019

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: jstockard@nifdi.org

Message: I wish to submit this for the beginning reading category. For some reason the category button above was not working.

From: WhatWorks
Sent: Friday, September 12, 2008 7:33 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse: correction

Dear Dr. Stockard,

We mistakenly stated that we could not access the study you submitted. We received the document & you do not need to submit anything further at this time.

Thank you for your submission.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: WhatWorks
Sent: Friday, September 12, 2008 7:30 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse

Hello,

Thank you for contacting the What Works Clearinghouse (WWC). We apologize that the document upload feature on our website did not work properly. We were unable to access the study you provided. Please reply to this email with the study attached.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

```
-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Thursday, September 11, 2008 6:32 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference
ID
Number: 894997019

info@whatworks.ed.gov, this email was automatically sent through the
Contact link on the WWC website.

From: jstockard@nifdi.org
```

Message: I wish to submit this for the beginning reading category.
For some reason the category button above was not working.

From: jstockard@nifdi.org
Sent: Monday, September 29, 2008 12:42 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number:
1718962892

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: The category is beginning reading - I couldn't get that box to work

Note that even though the dependent variable is reading achievement in 5th
grade, the intervention occurred in first grade. According to the WWC beginning reading protocol, this sample is relevant (subpoint 1 under "sample relevance", p. 3 of the WWC protocol updated 11/28/06).

From: jstockard@nifdi.org
Sent: Monday, September 29, 2008 12:53 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Check on the Status of a Submission, Reference ID
Number: 1795680584

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: I just tried to submit a study using a pdf format, but am unsure if
it loaded.
The number is 1718962892.
Just so you know, the browser would not do anything with the file when the
"add" button was pushed.
So then I just clicked on done.

Thanks for checking on this.
Jean Stockard

From: WhatWorks
Sent: Monday, September 29, 2008 10:37 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse study submission
Dear Dr. Stockard,

Thank you for your submission. We received your attached PDF of the study.

Please note that studies submitted to the WWC could be considered for review under the next round of reviews as determined by IES. We are unable to provide a timeline for the review process at this time.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Monday, September 29, 2008 12:53 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Check on the Status of a Submission, Reference ID Number: 1795680584

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: jstockard@nifdi.org

Message: I just tried to submit a study using a pdf format, but am unsure if it loaded.
The number is 1718962892.
Just so you know, the browser would not do anything with the file when the “add” button was pushed. So then I just clicked on done.

Thanks for checking on this.
Jean Stockard

-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Monday, September 29, 2008 12:42 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number: 1718962892

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: jstockard@nifdi.org

Message: The category is beginning reading - I couldn't get that box to work

Note that even though the dependent variable is reading achievement in 5th grade, the intervention occurred in first grade. According to the WWC beginning reading protocol, this

sample is relevant (subpoint 1 under “sample relevance”, p. 3 of the WWC protocol updated 11/28/06).

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

September 2008

Technical Report 2008-03



© 2008 National Institute for Direct Instruction
PO Box 11248 • Eugene, OR 97440
1-877-485-1973

Table of Contents

	Page
Executive Summary	iv
List of Tables	v
List of Figures	vii
Report	1
A: First Grade Mathematics Achievement	1
B. The Relationship of First Grade Direct Instruction and Fifth Grade Mathematics Achievement	8
C. Summary and Discussion	13
Appendix I: Sample and Measures – First Grade Analysis Sample Measuring Achievement Measuring School Context	15
Appendix II: Detailed Statistical Results – First Grade Analysis Methodology Research Question One: Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools? Research Question Two: Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized? Research Question Three: Did the impact of being in a DI School increase over time as reforms became institutionalized within a school? Research Question Four: What is the Magnitude of the Effect of Direct Instruction?	20
Appendix III: Panel Sample: Changes from First to Fifth Grade	33
Appendix IV: Detailed Statistical Results: Fifth Grade Achievement Analysis Methodology Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools? Research Question Two: Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the	37

socio-demographic context of the school and first grade achievement are controlled?
Research Question Three. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

References 50

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Executive Summary

In the mid 1990s, in response to very low achievement scores, the Baltimore City Public School System (BCPSS) implemented curricular reforms. Sixteen schools used Direct Instruction. They began by implementing the reading curriculum and, one to two years later, added mathematics Direct Instruction, using *Distar Arithmetic* and *Connecting Math Concepts*. While all of these schools initially received technical support from the National Institute for Direct Instruction, five eventually ceased this support and used an alternative provider. The Comprehensive Test of Basic Skills (CTBS) was administered to all first graders and all fifth graders in BCPSS in the spring of 6 school years: 1997-1998 through 2002-2003. Data were obtained from over 40,000 first grade students on mathematics achievement in computations, concepts and applications, and a composite score. Of the students who were in first grade in 1997-98 and 1998-99, more than 4,500 remained in the system and in the same schools in 2001-02 and 2003-03.

This study compares students' mathematics achievement in 1) schools with NIFDI support, 2) schools that implemented Direct Instruction without NIFDI support (other DI schools), and 3) schools without Direct Instruction (the Control schools). Schools in the three groups had similar demographic characteristics, although one of the Other DI schools had a markedly lower rate of poverty and its data are examined separately. Key findings are listed below.

- Combining data across all years (1998-2003) first grade students in schools that implemented Direct Instruction had significantly higher mathematics achievement than students in other schools.
- Combining data across all years (1998-2003) students in NIFDI-supported schools had significantly higher achievement than students in the demographically similar high poverty direct instruction school and the control schools.
- When the socio-economic characteristics of the schools were controlled, the students in the NIFDI-supported schools had significantly higher achievement scores than students in both the Other DI schools and the control schools on all measures of achievement.
- First grade students in all 3 groups of schools had higher achievement scores in 2003 than in 1998, but the increase was significantly larger for students in NIFDI-supported schools than for students in the other schools. On average, first grade composite mathematics achievement scores in the NIFDI-supported schools increased by 135 percent from 1998 to 2003.
- The magnitude of the effect on first grade achievement from attending a NIFDI-supported school was statistically significant and substantively large.
- Among the students who remained in the same school from first grade to fifth grade, those in the NIFDI-supported schools had significantly lower achievement in first grade than those in the other schools. By fifth grade, however, these differences had disappeared.

List of Tables

Table Title	Page
A-1: Schools in Treatment Group and Start Dates of Reading and Mathematics Direct Instruction	16
A-2: Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels by Group	17
A-3 Factor Analysis of School Characteristics, Elementary Schools, BCPSS, 1998-2003	18
A-4: Average Mathematics Achievement Scores of Students With and Without Direct Instruction, All Years, First Graders, BCPSS	21
A-5: Variables in Models and Measures of Fit, Analysis of First Grade Mathematics Achievement, BCPSS, 1998-2003	23
A-6: Coefficients Associated with Mixed Model Regressions of First Grade Mathematics Achievement, BCPSS, 1998-2003	26
A-7: Predicted Achievement Scores Assuming Equal School Contexts	27
A-8: First Grade Mathematics Achievement Scores by Group and Year, Adjusting for School Advantage, BCPSS	28
A-9: Mathematics Achievement Scores by Year and Group, First Graders, BCPSS, 1998-2003	30
A-10: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Mathematics Achievement, BCPSS, 1998-2003	32
A-11: Turnover in the BCPSS, 1997-98, 1998-99 to 2001-02, 2002-03	33
A-12: Schools in Treatment Group by Year for Panel Analysis	34
A-13: Persistence in the Same School, by School Group	35
A-14: Comparison of Achievement Scores of Panel Group and Other First Graders, DI and Control Schools	35
A-15: Descriptive statistics and Paired t-test Results, 1st and 5th Grade Mathematics Achievement Norm Equivalent Scores, DI and Control Schools, Full and Reduced Samples, BCPSS, 1998-2003	38

A-16: Independent t-tests, Average Mathematics Achievement, DI and Control Schools (Full and Reduced Samples), 1st and 5th Grade, BCPSS, 1998-2003 40

A-17: Repeated Measures Analysis of Variance, Norm Equivalent Mathematics Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools 41

A-18, Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Mathematics Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction 42

A-19: Coefficients Associated with Mixed Model Results, Fifth Grade Mathematics Achievement Regressed on First Grade Achievement, School Advantage/Disadvantage, and Attending a DI School 45

A-20: Simulated Fifth Grade Mathematics Achievement Scores (Norm Equivalent and Percentile of Hypothetical Average Student), DI and Control Schools, Controlling for School Context and First Grade Achievement, BCPSS, 2001-03 47

A-21: Effect Sizes of Change, First to Fifth Grade, by Group, Mathematics Achievement Norm Equivalent Scores 49

List of Figures

Figure Title	Page
1: Percentile Scores, Mathematics, Average Student, DI and Control Schools	3
2: Percentile Scores, Average Student by Group, Adjusting for School SES	4
3: Computation Achievement, Average Student, BCPSS, 1998-2003 by Group	5
4: Concepts and Applications Achievement, Average Student, BCPSS, 1998-2003 by Group	5
5: Composite Achievement, Average Student, BCPSS, 1998-2003 by Group	5
6: Percentage Increase in Mathematics Achievement, 1998-2003, by Group	6
7: Effect Sizes, Mathematics Achievement, All Years and 2003	7
8: Effect Sizes, Mathematics Achievement, Adjusting for School Context, All Years and 2003	8
9: Computation, Concepts and Applications, and Composite Mathematics Percentile Achievement Scores of an Average Student, First and Fifth Grade, DI and Control Schools, BCPSS, 1998-2003	10
10: Percentage Gain in Achievement Scores, First to Fifth Grade, by Group	10
11: Mathematics Percentile Scores of an Average Fifth Grade Student, Adjusting for School Context and First Grade Achievement, DI and Full Set of Control Schools	11
12: Mathematics Percentile Scores of an Average Fifth Grade Student, Adjusting for School Context and First Grade Achievement, DI and Reduced Set of Control Schools	12
13: Effect Sizes, Change from First to Fifth Grade in Mathematics Achievement Scores, DI and Control Schools	12

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in Baltimore City Public Schools

Low achievement in mathematics has been a major concern of school districts throughout the country. The Baltimore City Public School System (BCPSS) is similar to many other large city school districts that serve students with high levels of poverty and struggle with low levels of achievement. In the late 1990s, curricular reforms were implemented in the BCPSS elementary schools to address this low achievement. Eleven schools in the district introduced Direct Instruction as part of a whole school reform under the guidance of the National Institute for Direct Instruction (NIFDI).[1] The schools began with the reading curriculum and, after one to two years, introduced the mathematics curriculum. Two mathematics curricula were used, depending upon the students' level and needs: *Distar Arithmetic* and *Connecting Math Concepts*. Other schools in the district used a variety of curricular materials.

A large body of literature, including well designed meta-analyses, has demonstrated the effectiveness of Direct Instruction in promoting high achievement, although somewhat fewer studies address the mathematics curricula with data from an extended time period and from a relatively large number of schools. This paper addresses this area by examining data on the mathematics achievement of first and fifth graders in the BCPSS from 1998 through 2003. The first part of this report focuses on the mathematics achievement of first graders in the BCPSS from 1998 through 2003, comparing the achievement of students in schools that used the Direct Instruction to the achievement of other students in the district. The second part focuses on the subset of students who were in the same schools in the BCPSS in first and fifth grade and examines the extent to which experiencing the Direct Instruction curriculum in first grade is associated with higher achievement in fifth grade. In this report results are summarized in graphs and are also expressed as effect sizes. Appendices provide full statistical details.

A. First Grade Mathematics Achievement

The analysis of first grade mathematics achievement uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Achievement measures were obtained each spring using the Comprehensive Test of Basic Skills (CTBS), a widely used, nationally normed standardized test. Subtest measures of "Computations" and "Concepts and Applications," as well as a Composite measure were available. Results in this part of the report are presented as percentiles. Percentiles can be simply interpreted as the percentage of students that would have scores lower than a given student. For instance, if a student has a score at the 60th percentile, 60 percent of all students had scores that were lower.

[1] One additional school was included in the original group that implemented Direct Instruction. However, it closed very soon after the implementation began and thus it is not included in this count or in any of the analyses. In addition, a few other schools in the system used Direct Instruction reading programs. They, however, did not use DI mathematics curricula.

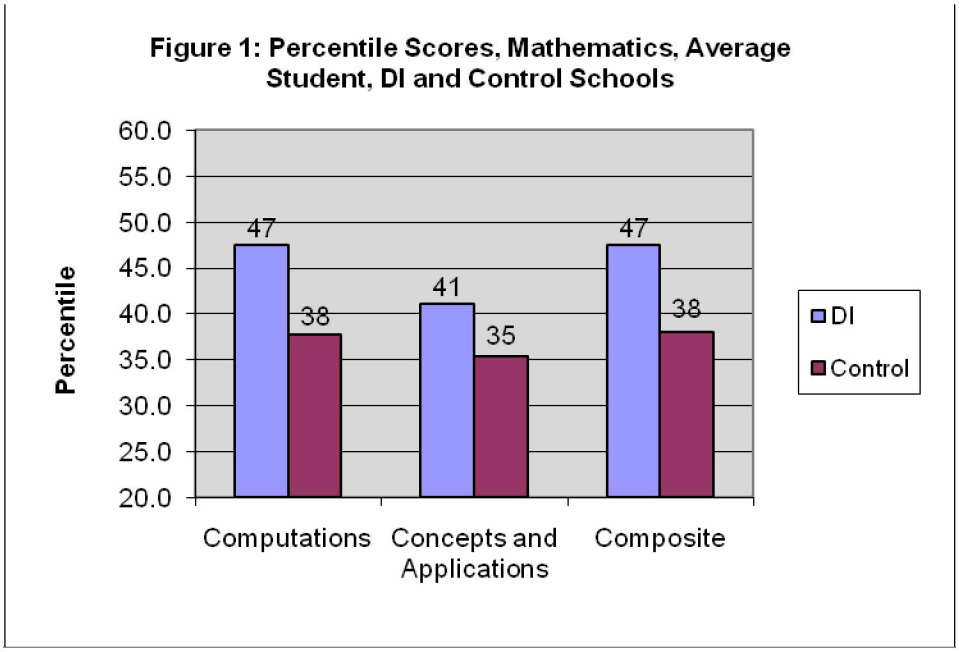
The average school in the BCPSS had large proportions of low income and minority students. On average, 75 percent of the students in a school were on free or reduced lunch. This varied however, from a low of 22 percent to a high of 93 percent. On average 84 percent of the students in average elementary school were African American, although the percentage varied from a minimum of 3 to almost 100 percent. Reflecting this high level of segregation, schools also varied in their representation of non-Hispanic white students, from having virtually no white students to being 94 percent white, with the average school having 14 percent of its students reported as being non-Hispanic white. Asian, Hispanic and Native American students were quite rare in the district, with an average representation of 1 percent in the student bodies. The DI Schools were similar to the Control Schools in racial-ethnic composition. However, the DI schools had proportionately more students who received free or reduced lunch: an average of 83 percent in the DI Schools compared to an average of 74 percent in the Control Schools.

The analysis focused on comparing the achievement of students who received Direct Instruction with those who did not. Full details on the statistical techniques and results are given in Appendix II. Four research questions were addressed:

- Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?
- Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?
- Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?
- What is the magnitude of the effect of have Direct Instruction on mathematics achievement?

1. Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?

Figure 1 summarizes achievement data for students who received Direct Instruction and those who did not on the three measures of mathematics achievement. These results combine the data from all the available years: 1997-98 through 2002-03 and represent the percentile score of the average student in a school. Combining data from all the years in the study, students who received Direct Instruction had significantly higher achievement scores than students in the Control Schools. The average first grader in a DI school scored at the 47th percentile, slightly below the national norm, in both the measure of computations and the composite score, but somewhat lower, at the 41st percentile, on the measure of concepts and applications. The average student in the Control schools had scores that were substantially lower, between the 36th and 38th percentile. These differences were statistically significant.

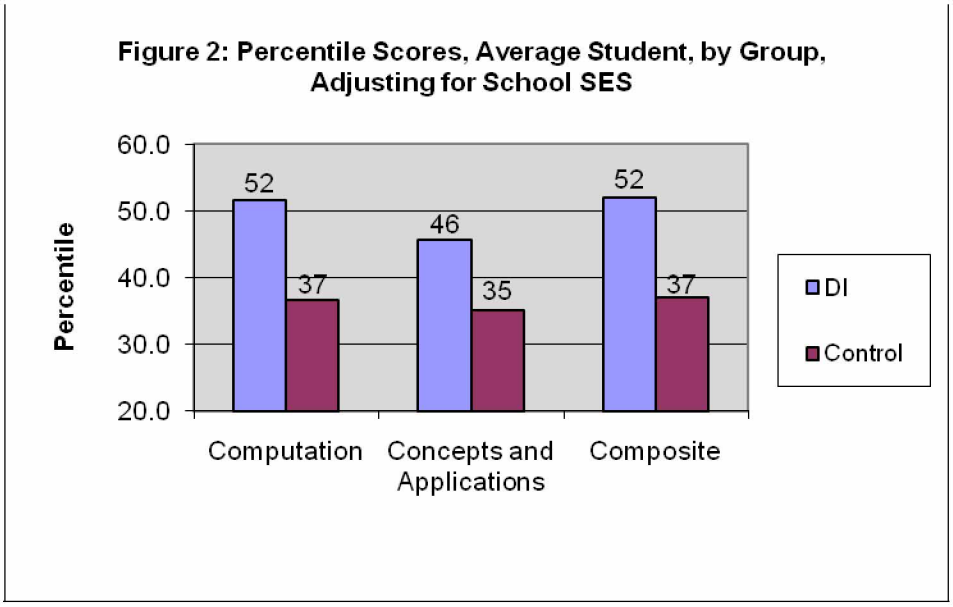


Note: Scores in this graph represent the percentile score of an average student in each type of school, averaging data over all years in the analysis: 1998-2003.

2. Did students in Direct Instruction Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?

Many studies have demonstrated that the socio-economic characteristics of a school have a strong influence on achievement, and thus it is important to control for this factor. As noted above, elementary schools in the Baltimore City Public School System varied substantially in their racial-ethnic composition and poverty rate, and, as in other studies students in BCPSS schools with fewer children from disadvantaged backgrounds had substantially higher average levels of achievement.

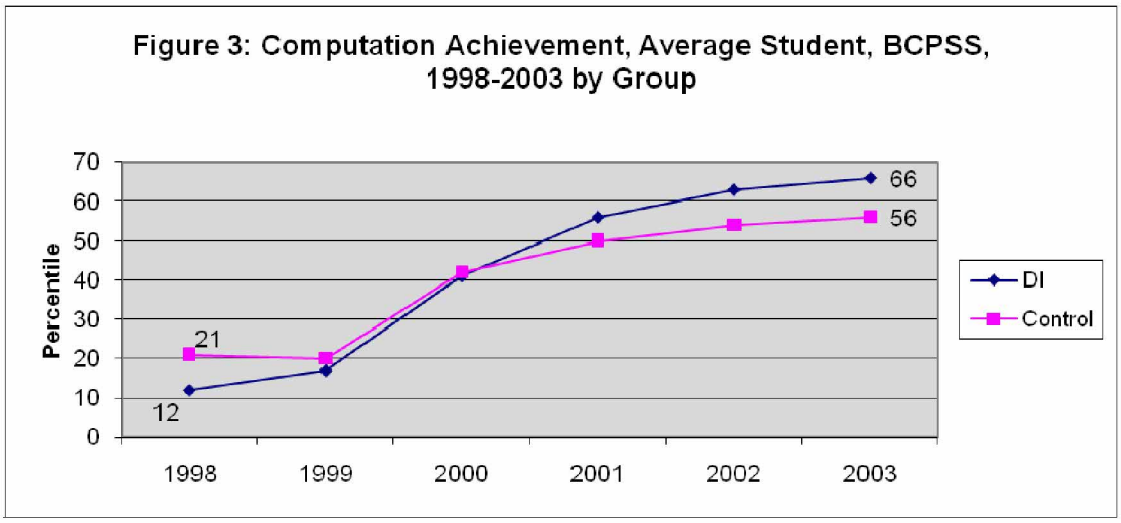
When the socio-economic characteristics of the schools were statistically controlled the differences in mathematics achievement between students in the DI schools and those in other schools became more marked. Figure 2 summarizes these results. The numbers in this graph represent the percentile score for an average student in each group if schools were equal in their racial-ethnic and poverty status. (Details regarding these calculations are in Appendix II.) After adjusting for the socioeconomic characteristics of the schools, the average student in a Direct Instruction School scored at the 52nd percentile on both the measures of computations and the composite score and at the 46th percentile on the measure of concepts and applications. In contrast, the average student at a Control School scored between the 35th and 37th percentile.

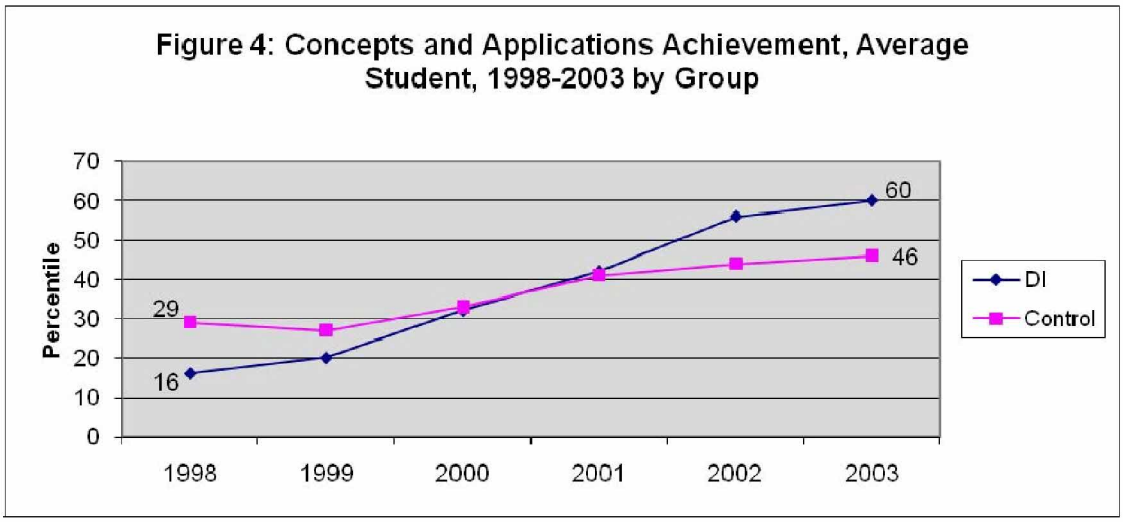


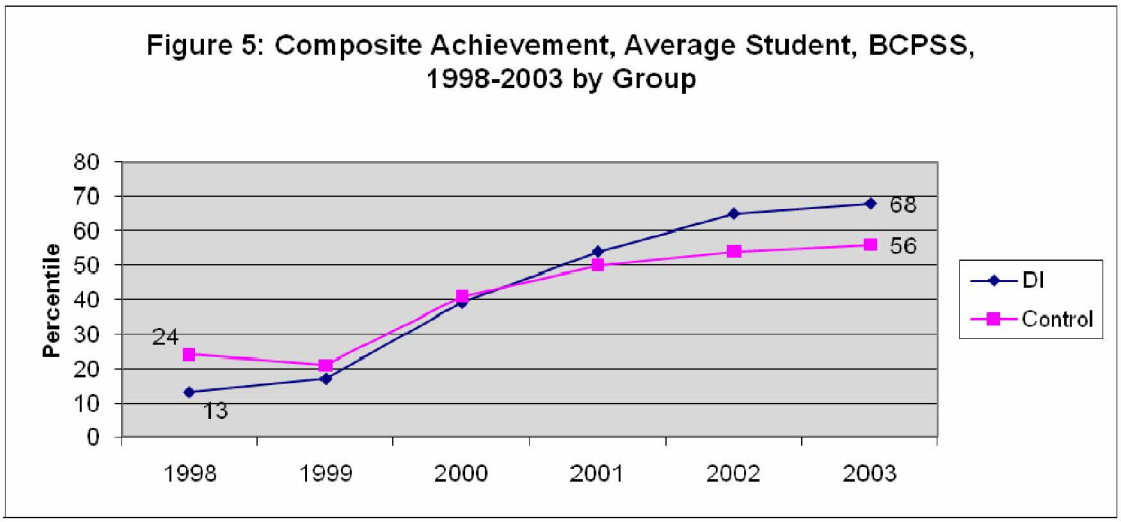
Note: Scores in this graph represent the percentile score of an average student in each type of school if the school contexts were equal in their socio-economic and demographic characteristics, averaging over all the years in the analysis: 1998-2003.

3. Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?

Changing school practices takes time and effort. It is reasonable to expect that the advantage of an effective instructional program would become more apparent as schools have more experience with a new curriculum, teachers gain more practice, and the procedures become part of the institutionalized and accepted practices within a school. During the time period under study the BCPSS paid concerted attention throughout the district to enhancing student achievement. The data indicate that these efforts were fruitful. Across the BCPSS first graders' mathematics achievement was significantly higher in 2003 than 1998. However, the changes over time were significantly larger for the students in the DI Schools than for those in the Control schools. Figures 3, 4, and 5 illustrate these changes. They show the percentile at which an average student in each group scored in each of the 6 years on each measure. Unlike the data in Figure 2, no adjustment was made for the school context of disadvantage.



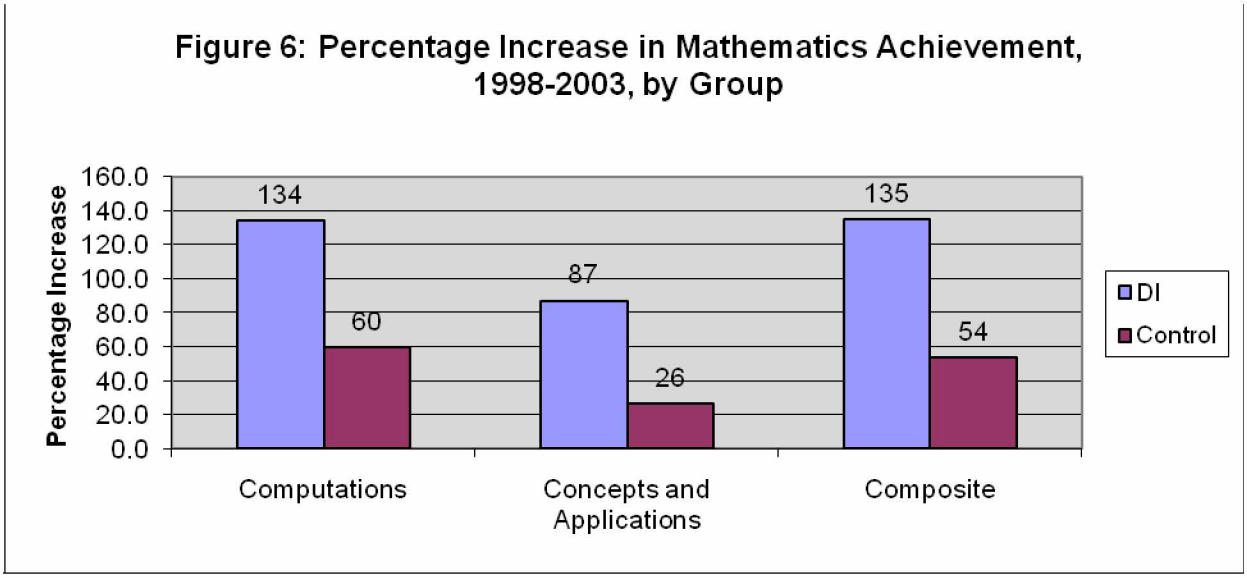




Note: Scores in Figures 3, 4, and 5 represent the percentile score of an average first grade student in each type of school for each year in the analysis.

At the beginning of implementation, in 1998, students in the DI Schools had significantly lower scores, on average, than students in the Control Schools. An average student in a DI school scored from the 12th (the computation and composite scores) to the 16th percentile (the concepts and applications score), while an average student in the Control Schools had scores ranging from the 21st to the 29th percentile. By 2003 the situation had changed markedly, with students in the DI schools having significantly higher scores than students in the Control Schools on all measures. By 2003, an average student in the DI Schools had scores were well above the national average, ranging from the 60th percentile on the measure of concepts and applications to the 68th percentile on the composite measure. In contrast, an average student in the Control Schools had scores ranging from the 46th percentile (on the measure of concepts and applications) to the 56th percentile.

Figure 6 illustrates the magnitude of these changes. Values in Figure 6 represent the difference between the 1998 and 2003 achievement scores as a percentage of the 1998 score, thus showing the percentage increase over time. For each measure, the changes over time were substantially larger for students in the DI schools than for students in the Control Schools. For instance, for the composite scores, first graders in DI schools had composite mathematics achievement scores in 2003 that were 135 percent higher than the scores of first graders in the same schools in 1998. In contrast, the scores of first graders in the Control Schools improved by only 54 percent. Similar differences appear with the other measures.



Note: The figures in this graph represent the percentage change in achievement from 1998 to 2003 for each of the groups of schools and each measure. These change figures were calculated using normal curve equivalent (NCE) scores rather than percentiles. As explained more fully in Appendix I, NCE scores are preferable for statistical procedures and were used for all calculations in the analysis. NCE scores were translated to percentiles for Figures 1-5.

4. What is the Magnitude of the Effect of NIFDI Supported Direct Instruction?

Unless specifically noted, all of the results presented in this report are statistically significant. That is, various statistical tests indicate that the results do not appear by chance. However, it is possible that statistical significance can simply reflect sample size. When data from large samples are subjected to statistical tests relatively small substantive differences can produce significant results. To combat this problem, researchers use measures of effect size. These measures provide descriptions of the magnitude of the results. One of the most commonly used measures is Cohen's *d*. It is simply calculated as the difference between two average values divided by the common standard deviation. Thus, it reports the magnitude of a difference between two groups as a proportion of the standard deviation. A d-value, or effect size, of .25 or larger has traditionally been considered educationally significant, indicating that an intervention has an important impact on students' achievement.

Figures 7 and 8 summarize the magnitude of the various effects reported earlier. Figure 7 summarizes the effects without any adjustment for school characteristics, using the data summarized in Figure 1. When data from all the years are examined together, the effect sizes comparing achievement of students in the DI Schools and those in the Control Schools are slightly lower than the .25 criterion (.13 for the measure of concepts and applications to .21 for the computations and composite scores). However, by 2003 the effect sizes were substantially larger, meeting or exceeding the .25 criterion: .25 for the measure of computations and.31 and .32 for the composite and concepts and applications measures.

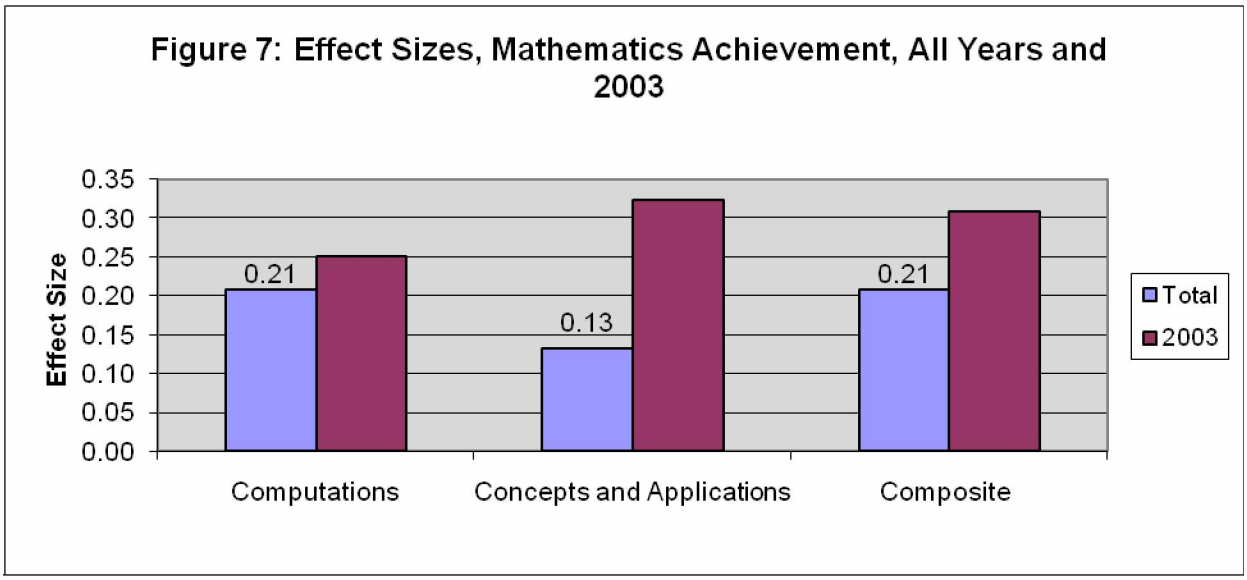
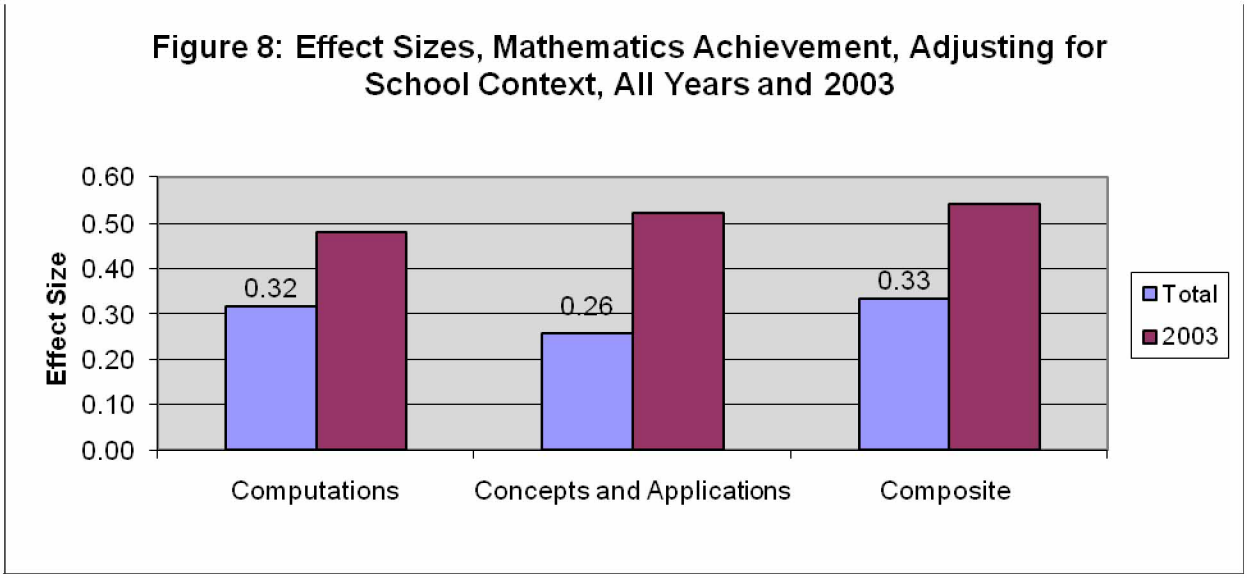


Figure 8 gives the effect sizes for the measure for all years and 2003 after the impact of school socio-economic status has been controlled (the data shown in Figure 2). Averaged across all years, the effect of being in a DI School relative to a Control School ranged from .26 for the measure of concepts and applications to 32 and .33 for computations and the composite score. At the end of the study period, the effect sizes were substantially larger: ranging from .48 to .54.



5. Summary

This analysis of the mathematics achievement of first grade students in the Baltimore City Public School System indicates that students in schools that implemented Direct Instruction had significantly higher achievement scores on all three indicators (computations, concepts and applications, and a composite score) than students in the Control Schools. These differences became even larger when the socio-demographic context of the schools was controlled and after several years of implementation of Direct Instruction. The effects were not just statistically significant, but substantively large and educationally important.

B. The Relationship of First Grade Direct Instruction and Fifth Grade Mathematics Achievement

Data were available for first and fifth graders in 1997-98 through 2002-03. While the first part of this report examined changes in first grade achievement over this time period, this section examines changes in achievement of individual students who began in the BCPSS in 1997-98 or 1998-99 and were still in the same schools five years later, in 2001-2 or 2002-3. As in the first part of the report, comparisons are made between students in the DI Schools and the Control Schools. (See Appendix III for full details on the sample used in the analysis.)

Two factors limited the size of the sample for this analysis: a high rate of turnover of students in the BCPSS and the fact that only six schools within the DI group had implemented Direct Instruction in mathematics by 1998-99. To provide additional insight into the results, two different control groups were used in the analysis: One includes all students in the Control Schools who started first grade in 1997-98 or 1998-99 and were in fifth grade in the same schools in 2001-02 or 2002-03. The second includes only students from schools with the lowest levels of achievement in the BCPSS in first grade in 1997-98 and 1998-99.[2] The DI group included 164 students, the total Control Group included 4,607 students, and the reduced Control Group included 1,820 students.

[2] Students who were retained, or held back a grade, were excluded from the analysis.

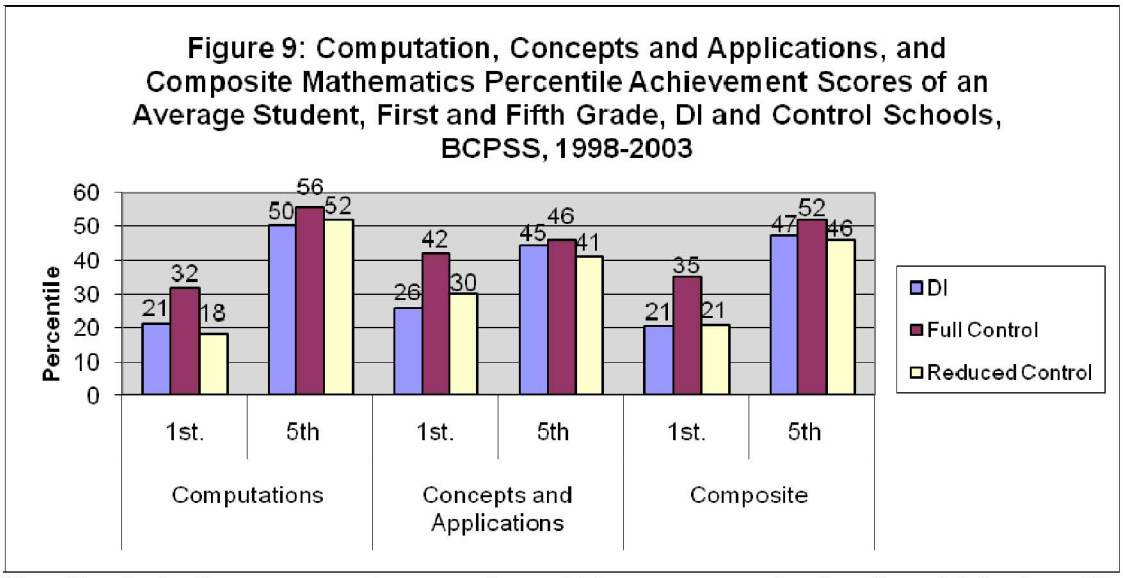
Three research questions were addressed:

- Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools?
- Did students who received Direct Instruction in first grade have higher mathematics achievement in fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?
- What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

1. Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools?

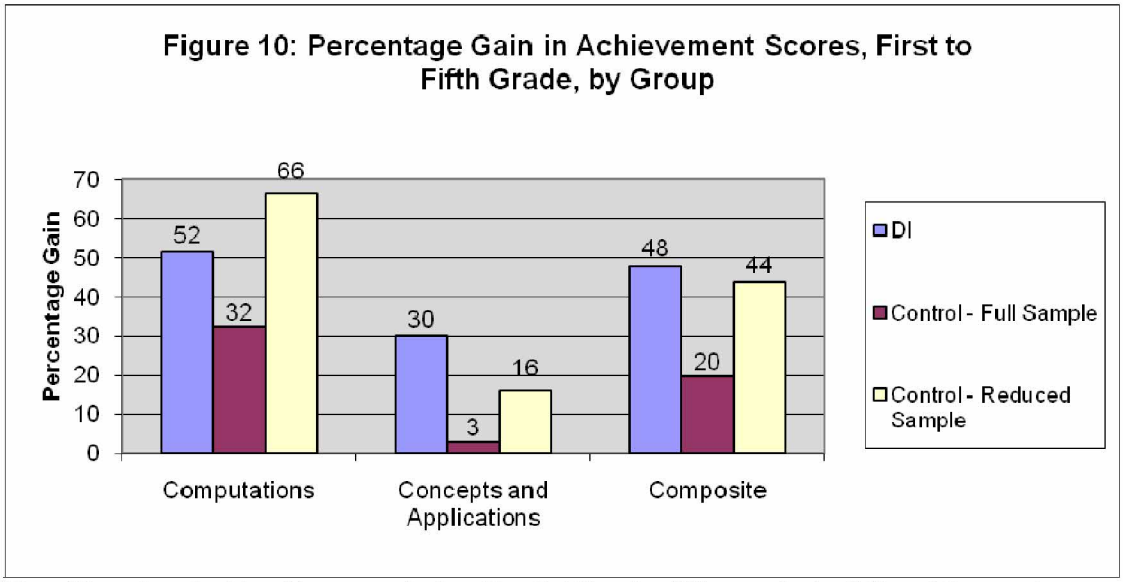
As explained fully in Appendix IV, several different statistical methods were used to examine differences in the changes in achievement from first to fifth grade of students who received DI and those who did not. All the results, whether with the full or reduced set of Control Schools, pointed to the same substantive conclusions: Students in the Direct Instruction schools had significantly greater gains than students in either of the set of Control schools on the measure of concepts and applications. Increases on the measure of computations and the composite measure did not differ significantly between the three groups. Figures 9 and 10 summarize these results.

Figure 9 reports the percentile score of an average student in each group of schools on each measure in both first grade and fifth grade. On average, all students, in both the DI and the Control Groups, had higher average achievement scores in fifth grade than in first grade. In first grade, the average student scored well below the national average, but by fifth grade, the average student, in all three groups, had scores that were much closer to, and in some cases above, the national average.



Note: Data in the figure represent the percentile at which an average student in a Control School or a DI school scored in first grade and in fifth grade. As with the analysis in Part A, all statistical calculations were conducted with Norm Equivalent Scores, which can be compared from one year to another, and the results were translated to percentiles for graphic display.

Figure 10 gives the percentage change in scores from first grade to fifth grade for students in each group on each measure. Students in the DI group had the highest average gain on both the measure of concepts and applications and the composite measure but, as noted above, these differences were only significant with the measure of concepts and applications. On average, students in the DI schools had concepts and applications scores in fifth grade that were 30 percent higher than their first grade scores, compared to an average increase of 3 percent for students in the full set of Control schools and 20 percent for students in the reduced set of Control schools.

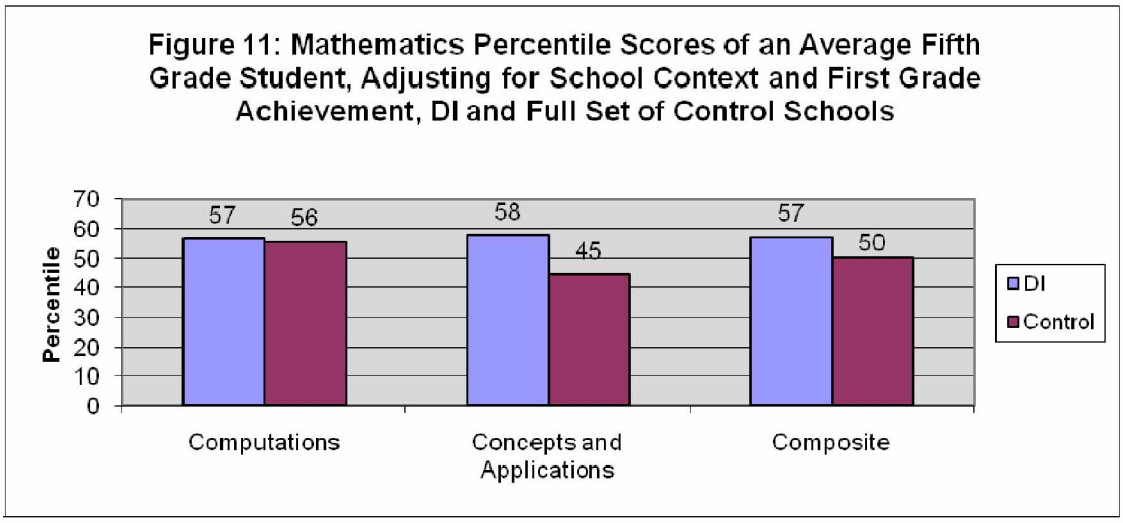


Note: The values in this table were calculated by dividing the difference in the fifth and first grade norm equivalent scores by the first grade score. Details on these calculations are in Appendix IV.

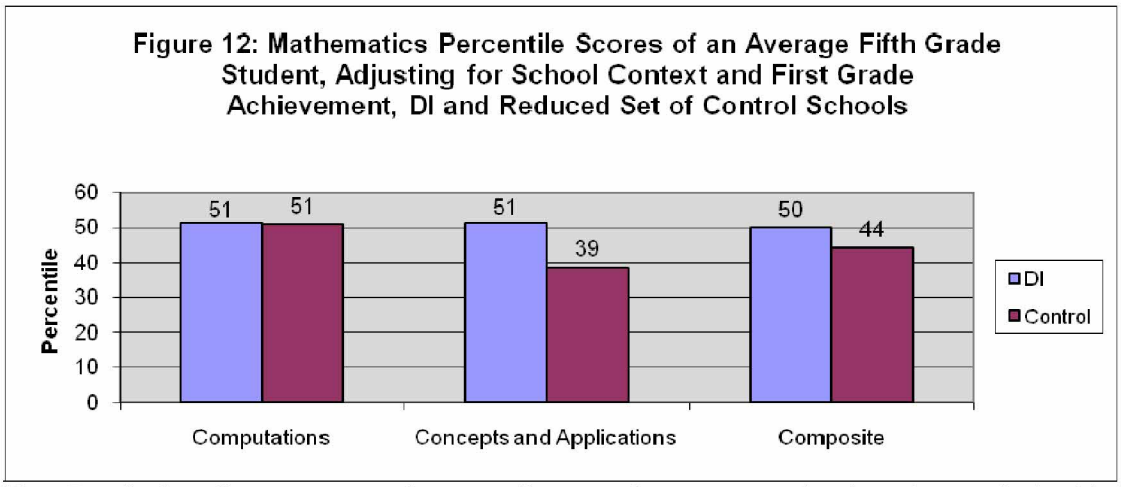
2. Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?

As described in Appendix IV, multivariate statistical techniques were used to address the second research question. This analysis compared the differences between students in the DI schools and those in the two sets of Control schools after statistically equalizing students on both their level of first grade achievement and the socio-demographic characteristics of their schools. The results were similar with analyses with both the full set of Control schools and the reduced set and parallel the results with Research Question 1: Students in Direct Instruction schools had significantly higher scores on the measure of mathematics concepts and applications than students in other schools, even when school context and first grade achievement were controlled. While the DI students also had higher scores on the measure of computations and the composite score, these differences were not statistically significant.

Figures 11 and 12 summarize the results of this analysis. The data in the figures are the percentile score that an average student in each group of schools would be expected to achieve if the schools had equal socio-demographic characteristics and if the students all had equal levels of achievement in first grade. Figure 11 compares the DI schools with the full set of Control schools, while Figure 12 compares the DI schools with the reduced sample of Control schools.[3] In both comparisons, the students in the DI Schools have higher levels of achievement, but the differences are substantially greater with the measure of mathematics concepts and applications.



[3] Different multivariate analyses were conducted with these two separate samples. The lower average values in Figure 12 reflect the fact that the reduced sample of Control Schools had lower levels of first grade achievement as well as a socio-demographic school context that was more high-risk.

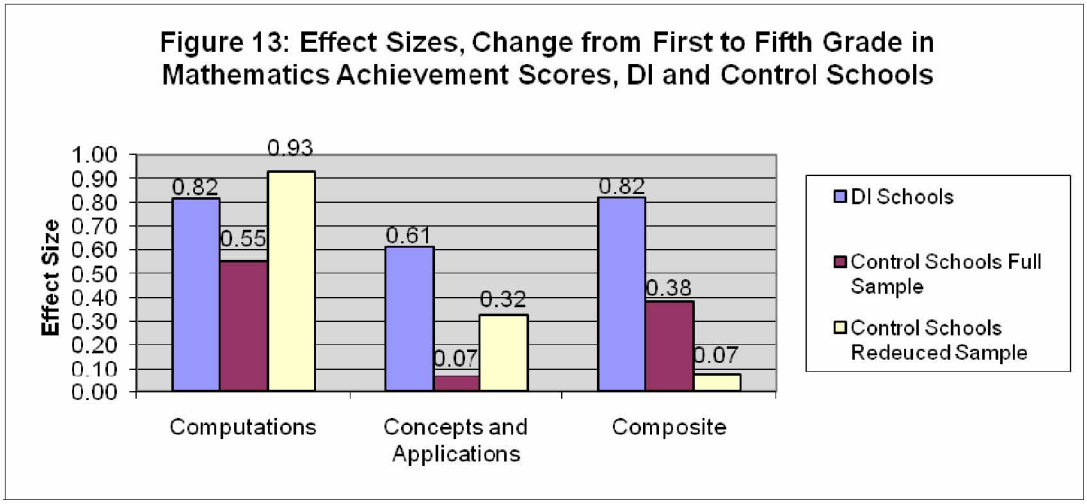


Note: Data in these figures represent the percentile score of an average student in each type of school in fifth grade if they had the same achievement scores in first grade and were in schools with equal socio-demographic characteristics.

3. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

Standard tests of statistical significance are influenced by the size of a sample in the calculations. To avoid this bias researchers sometimes use measures of effect sizes. With data involving change over time these calculations consider both the size of the scores at the two time points as well as the correlation between these two measures. (Details are in Appendix IV.) Traditionally, measures of effect size of .25 or larger are considered educationally important.

Figure 13 gives the effect sizes associated with the changes from first to fifth grade for each of the measures of achievement for students in the DI schools and for those in the two sets of Control schools. All of the effect sizes associated with the DI schools far surpass the .25 criterion. For the full set of Control schools, effects surpass the .25 criterion for both the measure of computations and the composite score. For the reduced set of Control schools, effects surpass the .25 criterion for the measure of computations and the composite score.



C. Summary and Discussion

The first section of the report examined changes in the average achievement of first graders in the BCPSS from 1998 to 2003, comparing students in schools that implemented Direct Instruction and students in other schools in the district. Results indicated that, averaging over all years, first grade students who received Direct Instruction had significantly higher levels of mathematics achievement than students who did not receive Direct Instruction. These differences became stronger when measures of the school context were statistically controlled and also became stronger when trends were examined over time. The students in the DI schools had achievement scores that were significantly lower than those in the Control schools at the beginning of the implementation (1998) but scores that were significantly higher by the end (2003). The percentage change in average first grade achievement was more than twice as high for students in the DI schools than in the Control schools. These differences in achievement were statistically significant and also met established criteria for educationally meaningful results.

The second part of the report looked at a smaller subset of students: those who began first grade in the BCPSS in 1997-98 or 1998-99 and remained in the same schools as fifth graders in 2001-02 or 2002-03. It and examined changes in the students' achievement from first to fifth grade. All students in the sample had higher levels of mathematics achievement in fifth grade than in first grade. However, students in the DI schools had significantly higher increases in their mathematics concepts and applications scores than other students. This result appeared when comparisons were made to students in all Control schools and in a more limited set of low achieving schools. They also continued to appear when strong controls for first grade achievement and school context were applied.

Several possible implications of these results should be noted: First, they provide additional support to the large body of studies that has documented the superiority of Direct Instruction in promoting mathematics achievement. Students in schools with Direct Instruction had significantly higher achievement than students in schools with other curricula. These results appeared in first grade on all measures and in fifth grade on the measure of mathematics concepts and applications. The differences met and, usually, surpassed the criteria of statistical significance and educational importance.

Second, the results show the necessity of examining a relatively long period of time to assess the full implications and results of school reform efforts. The changes in first grade achievement only occurred after several years of implementation within a school. In addition, it should be noted that all of the DI schools had strong technical support from the National Institute of Direct Instruction. Other work on reading achievement has shown that such strong technical support is important for helping to ensure that strong achievement gains occur. This result also may address a common claim that Direct Instruction involves a relatively mindless application of a script with little intellectual engagement or skill required by the teacher. The fact that strong changes in achievement only appeared over an extended period of time suggests that learning to teach Direct Instruction involves a great deal of skill and practice. Those that are most successful have more practice and more skilled, technical guidance.

Finally, it is important to note the impact of Direct Instruction on the measure of concepts and applications, typically assumed to be tapping more of a "higher order" skill than the measure of computations, the other element of the composite score. Some have suggested that Direct Instruction is useful for teaching rote, elementary skills such as computations, but has more limited utility in teaching higher level skills such as those tapped by the mathematics concepts and applications measure. The results in this paper would contradict that view, for it was in the area of concepts and applications that the long-term impact of Direct Instruction was most apparent.

Appendix I
Sample and Measures – First Grade Analysis

This appendix provides details on the sample and the measures used in the analysis of first grade mathematics achievement.

Sample

The analysis uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Eleven schools implemented Direct Instruction with technical assistance provided by NIFDI and are referred to below as the DI schools.[4] The remaining 103 schools are termed the Control schools. Schools in the intervention group are listed in Table A-1[5]

Table A-1 also lists the date at which programs were implemented in each school. All schools that used Direct Instruction began their implementation with the reading programs. They added DI instruction in mathematics in a subsequent year. In all but two schools (City Springs and Collington Square) the Direct Instruction mathematics curriculum was added one year after beginning reading instruction.

In the analysis schools were designated as belonging to a given condition only during the years in which their school was receiving the mathematics implementation. For instance, Collington Square had no mathematics DI instruction in either the first or second year of data collection (1997-98 and 1998-99), but began implementing Direct Instruction mathematics programs in 2000-2001. Thus, first grade students in Collington Square were included in the mathematics intervention group in 2000-2001 and subsequent years, but in the control condition in the two earlier years.

The analysis also considers the amount of time that a school had implemented the program. For instance, Collington Square was considered to be in its first year of implementation of Direct Instruction in 2000-2001, its second in 2001-2002, etc.

Data were available for a total of approximately 45,000 students (n = 44,766 for the composite measure, n = 45,287 for the measure of computation, and n = 45,535 for the concept and applications measure). Enrollment in the BCPSS declined from 1997-98 to 2000-2003, with data available for approximately 9000 students for the first year and 6400 for the last year. On average, data were available for approximately 380 students per school. Only four schools, all in the control group, had fewer than one hundred students in the sample.

---

[4] One additional school, Charles Carroll, was part of the original intervention group but was closed shortly after the start of the study period. Because data are not available throughout the time span of the study, data from that school are not included.

[5] There were two additional schools in the set of control schools, but they, unfortunately, had the same name: Highland Town. The number assigned to the schools was not available for all years, and alternative spellings of the schools' name across years made it impossible to clearly differentiate them. Thus these two schools were eliminated from the analysis. Data were available for one home-schooled student and that was also omitted.

Table A-1: Schools in Treatment Group and Start Dates of Reading and Mathematics Direct Instruction

	Reading Start Year	Math Start Year
Arundel	96 - 97	97 – 98
CC Barrister	97 - 98	98 – 99
City Springs	96 - 97	98 – 99
Collington Square	98 - 99	00 – 01
Dickey Hill	98 - 99	99 – 00
Federal Hill	97 - 98	98 – 99
General Wolfe	96 - 97	97 – 98
Hampstead Hill	96 - 97	97 – 98
Langston Hughes	98 - 99	99 – 00
Margaret Brent	98 - 99	99 – 00
Dr. Rayner Browne	98 - 99	99 – 00

Measuring Achievement

The Comprehensive Test of Basic Skills (CTBS), a widely used standardized achievement test, was administered to all first graders in the spring of each year, from 1997-1998 through 2002-2003. The 4th edition was administered in the spring of 1998 and 1999, and the 5th edition was administered in the spring of 2000 through 2003. Two subtest scores, Mathematics Concepts and Applications and Mathematics Computations, and a Composite Mathematics Achievement score were analyzed.

Normal Curve Equivalent (NCE) scores were used for all statistical analyses. NCE scores are calculated by translating the percentile scores to a distribution that is normally distributed. Like percentile scores, NCE scores range from 1 to 99 with a mean of 50. However, while percentile scores are evenly distributed (the graph of percentile scores would look like a rectangle), NCE scores comprise a normal distribution. The transformation results in scores that can be meaningfully added and subtracted, so that the difference of an NCE score of 50 and 55 (=5) is the same as the difference between 30 and 35. This interval scale allows computations of most common statistics. Percentiles and raw scores do not meet the technical requirements of interval scales, and thus statisticians urge researchers to analyze data based on NCE scores rather than percentiles or raw scores. All statistical calculations for this report were done using NCE scores.

The meaning of NCE scores is, however, not intuitively obvious. For this reason, results in the body of the text and the executive summary have been translated into percentile scores. After completing statistical calculations with NCE scores the resulting descriptive information

was converted into percentiles for display in the graphs in the body of the report. This conversion was done using a standard conversion table.[6]

Measuring School Context

Given the strong influence of school context on student achievement, it was important to develop an efficient, yet strong, measure of the demographic context of the schools in the sample. Preliminary analysis indicated that the demographic characteristics of schools were very highly correlated from one year to another. Thus, one summary measure was developed for each school that would be valid for all the years included in the data set.

There were a few cases in which data were not available for a school for all years. A regression-based method was used to predict values of missing cases from other years (e.g. predicting 1998 levels of proportion African-American from levels in 1999 through 2003).[7] Values for each variable were averaged across the years to produce aggregate measures of the demographic characteristics of the school.

Table A-2 gives means and standard deviations on each of the demographic characteristics for the DI and the Control schools. Results of t-tests indicated no significant differences between the groups of schools in their racial-ethnic composition. However, the DI schools had significantly more students who received free or reduced lunch: 83 percent in the DI schools versus 74 percent in the Control schools.

A factor analysis was conducted to develop summary measures of school context. Six variables were included in this analysis: the average proportion of Asian, Hispanic, African American, Native American, and non-Hispanic White students as well as the average proportion of students receiving free and reduced lunch. A principle components extraction method with varimax rotation was used. Two significant factors (with eigenvalues greater than 1.0) were found.[8] Factor scores were computed for each school and saved.

Results for the factor analysis are given in Table A-3. The commonalities indicate that, except for the measure of proportion of Asian students, there is a great deal of shared variance among the indicators, ranging from .66 for the proportion of students receiving free and reduced lunch to .94 for the proportion of African American students. Eigenvalues are also relatively high, with 73.5 percent of the total variance between schools explained by the two factors.[9]

---

[6] The conversion can also be accomplished manually through using a normal curve table. The NCE scores can be converted to z-scores (z = ((nce-50)/21.06))). The percentile that corresponds to the z score can then be found in the normal curve table.

[7] There were two cases where a predicted value fell outside of the theoretical range. Both of these involved the predictions for proportion white, where the predicted values were less than zero. For these cases (Malcolm X and Mildred Monroe schools), an average of the other years in the data file was used as the predictor. The average values for both cases were .01 or less.

[8] The factor analyses were also conducted with data for each year separately (i.e. the non-aggregated data), and the results were virtually identical to those obtained with the aggregated data.

[9] N = 120 for these analyses, for Carroll School was included.

Table A-2: Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels By Group

	DI		Control			
	Mean	S.D.	Mean	S.D.	t	prob.
Asian	0.01	0.03	0.01	0.01	-0.88	0.40
Black	0.75	0.34	0.85	0.26	1.12	0.27
Hispanic	0.05	0.12	0.01	0.01	-1.10	0.30
Native American	0.01	0.02	0.002	0.004	-1.31	0.22
Non-Hispanic White	0.17	0.25	0.13	0.25	-0.49	0.62
Free and Reduced Lunch	0.83	0.06	0.74	0.15	-4.02	<.001
Factor 1	-0.12	0.96	0.02	1.01	0.44	0.66
Factor 2	1.18	2.82	-0.12	0.46	-1.53	0.16
N	11		108			

The first factor, which accounts for 44% of the total variance, has a strong positive loading of the proportion of non-Hispanic White students, a slightly smaller positive loading for the proportion of Asian students, and strong negative loadings for the proportion of African American students and the proportion receiving free and reduced lunch. Thus, schools with positive scores on this factor would have proportionately more non-Hispanic White and Asian students and many fewer African-American and poor students. As can be seen in Table A-2, there are no significant differences between the two groups on this factor.

Table A-3: Factor Analysis of School Characteristics, Elementary Schools, BCPSS,1998-2003

	Rotated Factor Loadings		Communalities
	Factor 1	Factor 2	
Proportion Asian	0.64	0.09	0.42
Proportion African-American	-0.88	-0.40	0.94
Proportion Hispanic	0.13	0.85	0.74
Proportion Native American	0.08	0.87	0.77
Proportion Non-Hispanic White	0.90	0.27	0.89
Proportion Free/Reduced Lunch Eligible	-0.80	0.14	0.66
Eigenvalues	2.66	1.75	
% of total variance	44.4	29.1	

N = 120 schools

The second factor, which accounts for 29% of the total variance, has positive loadings on the proportion of Hispanics and the proportion of Native Americans in a school, and a negative loading, of a somewhat smaller magnitude, of the proportion of African Americans. Thus, schools with a positive score on this factor would have proportionately more Hispanic and Native American students and somewhat fewer African American students. The results in Table A-2 again indicate no significant differences between the groups on this factor.

Appendix II: Detailed Statistical Results – First Grade Analysis

This appendix expands upon the material presented in the text including an extended discussion of the analysis techniques employed and the results that address each research question.

Methodology

As noted in the text, the data analysis focused on comparing the mathematics achievement of first grade students who received Direct Instruction with those in the Control Schools. Four research questions were addressed.

The first research question was, "Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?" Simple t-tests were used to answer this question, comparing the average achievement scores of students in the DI schools with scores of students in the Control schools, combining data from all years of implementation.

The second research question asked, "Did students who received Direct Instruction have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?" Mixed model regression, including measures of school advantage and disadvantage from the factor analysis as predictors, was used to answer this question. To the extent that Direct Instruction is more effective at promoting student achievement, it would be expected that students in the DI Schools would have higher achievement scores than those in the Control Schools, even when school characteristics were controlled.

Mixed models are particularly appropriate for analyzing multi-level data, such as data regarding students and the schools that they attend. In these models a "random variable" is used to control for differences between schools (often termed the level 2 entity) while calculating regression coefficients regarding the impact of variables from both students and schools on achievement. The random variable is equivalent to having a separate intercept in the regression equation for each school. The coefficients associated with the various individual and school related variables are then calculated while this between school variance is controlled. The analysis also allows one to calculate the amount of variance in the dependent variable that occurs between schools and the extent to which various independent variables can account for this between school variance.

The third research question asked, "Did the impact of being in a DI school increase over time as reforms became institutionalized within a school?" It was expected that the impact of the DI implementation would become stronger over time. This question was analyzed both through examination of mean scores over time, with independent sample t-tests to compare the means, as well as through adding variables regarding length of implementation to the mixed models.

The final research question asked, "What is the magnitude of the effect of Direct Instruction on mathematics achievement." To address this question, the results were translated into effect sizes. Effect sizes are commonly used to summarize the magnitude of differences

between two groups. They are calculated by simply dividing the difference in the means of two groups by their common standard deviation. The resulting figure thus describes the magnitude of a difference between two groups as a percentage of the standard deviation. Unlike tests of significance, which can be heavily influenced by sample size, effect sizes are unaffected by the size of a sample. Effect sizes greater than .25 have traditionally been considered educationally significant.

Research Question 1: Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?

The first research question addressed was, "Do students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?" This was answered through comparing the average achievement scores of students in the DI schools with students in the Control schools. In this analysis data for all years in the sample (1997-98 through 2002-03) were combined.

Table A-4 summarizes the information used for these comparisons. Panel A reports the means and standard deviations of the normal curve equivalent scores for students who received Direct Instruction and students in the Control schools on each of the measures of achievement as well as the results of t-tests examining the hypothesis that any differences occur by chance. Panel B translates the average Normal Curve Equivalent scores in Panel A to the corresponding percentiles. The values in Panel B may be interpreted as the score of an average student in each type of school over all the years in the study and were used to construct Figure 1 in the text.

Table A-4: Average Mathematics Achievement Scores of Students With and Without Direct Instruction, All Years, First Graders, BCPSS

Panel A: Descriptive Statistics, Norm Equivalent Scores, and t-test Results

Computations	Control Schools	Direct Instruction
Mean	43.4	48.6
S.D.	24.9	25.8
N	42838	2449
t-value	9.73	
d.f.	2716	
prob.	<.001	
Concepts and Applications		
Mean	42.3	45.2
S.D.	22.1	22.9
N	43072	2463
t-value	6.40	
d.f.	45533	
prob.	<.001	

Composite

Mean	43.6	48.6
S.D.	24.0	25.4
N	42340	2426
t-value	9.47	
d.f.	2679	
prob.	<.001	

Panel B: Percentile Scores Corresponding to Mean Values

	Control School Students	Direct Instruction Students
Computations	37.7	47.4
Concepts and Applications	35.7	41.0
Composite	38.0	47.4

Note: For the composite measure and the measure of computations, the F-test comparing the variances of the two groups indicated that the hypothesis of equal variances should be rejected. Thus the t-values and degrees of freedom reflect the use of the unequal variance formulas for those two analyses.

The results in Table A-4 show that students in schools with Direct Instruction had significantly higher achievement scores than those in the Control Schools on all three measures: Computations, Concepts and Applications, and the Composite score. The average student in the Control schools, across all the years in the study, had scores ranging from the 35th to the 38th percentile. In contrast, the average student in a DI school had scores ranging from the 41st to the 47th percentile. The lowest scores for both groups occurred with the measure of Concepts and Applications. In neither of the groups did the average student reach the national average of the 50th percentile when scores were combined across all years.

Research Question 2: Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized?

Research Question 2 deals with the extent to which students in Direct Instruction schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized. (See Stockard and Mayberry, 1992, for an extensive discussion of the literature regarding such "school effects.") This question is addressed with mixed model regressions. Table A-5 summarizes the models that were tested and gives summary information on the results with each model. Panel A gives the variables that are included and Panel B gives the model fit statistics.

Table A-5: Variables in Models and Measures of Fit, Analysis of First Grade Mathematics Achievement, BCPSS, 1998-2003

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3	Model 4
Random intercept for Schools	x	x	x	X
Year		x	x	X
Factors 1 and 2		x	x	X
DI School			x	X
DI * Year				X

Panel B: Model Fit Statistics

Computations				
Random Effects Estimate	47.5	37.6846	38.0	37.8873
s.e.	6.5	5.1378	5.2	5.1685
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	582.9	515.3	514.5	514.3
s.e.	3.9	3.4	3.4	3.4
sig.	<.0001	<.0001	<.0001	<.0001
BIC	417320	411740.3	411646	411661.3
- 2 Log Likelihood	417306	411711.6	411660	411623.1
Change in LL		5594	52	37
Df		3	1	1
P		<.0001	<.0001	<.0001
PRE measure		0.12	0.001	0.0005
Correlation ratio	0.08			
Concepts and Applications				
Random Effects Estimate	46.4	34.8	34.5	34.6
s.e.	6.4	4.8	4.8	4.8
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	451.6	434.5	434.1	433.2
s.e.	3.0	2.9	2.9	2.9
sig.	<.0001	<.0001	<.0001	<.0001
BIC	408014	406238.6	406202.5	406112.9
-2 Log Likelihood	407999.7	406209.9	406169	406074.7
Change in LL		1790	41	94
Df		3	1	1
P		<.0001	<.0001	<.0001
PRE measure		0.04	0.001	0.0021
Correlation ratio	0.09			
Composite				

Random Effects Estimate	53.4	40.5	40.5	40.4
s.e.	7.3	5.5	5.5	5.5
sig.	<.0001	<.0001	<.0001	
Residual Estimate	533.2	474.8	474.0	473.47
s.e.	3.6	3.2	3.2	3.2
sig.	<.0001	<.0001	<.0001	
BIC	408554	403360	403293	403245
-2 Log Likelihood	508539	403331	403259	403207
Change in LL		105209	71	53
Df		3	1	1
sig.		<.0001	<.0001	<.0001
PRE measure		0.11	0.002	0.0012
Correlation ratio	0.09			

Note: The correlation ratio is computed by dividing the random effects estimate for schools by the sum of the estimate for schools and the residual. Thus, the ratio represents the proportion of total variation that is between schools. The PRE measure is the ratio of the difference of residual estimates of two models divided by the estimate from the less complex model. Thus it tells the proportionate change in the variance that occurs by adding more variables to a model.

Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average reading achievement. The correlation ratio attached to Model 1 is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that from eight to nine percent of the variance in achievement is between schools rather than simply between students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model and school differences are controlled. These null hypotheses can be easily rejected with Model 1 and with all subsequent models. There is significant variation between schools and also between students in all models that we examine. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those available to in this analysis.

The models become incrementally more complex, with each subsequent model adding one or more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the year in which data were collected and the two factor scores to the baseline model. Year was included to test the hypothesis that test scores vary over the years in the study (1997-98 through 2002-03). This is important to control for any general changes within the district. Including the two factor scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African

American students and more non-Hispanic white students, would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient).

Model 3 then adds the grouping variable – the indicator of attending first grade in a DI school or a Control School. It was expected that, net of the year data were gathered and the measures of school context, students in DI schools would have higher levels of achievement (a positive coefficient).

The -2 log likelihood measures and the BIC values in Panel B of Table A-5 can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the comparisons between these values are in the bottom part of each section of Panel B of Table A-7. For example, Model 2 provides a significantly better fit to the data than Model 1 for all measures. For computations, the change in the -2 Log Likelihood = 5594 (417,306 -41712). With three degrees of freedom (because three new variables were added to Model 2 compared to Model 1), this result is highly significant. The comparisons of Model 3 with Model 2 and of Model 4 with Model 3 also indicate that adding the variables in the more complex model significantly improves the fit. (See the significance associated with the change in the -2 log likelihood.)

The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. Looking at all the models in Table A-4 it may be seen that the lowest BIC values appear for Model 4. The Proportionate Reduction of Error (PRE) measures are another descriptive measure of the incremental changes in fit of the models and simply reflect the proportionate changes in the residual variance from one model to the next. The greatest changes occur from Model 1 to Model 2 and the proportionate reduction of error in prediction is substantially less for each of the more complex models. Again, this could be expected given the large number of factors that affect student achievement.[10]

Table A-6 gives the coefficients associated with each of the models in Table A-5, beginning with Model 2. As expected, the coefficients associated with Factor 1 (where higher scores indicate schools with fewer students on free and reduced lunch and more non-Hispanic whites) are positive and highly significant and the coefficients associated with Factor 2 (where higher scores indicate more Hispanic and Native American children) are negative and also statistically significant. The coefficient for year is positive and significant in all models, indicating that, over time, students' achievement scores increased.

The coefficients in Model 3 directly test the second research question by including both a dummy variable for treatment group and the controls for the measures of school context. It should be recalled that students were only coded as having the treatment if they were in a school in a year when the school was in a treatment condition. The coefficients associated with being in a DI school are positive and highly significant ($p < .0001$) in analyses of all three measures of

[10] It should be noted that the values of BIC, the changes in the Log-Likelihood Ratio, and the PRE measures would alter if the order in which variables were introduced were changed.

achievement. In other words, once the demographic context of the school is controlled, students who received Direct Instruction in mathematics had significantly higher achievement than students in the Control Schools.

Table A-6: Coefficients Associated with Mixed Model Regressions of First Grade Mathematics Achievement, BCPSS, 1998-2003

	Model 2			Model 3			Model 4		
Computations	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-9754.1	127.1	<.0001	-9568.6	129.1	<.0001	-9442.3	131.7	<.0001
Year	4.9	0.1	<.0001	4.8	0.1	<.0001	4.7	0.1	<.0001
Factor 1	1.6	0.6	0.005	1.6	0.6	0.005	1.7	0.6	0.004
Factor 2	-0.8	0.6	0.16	-1.7	0.6	0.004	-1.5	0.6	0.009
DI	-----	-----	-----	7.9	1.0	<.0001	-3090.9	653.0	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	1.5	0.3	<.0001
Concepts and Applications	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-4902.4	116.6	<.0001	-4768.7	118.4	<.0001	-4532.7	120.7	<.0001
Year	2.5	0.1	<.0001	2.4	0.1	<.0001	2.3	0.1	<.0001
Factor 1	2.6	0.6	<.0001	2.7	0.6	<.0001	2.7	0.6	<.0001
Factor 2	-1.0	0.6	0.06	-1.7	0.6	0.003	-1.4	0.6	0.01
DI	-----	-----	-----	5.7	0.9	<.0001	-5824.1	599.9	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	2.9	0.3	<.0001
Composite	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-9060.7	122.7	<.0001	-8873.5	124.6	<.0001	-8687.6	127.2	<.0001
Year	4.6	0.1	<.0001	4.5	0.1	<.0001	4.4	0.1	<.0001
Factor 1	2.2	0.6	0.0002	2.3	0.6	0.0001	2.3	0.6	0.0001
Factor 2	-1.0	0.6	0.09	-1.9	0.6	0.002	-1.7	0.6	0.005
DI	-----	-----	-----	8.1	1.0	<.0001	-4568.5	631.2	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	2.3	0.32	<.0001

Table A-7 shows the predicted achievement scores for students in the DI and Control Schools based on the results in Model 3. These values were calculated by substituting average values for the Factor Scores and Year in the models and calculating predicted values for students in each group. Thus, the resulting values give the achievement score that would be predicted if students were equalized on the factor scores and time. Panel A gives the Normal Curve Equivalent scores and Panel B gives the percentile equivalent of these scores. Values in Panel B may be interpreted as the percentile at which an average student in each of group would be expected to achieve if schools were equal in their demographic characteristics. The values in Panel B were used in Figure 2 in the body of the text. The results in Table A-7 indicate the

magnitude of the advantage that accrued to students in the DI Schools for they have higher average scores on all three measures and surpass the national mean on both the measure of computations and the composite score.

Table A-7: Predicted Achievement Scores Assuming Equal School Contexts

Panel A: Norm Equivalent Scores

	DI	Control
Computation	50.8	42.8
Concepts and Applications	47.7	42.0
Composite	51.1	43.0
Panel B: Percentile Scores		
Computation	51.5	36.6
Concepts and Applications	45.6	35.2
Composite	52.0	37.0

Note: Scores were computed by substituting average values for the entire sample for Year and Factors 1 and 2 into the equations associated with Model 3 in Table A-6. The average value for year was 2000.3. The average value for Factor 1 was -.03203, the average value for Factor 2 was -.04252. The value for year does not equal 2000.5 (the average of 1998-2003) because there were fewer students in the BCPSS in later years than in earlier years. The average values for Factors 1 and 2 do not equal zero because the unit of analysis for computing the average values was students rather than schools. If the alternative values were used in computations, the substantive results would not differ.

Research Question 3: Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?

Research question three tests the hypothesis that the advantage to students in a DI School would become stronger as practices and procedures were institutionalized within a school. As noted in the text, it takes time for teachers and other school staff to fully adjust to and incorporate the nuances of a new curriculum. Thus, it would be expected that the advantages accruing to attending a DI School would become greater over time.

Model 4 in Tables A-5 and A-6 addresses this hypothesis by adding the interaction of attending a DI school and year. It was expected that the coefficient associated with the interaction of year and DI would be positive, indicating stronger effects of Direct Instruction as years of treatment increased. This hypothesis was confirmed. In each analysis the interaction term is positive and highly significant.

The coefficients associated with Model 4 in Table A-7 can be used to calculate the average change expected in the three groups of schools from one year to the next. For instance, first grade students in the control schools had an average gain from one year to the next of 4.4 points in their normal curve equivalent scores on the composite measure of mathematics achievement (the coefficient associated with year), controlling for differences in the demographic context of the schools. In contrast, the scores of first grade students in the DI Schools had an average gain of 6.7 normal curve equivalent points with each additional year of implementation of Direct Instruction (summing the impact of year and the impact of years of implementation: 4.4 + 2.3).

Table A-8 gives the scores that would be predicted for students in the two groups of schools if they attended schools with similar demographic characteristics in 1998, the first year for which data are available, and 2003, the last year. These scores were obtained by using the regression coefficients in Table A-6 for Model 4 and substituting the average factor scores, as was done for the calculations reported in Table A-9, and either 1998 or 2003. Panel A gives the resulting normal curve equivalent scores and Panel B translates these scores into the corresponding percentiles. The scores in Panel B can be interpreted as the percentile that an average student in a group of schools would be expected to achieve at a given year.

Table A-8: First Grade Mathematics Achievement Scores by Group and Year, Adjusting for School Advantage, BCPSS

Panel A: Norm Equivalent Scores

	Year	DI	Control
Computations	1998	35.0	32.0
	2003	66.5	55.7
Concepts and Applications	1998	33.2	36.7
	2003	59.2	48.2
Composite	1998	33.8	33.1
	2003	67.1	54.9

Panel B: Expected Percentile Score of Average Student

	Year	DI	Control
Computations	1998	24	20
	2003	78	61
Concepts and Applications	1998	21	26
	2003	67	47
Composite	1998	22	21
	2003	79	59

Even after just one year of implementation, in 1998, the average scores of students in the DI schools, after adjustments for the socio-economic status of the schools, are higher than those of students in the other schools on the computation and composite measures. By 2003 these differences become much stronger and more dramatic. Assuming that they attended schools with similar socio-economic contexts, by 2003 the average student in a DI School would be expected to score between the 67thth and 79th percentile on the various measures, while the average student at Control School would score from the 47th to the 61st percentile, depending on the measure of achievement.

Table A-9 presents the data that were used to construct Figures 3, 4, and 5 in the text. Unlike data in Tables A-7 and A-8, these data are simply the average achievement scores in each year on each measure for each group. No adjustments were included for school advantage/disadvantage. They simply compare the average scores in the schools in the two groups from 1998 to 2003. The t-tests in each row test the null hypothesis that the mean NCE scores of students in the two groups are equal.

The results show that students in the DI schools had significantly lower scores than students in the control schools in the first two years for which data are available (1998 and 1999) for the concepts and applications and the composite measures and the first year (1998) for the computation measure. However, by the end of the study period the situation had totally reversed. Students in the DI Schools had significantly higher scores than students in the Control Schools on both the computation and composite measure from 2001 to 2003 and on the concepts and applications measure in 2003.

Another way to examine the change in scores is to calculate the percentage change in average NCE scores from 1998 to 2003. The last row of each section of Panel A gives these figures for each group and measure. The percentage change in average scores was much larger for students in the DI Schools than in the Control Schools. For instance, with the composite achievement measure, average normal curve equivalent scores for students in the DI Schools increased by 135 percent, while scores for students in the Control Schools increased by only 54 percent, less than half as much. Differences between the three groups were most marked in the measure of Concepts and Applications. Scores of students in the DI Schools increased by 87% over the 6 year period, while those of students in the Control Schools increased by only 26 percent. (These data were used to create Figure 6 in the text.)

Panel B of Table A-9 translates the Normal Curve Equivalent scores in Panel A into the corresponding percentiles. (These percentiles were used to create Figures 3 through 5 in the text.) The percentile scores confirm the pattern that appears in the other analyses. Students scores increased in both groups, but the increases were substantially stronger for students in the DI Schools. For instance, on the composite measure of mathematics achievement the average student in the DI Schools scored at the 13th percentile in 1998, but by 2003 the average student in these schools scored at the 68th percentile. In contrast, the average student in the Control Schools went from the 24th percentile in 1998 to the 56th percentile in 2003. In general, the raw, unadjusted scores confirm the more efficient and parsimonious results obtained with the mixed models, which incorporate the statistical controls for school context.

Table A-9: Mathematics Achievement Scores by Year and Group, First Graders, BCPSS, 1998-2003

Panel A: Means and t-test Results

	Group				
	DI	Control	t-value	df	prob.
Computations					
1998	25	33	-4.76	176	<.001
1999	30	32	-1.27	343	0.21
2000	45	46	-0.56	7700	0.58
2001	53	50	2.26	608	0.02
2002	57	52	4.57	570	<.001
2003	59	53	5.01	530	<.001
Percent Increase	134	60			
Concepts and Applications					
1998	29	38	-5.64	176	<.001
1999	32	37	-4.42	342	<.001
2000	40	41	-0.10	7851	0.92
2001	46	45	0.91	7352	0.36
2002	53	47	5.97	6805	<.001
2003	55	48	5.98	518	<.001
Percent Increase	87	26			
Composite					
1998	26	35	-5.71	174	<.001
1999	30	33	-3.09	334	0.002
2000	44	45	-0.51	7610	0.61
2001	52	50	1.93	7248	0.05
2002	58	52	5.58	568	<.001
2003	60	53	5.89	523	<.001
Percent Increase	135	54			

Panel B: Percentile Equivalents of Norm Equivalent Scores

	DI	Control
Computations		
1998	12	21
1999	17	20
2000	41	42
2001	56	50
2002	63	54
2003	66	56

Concepts and Applications		
1998	16	29
1999	20	27
2000	32	33
2001	42	41
2002	56	44
2003	60	46
Composite		
1998	13	24
1999	17	21
2000	39	41
2001	54	50
2002	65	54
2003	68	56

Note: In Panel A, the smaller degrees of freedom reflect cases where F tests revealed that the variances of the two groups were unequal and the alternative formula for calculating t and degrees of freedom was used.

Research Question 4: What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?

The fourth and final question asked, "What was the magnitude of the effects that were found?" Cohen's d, a standard measure of effect size, was calculated to provide a descriptive measure of the extent to which students in the NIFDI supported schools had achievement scores that were superior to students in the other schools. Cohen's d is simply the difference between the average score of two groups divided by the common standard deviation. Traditionally, d scores of .25 or larger have been deemed educationally significant, although McLean and associates (2000) caution that effect sizes calculated with normal curve equivalent scores are conservative in nature.

Table A-10 gives the d scores comparing the average achievement of students in the DI and Control Schools**.** Effect sizes are given in Panel A for the unadjusted scores. Effect sizes in Panel B were calculated with school context controlled, based on the results in Tables A-7 and A-8. The standard deviations for the total group were used for all calculations, and all calculations were made with the norm equivalent scores.

With the unadjusted scores combined over all years, the effect sizes are positive, but below the .25 threshold. The effect sizes for 1998 are negative and surpass the .25 criterion, indicating that the students in the DI Schools had substantially lower achievement than students in the Control Schools at the beginning of the implementation. However, after 2000 all of the

effect sizes are positive and by 2003 they all surpass the .25 threshold. In other words, while the d score at the beginning of the study period indicates an educationally significant disadvantage for students in the Direct Instruction Schools, this had totally reversed by the end of the study period and the students in the DI Schools had an educationally significant advantage.

Table A-10: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Mathematics Achievement, BCPSS, 1998-2003

Panel A: Gross Effect Sizes

	Computations	Concepts and Applications	Composite
Total	0.21	0.13	0.21
1998	-0.33	-0.38	-0.39
1999	-0.07	-0.22	-0.16
2000	-0.03	0.00	-0.02
2001	0.11	0.04	0.09
2002	0.23	0.28	0.27
2003	0.25	0.32	0.31

Panel B: Effect Sizes Independent of School Context

	Computations	Concepts and Applications	Composite
Total	0.32	0.26	0.33
1998	0.12	-0.15	0.03
2003	0.48	0.52	0.54

Similar, but somewhat stronger, results occur with the measures derived from the results where school context was statistically controlled (see Panel B). These effect sizes far surpass the threshold of educational significance by the 7th year of implementation, ranging from .48 to .54.

Thus, the results with the effect sizes parallel the results with the other statistical analyses. Students in DI schools had higher achievement than students in the Control Schools, and the magnitude of this advantage increased substantially over the years of implementation.

Appendix III
Panel Sample: Changes from First to Fifth Grade

This appendix provides details on the sample used in the analysis of changes in achievement from first to fifth grade. The potential pool included only students who were in the BCPSS in 1997-98 or 1998-99 and thus could still be in the system in fifth grade in 2001-02 or 2002-03, the last years for which data were available. Thus, data for this study included student cohorts who began first grade in the BCPSS in 1997-98 or 1998-99 and were in the same schools in fifth grade in either 2001-02 (for the first graders in 1997-98) or 2002-03 (for the first graders in 1998-99). Students who were held back were not included.

There was substantial student turnover during the study period, as summarized in Table A-4. There were almost 18,000 students in first grade in the BCPSS in 1997-98 and 1998-99. Five years later over half of these students (52.8%) had left the BCPSS, and an additional 20 percent had transferred to another school within the system. Only 27 percent of the students who were in first grade in the system in 1997-98 and 1998-99 were still in the same school in 5th grade in 2001-2 or 2002-3. The panel sample included a total of 4,771 students. Of these, 164 were in DI Schools and the remaining 4,607 were in the Control Schools.

Table A-11: Turnover in the BCPSS, 1997-98, 1998-99 to 2001-02, 2002-03

	N	Percent
Left BCPSS (In first grade, but not in fifth)	9362	52.80%
Transferred within BCPSS (In BCPSS in fifth grade, but in a different school)	3595	20.30%
Panel Sample (Same School, first and fifth grade)	4771	26.90%
Total Number of Students In First Grade 1997-98 or 1998-99	17,728	100.00%

Students were considered to have received an intervention only if they could be considered to have begun receiving Direct Instruction in first grade.[11] As summarized in Table A-12, this limitation of the sample resulted in students in several of the schools listed in Table A-1 not being included within the intervention group. For example, students who were fifth graders

[11] Analyses were also conducted that considered students who began DI instruction in second grade as part of the implementation group as well as at any time in their elementary career. The results with these comparisons were substantively similar to those reported here. It was decided, however, to limit the intervention sample only to those who had received DI throughout their first 5 elementary years to provide the least contaminated test possible of the effect of the curriculum.

in 2001-02 or 2002-03 at Collington Square would not have had Direct Instruction in first grade (the 1997-98 or 1998-99 school year) and thus were included in the control condition. Students who were fifth graders in 2001-02 in City Springs, Barrister, and Federal Hill did not have Direct Instruction in mathematics in first grade and were included in the control group. But, students who were fifth graders in 2002-03 in these schools did have first grade Direct Instruction and were included in the treatment group.

Table A-12: Schools in Treatment Group by Year for Panel Analysis

School	Year Started DI in Math	Fifth Graders in 2001-02	Fifth Graders in 2002-03	N in Panel Sample
Arundel	97 – 98	yes	yes	37
General Wolfe	97 – 98	yes	yes	23
Hampstead Hill	97 – 98	yes	yes	55
CC Barrister	98 – 99	no	yes	23
City Springs	98 – 99	no	yes	16
Federal Hill	98 – 99	no	yes	10
Collington Square	00 – 01	no	no	
Dickey Hill	99 – 00	no	no	
Langston Hughes	99 – 00	no	no	
Margaret Brent	99 – 00	no	no	
Dr. Rayner Browne	99 – 00	no	no	

The pattern of persistence within the same school differed somewhat between the groups of schools, as shown in Table A-13, with the Direct Instruction Schools having a higher proportion of their students remaining within the same schools from first to fifth grade[12] About a third of the students in first grade remained in the same schools in fifth grade, compared to slightly over one-fourth of those in the Control Schools. These differences were statistically significant.

As would be expected the students in the panel group (those that were in the same schools in 1st and 5th grade) differed from other students in their achievement levels in first grade. As shown in Table A-14, students who persisted in the schools in which they attended first grade had significantly higher first grade mathematics achievement on all three scores. These differences appeared for both the Direct Instruction Schools and the Control Schools, but were slightly stronger for the Control Schools than for the DI Schools (see the marginally significant interaction effects for concepts and applications and composite scores).

[12] There was no way to control for the possibility that a student had attended multiple other schools between first and fifth grade. It is possible that some of the students in the panel sample were in the targeted schools in both first and fifth grade, but had attended other schools in the interim. Assuming that such children would have less of the "full treatment," this would bias results in a conservative direction.

Table A-13: Persistence in the Same School, by School Group

	DI	Control Schools	Total
Not in Panel Sample	66.8	73.3	73.1
In Panel Sample	33.2	26.7	26.9
Total (%)	100	100	100
N (first grade)	494	17234	17728
N (panel)	164	4607	4771

chi-square = 10.209, df = 2, p = .001

The results in Table A-14 also indicate that the students in the DI Schools had substantially lower average levels of first grade achievement on each of the measures than students in the Control Schools. These differences were highly significant.

Table A-14: Comparison of Achievement Scores of Panel Group and Other First Graders, DI and Control Schools

	Computations		Concepts and Applications		Composite	
DI Schools	Mean	s.d.	Mean	s.d.	Mean	s.d.
Non-Panel Sample	26.2	22.9	28.8	19.5	25.7	20.1
Panel Sample	33.3	21.9	36.2	19.0	33.0	20.1
Control Schools						
Non-Panel Sample	29.9	24.1	34.7	22.8	30.9	22.7
Panel Sample	40.2	25.1	46.2	22.2	42.2	23.0
ANOVA Results	F	p	F	p	F	p
Panel vs. Non-Panel	53.16	<.001	71.66	<.001	68.34	<.001
Treatment Group	20.14	<.001	50.89	<.001	40.49	<.001
Interaction	1.79	0.18	3.25	0.07	3.21	0.07

Note: Mean and standard deviation values in the table are computed from Normal Curve Equivalent (NCE) Scores.

In order to minimize the differences between the Control Schools and the DI schools the sample of schools included within the control group was also limited to produce a subsample that had a mean as close as possible to that of the DI schools. This was accomplished by rank

ordering all schools by their average composite mathematics achievement score and then selecting the subsample of schools (beginning with the lowest) that produced an average as close as possible to that of the Direct Instruction schools. The resulting sample includes the 164 students in the DI Schools, as noted in Tables A-12 through A-14, and 1,820 students in the Control Schools. The sample includes 50 different schools. The minimum number of cases from a school is 16, and the maximum is 101. There are no significant differences in first grade achievement between these two groups. Results with both groups of Control Schools are presented in Appendix IV and in the body of the text.

Appendix IV: Detailed Statistical Results – Fifth Grade Achievement

This appendix expands upon the material presented in the text, including a more extended discussion of the analysis techniques employed and the results related to each of the three research questions regarding the impact of DI instruction in first grade on mathematics achievement in fifth grade.

Analysis Methodology

The analysis focused on changes in students' mathematics achievement from first to fifth grade and the extent to which students in schools with Direct Instruction had greater gains over time. Three research questions were addressed. The first was, "Do students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools?" Several comparisons were used to examine this question. Simple paired t-tests were used to examine changes from first to fifth grade in both sets of schools. Second, independent sample t-tests were conducted to compare the average achievement scores of students in the DI and Control schools in both first and fifth grade. Third, a repeated measures analysis of variance was conducted with the first and fifth grade scores as repeated measures and the school as a factor. If gains were greater in the DI schools, the interaction between the repeated measures would be expected to be significant.

The second research question was, "Do students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school is controlled?" This question was addressed by employing a mixed model regression, having schools as a random variable and regressing students' fifth grade scores on first grade scores, the measures of a school context of advantage/ disadvantage, and school type (DI or Control). Successively more complex models were used and differences between these models were examined.

The third research question asked, "What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?" To answer this question, the results were translated into effect sizes. Unlike tests of significance, effect sizes are not influenced by the size of a sample. Because the issue of interest in this report is the change in scores from one year to another the standard method of calculating effect sizes (dividing the difference between two means by the common standard deviation) is not appropriate. Instead, the correction suggested by Dunlap and associates (1996) is used. The calculations involve an adjustment for the correlation between the two scores. Results may be interpreted in the same manner as Cohen's d. Effect sizes greater than .25 have traditionally been considered educationally significant (Fashola and Slavin 1990). However, McLean and associates (2000) caution that effect sizes calculated with norm equivalent scores, as with this analysis, are inherently smaller than with other metrics. Thus the effect sizes presented are a conservative estimate of the results.

For all analyses results are presented comparing the students in the Direct Instruction Schools with students in all the Control Schools as well as those in the schools with more similar levels of achievement in first grade. The latter group is referred to as the "reduced sample" of control schools.

Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools?

Three different approaches were used to address the first research question: paired t-tests, independent sample t-tests, and repeated measures analysis of variance.

Difference Scores and Paired t-tests

Table A-15 summarizes the data used to examine differences in achievement gains from first to fifth grade for students in the Direct Instruction and Control Schools using paired t-tests. Panel A gives descriptive statistics for each measure of mathematics achievement using norm equivalent scores. Panel B gives the average change from first to fifth grade, the correlations of first and fifth grade scores, and results of paired t-tests that examine the null hypothesis that there was no change, on average, in achievement from one grade to the other. Panel C translates the averages in Panel A into percentile scores. The data in Panel C can be interpreted as the percentile score that the average student obtained in the two groups of schools in first grade and fifth grade. The data in Panel C were used to construct Figure 9 in the text.

Table A-15: Descriptive statistics and Paired t-test Results, 1st and 5th Grade Mathematics Achievement Norm Equivalent Scores, DI and Control Schools, Full and Reduced Samples, BCPSS, 1998-2003

A: Descriptive Statistics (Norm Equivalent Scores)

		DI Schools Full Sample		**Control Schools Full Sample**		**Control Schools Reduced Sample**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Computations	1st.	33.1	22.1	40.3	25.1	30.8	21.7
	5th	50.2	19.5	53.4	22.1	51.2	22.0
	% ch.	0.52		0.32		0.66	
Concepts and Applications	1st.	36.2	19.0	46.2	22.2	38.7	20.3
	5th	47.1	16.2	47.5	17.4	44.8	16.9
	% ch.	0.30		0.03		0.16	
Composite	1st.	32.8	20.2	42.3	22.9	33.3	19.9
	5th	48.5	17.8	50.6	20.3	47.9	19.9
	% ch.	0.48		0.20		0.44	

B: Average Differences, Correlations, paired (dependent sample) t-scores

	Ave. Dif.	r	t	p	N
DI Schools					
Computations	17.1	0.45	9.85	<.001	159
Concepts and Applications	10.9	0.46	7.44	<.001	159
Composite	15.7	0.54	10.58	<.001	155

Control Schools Full Sample					
Computations	13.1	0.38	33.26	<.001	4495
Concepts and Applications	1.3	0.53	4.59	<.001	4498
Composite	8.3	0.52	25.80	<.001	4437
Reduced Sample					
Control Schools Reduced Sample					
Computations	20.4	0.34	33.94	<.001	1755
Concepts and Applications	6.1	0.53	14.06	<.001	1756
Composite	14.6	0.49	3.02	<.001	1726

Results indicate that students in both the DI and the Control Schools had significantly higher achievement from in fifth grade than in first grade. For both groups and each measure the average change in norm equivalent scores from first to fifth grade was statistically significant. Changes on all three scores were larger for students in the DI group than for those in the full sample of Control Schools. Changes for students in the DI group were larger than those for students in the reduced sample of Control Schools on both the measure of Concepts and Applications and the Composite Score. This can be seen in the average differences reported in Panel B as well as the percentage gain reported in Panel A. The largest differences between students in the DI Schools and those in the Control Schools occurred with the measure of concepts and applications. On average students in the DI schools had concepts and application scores that were 30 percent higher in fifth grade than in first grade compared to an increase of only 3 percent for students in the total group of Control Schools and 16 percent for those in the reduced sample of Control Schools. These data were used to create Figures 9 and 10 in the text.

Independent t-tests

Table A-16 provides an alternative way to examine the data by looking at the difference in average achievement scores of students in the DI schools and Control schools in first grade and in fifth grade. While the computations reported in Table A-15 compared the average difference between students' first and fifth grade scores, the computations in Table A-16 compare average scores of students in DI and Control Schools in first grade and in fifth grade.

Panel A includes comparisons of students in the DI Schools with those in the full sample of Control Schools. Results indicate that scores of those in the Control Schools were significantly higher than those in the DI schools on all measures in first grade but only on the measure of computations in the fifth grade. Panel B includes comparisons with the reduced sample of Control Schools. None of the comparisons in Panel B are statistically significant, although the results for Concepts and Applications in fifth grade approaches significance ($p = .12$).

Table A-16: Independent t-tests, Average Mathematics Achievement, DI and Control Schools (Full and Reduced Samples), 1st and 5th Grade, BCPSS, 1998-2003

Panel A: DI Schools vs. Full Sample Control Schools

First Grade	t	df	p
Computations	3.94	177	<.001
Concepts & Apps.	6.46	174	<.001
Composite	5.625	172	<.001
Fifth Grade			
Computations	2.01	174	0.05
Concepts & Apps.	0.37	4756	0.71
Composite	1.42	174	0.16

Panel B: DI Schools vs. Reduced Sample Control Schools

First Grade	t	df	p
Computations	-1.44	1934	0.15
Concepts & Apps.	1.46	1920	0.15
Composite	0.09	186	0.93
Fifth Grade			
Computations	0.53	197	0.60
Concepts & Apps.	-1.56	1974	0.12
Composite	-0.42	1957	0.68

Note: F tests indicated that the t-value assuming unequal variances between the groups should be used for all comparisons with the total group of control schools but the one for concepts and applications for fifth grade. The t-value assuming unequal variances was used with comparisons with the reduced sample for the composite score for first grade and the computations score for fifth grade. Negative t values indicate that students in the DI schools had higher average scores than students in the Control schools. Positive values indicate that the Control school students had higher scores.

Repeated Measures ANOVA

Table A-17 gives the results of a repeated measures analysis of variance with the first and fifth grade scores as a repeated measure and the type of school (DI or Control) as a factor. This approach is more parsimonious than the separate t-tests reported in Tables A-15 and A-16. Panel A gives results with comparisons with the total set of Control Schools and Panel B gives results with the reduced sample of Control Schools. If exposure to Direct Instruction in the first grade is related to higher achievement in fifth grade, a significant interaction effect would be expected. With comparisons with the total sample of Control Schools (Panel A) the results are as expected with the measure of concepts and applications and the composite score. With comparisons with the reduced sample, the results are as expected only with the measure of concepts and applications.

Table A-17: Repeated Measures Analysis of Variance, Norm Equivalent Mathematics Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools

Panel A: DI Schools versus Full Set of Control Schools

	Computations	Concepts and Applications	Composite
F-Mathematics	203.16	60.02	189.09
P	<.001	<.001	<.001
F- Group	10.76	13.94	14.32
P	0.001	<.001	<.001
Interaction	3.59	36.56	17.87
P	0.06	<.001	<.001
Df	1,4652	1, 4655	1, 4590

Panel B: DI Schools versus Reduced Set of Control Schools

	Computations	Concepts and Applications	Composite
F-Mathematics	329.69	126.76	327.31
P	<.001	<.001	<.001
F- Group	0.20	0.004	0.003
P	0.66	0.95	0.95
Interaction	2.59	10.03	0.43
P	0.11	0.002	0.51
Df	1, 1912	1,1913	1,1879

Note: The F test associated with mathematics tests the null hypothesis that the change from first grade to fifth grade equals zero. The F test associated with group tests the null hypothesis that the average mathematics score of students in the DI schools. The F test associated with the interaction tests the null hypothesis that the differences between groups are the same for each grade or, alternatively, that the difference between the two grades is the same for each group.

Research Question Two: Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?

Tables A-18 and A-19 report the results of the mixed model regressions. Panel A of Table A-18 summarizes the models that were used and Panels B and C give summary statistics associated with each model for the full set of Control Schools (Panel B) and the reduced set (Panel C). The models become increasingly more complex. Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average mathematics achievement in fifth grade. The correlation ratio attached to Model 1 is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that between 9 and 12 percent of the variance in fifth grade achievement is between schools rather than simply between

students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model and school differences are controlled. These null hypotheses can be easily rejected with all three models. There is significant variation between schools and also between students even when the explanatory variables are included. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those in this analysis.

The models become incrementally more complex, with each subsequent model adding more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the norm equivalent reading achievement score for year one and the two factor scores measuring the advantage/disadvantage of the schools the students attended. It was expected that students' first grade achievement would be positively and significantly associated with their achievement in fifth grade. Including the two factor scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African American students and more non-Hispanic white students, would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient).

Model 3 adds the treatment group, testing the hypothesis that students in DI Schools have higher achievement than students in the Control Schools when first grade achievement and school advantage/disadvantage are controlled. If having Direct Instruction in first grade influences students' fifth grade achievement, independent of their first grade achievement and school context, we would expect positive associations with this variable. It should be recalled that students were only coded as having Direct Instruction if they were in a school in a year when the school was in a treatment condition.

Table A-18, Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Mathematics Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3
Random intercept for Schools	x	x	x
First Grade Achievement		x	x
Factors 1 and 2		x	x
DI in First Grade			x

Panel B: Model Fit Statistics Comparing DI Schools and Full Sample of Control Schools

		Model 1			Model 2			Model 3	
Computations	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects Estimate	54.8	9.0	<.0001	51.4	8.4	<.0001	51.4	8.4	<.0001
Residual Estimate	434.6	9.0	<.0001	363.3	7.6	<.0001	363.2	7.6	<.0001
BIC	42452			40873			40877		
- 2 Log Likelihood	42438			40845			40845		
Change in LL				1593			0.0		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.11								
Concepts and Applications									
Random Effects Estimate	35.1	5.8	<.0001	34.5	5.5	<.0001	34.7	5.5	<.0001
Residual Estimate	267.7	5.6	<.0001	181.3	3.8	<.0001	181.0	3.8	<.0001
BIC	40303			37690			37686		
- 2 Log Likelihood	40297			37662			37653		
Change in LL				2634			9.0		
df				3			1		
p				<.001			<.01		
Correlation ratio	0.12								
Composite									
Random Effects Estimate	49.8	8.1	<.0001	53.2	8.3	<.0001	53.2	8.3	<.0001
Residual Estimate	361.6	7.5	<.0001	247.3	5.2	<.0001	247.2	5.2	<.0001
BIC	41535			38603			38606		
- 2 Log Likelihood	41522			38575			38573		
Change in LL				2947			1.6		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.12								

Panel C: Model Fit Statistics Comparing DI Schools and Reduced Sample of Control Schools

		Model 1			Model 2			Model 3	
Computations	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects Estimate	53.2	12.8	<.0001	57.4	13.5	<.0001	57.4	13.5	<.0001
Residual Estimate	422.0	13.6	<.0001	361.6	11.8	<.0001	361.6	11.8	<.0001
BIC	17527			16825			16829		
- 2 Log Likelihood	17515			16801			16801		
Change in LL				714			0		
df				3			1		
p				<.001			n.s.		

Correlation ratio	0.11								
Concepts and Applications									
Random Effects Estimate	26.4	6.7	<.0001	36.7	8.3	<.0001	38.3	8.7	<.0001
Residual Estimate	262.5	8.5	<.0001	168.8	5.5	<.0001	167.9	5.5	<.0001
BIC	16704			15389			15384		
- 2 Log Likelihood	16692			15365			15357		
Change in LL				1327			8.5		
df				3			1		
p				<.001			<.01		
Correlation ratio	0.09								
Composite									
Random Effects Estimate	42.7	10.4	<.0001	57.9	13.0	<.0001	58.6	13.2	<.0001
Residual Estimate	350.4	11.3	<.0001	237.6	7.9	<.0001	237.4	7.8	<.0001
BIC	17134			15764			15767		
- 2 Log Likelihood	17123			15741			15740		
Change in LL				1382			1.1		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.11								

The -2 log likelihood measures and the BIC values in Panels B and C of Table A-18 can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the comparisons between these values are also included. For both sets of comparisons the results indicate that Model 2 is the best fitting model for the analyses of computations and the composite scores, but Model 3 is the best fitting for the analysis of concepts and applications.

Table A-19 gives the coefficients associated with variables in Models 2 and Model 3 for each of the measures of achievement and for both comparisons. The coefficients associated with first grade achievement are positive and highly significant in all models for all three dependent variables, indicating that, as expected, students with higher mathematics achievement scores in first grade also have higher mathematics achievement scores in fifth grade. As expected, the coefficients associated with factor 1 are positive and statistically significant with all analyses for the total group of Control Schools, indicating that students have higher achievement scores in schools with more non-Hispanic white children and fewer African American children and fewer children receiving free or reduced lunch. Associations with Factor 2 are insignificant in all analyses. Coefficients associated with Factor 1 are not significant for the analyses in Panel B, perhaps because the variation in the factor is somewhat smaller in this group.

Table A-19: Coefficients Associated with Mixed Model Results, Fifth Grade Mathematics Achievement Regressed on First Grade Achievement, School Advantage/Disadvantage, and Attending a DI School

Panel A: Coefficients for Comparisons of DI Schools and Total Sample of Control Schools

	Model 2			Model 3		
Computations	B	s.e.	p.	b	s.e.	p.
Intercept	38.5	0.9	<.0001	38.5	0.9	<.0001
First grade score	0.4	0.01	<.0001	0.4	0.01	<.0001
Factor 1	1.4	0.7	0.055	1.4	0.7	0.06
Factor 2	-0.2	0.7	0.795	-0.3	0.8	0.76
Had DI in First Grade	------	------	------	0.5	3.2	0.87
Concepts and Applications						
Intercept	27.1	0.7	<.0001	26.9	0.8	<.0001
First grade score	0.4	0.01	<.0001	0.44	0.01	<.0001
Factor 1	1.7	0.6	0.004	1.7	0.6	0.004
Factor 2	0.1	0.6	0.90	-0.7	0.6	0.25
Had DI in First Grade	------	------	------	7.0	2.3	0.003
Composite						
Intercept	29.5	0.9	<.0001	29.4	0.9	<.0001
First grade score	0.5	0.01	<.0001	0.5	0.01	<.0001
Factor 1	1.5	0.7	0.03	1.5	0.7	0.03
Factor 2	0.1	0.7	0.89	-0.31	0.8	0.70
Had DI in First Grade	------	------	------	3.6	2.8	0.20

Panel B: Coefficients for Comparisons of DI Schools and Reduced Sample of Control Schools

	Model 2			Model 3		
Computations	b	s.e.	p.	b	s.e.	p.
Intercept	39.4	1.4	<.0001	39.4	1.4	<.0001
First grade score	0.4	0.02	<.0001	0.4	0.02	<.0001
Factor 1	-0.04	1.3	0.97	-0.04	1.3	0.97
Factor 2	-0.2	0.8	0.79	-0.2	0.9	0.80
Had DI in First Grade	------	------	------	0.1	3.3	0.97
Concepts and Applications						
Intercept	26.5	1.1	<.0001	26.1	1.1	<.0001
First grade score	0.5	0.02	<.0001	0.47	0.02	<.0001
Factor 1	1.0	1.0	0.32	0.9	1.0	0.35
Factor 2	0.3	0.7	0.69	-0.6	0.7	0.43
Had DI in First Grade	------	------	------	6.8	2.3	0.004
Composite						
Intercept	29.9	1.3	<.0001	29.7	1.3	<.0001
First grade score	0.5	0.02	<.0001	0.5	0.02	<.0001
Factor 1	0.4	1.2	0.76	0.4	1.2	0.77
Factor 2	0.1	0.8	0.91	-0.3	0.9	0.75
Had DI in First Grade	------	------	------	3.0	2.8	0.29

The coefficients associated with attending a DI school in first grade are positive in all analyses, indicating that students in the DI schools have higher fifth grade achievement when other variables are equalized. However, the coefficients are significant only in the analyses of the measure of concepts and applications. These results appear both in the comparison with the total group of Control schools and with the reduced sample. As expected, students who attended a DI school in first grade had significantly higher scores in the measure of concepts and applications than other children, even when the socio-demographic context of the school and their first grade achievement scores were controlled.

Table A-20 summarizes the impact of these effects assuming that students had the same first grade reading score and that all schools were equivalent in their socio-economic characteristics. Average scores for the total group were substituted into the regression equations for Model 3 in Table A-19 for these calculations.[13] Results are given both in the norm equivalent

[13] The average value for the total group for computations was 39.97; for concepts and applications, 45.81; and for the composite score, 41.9. The average value for factor 1 was .1495 and for factor 2 was -.0497. These values do not equal zero for two reasons. The unit of analysis for these computations is students, rather than schools; and the population of students in the panel sample, as described above, is more advantaged than the general population and thus comes from schools with higher scores on Factor 1 and lower scores on Factor 2 than did the original sample of students in first grade.

scores and as the corresponding percentile score for a student scoring at this average. It can be seen, with both sets of comparisons, that, if they were to have equal scores at first grade and to attend schools with similar socio-demographic characteristics, students in the Direct Instruction schools and the Control Schools would be expected to have almost identical computation scores. However, students in the Direct Instruction school would have scores on the composite and concepts and applications measure that were higher than those of students in the Control Schools. The difference in the concepts and applications measure is the largest. With calculations based on the entire population of students, an average student in a DI school would score at the 58th percentile on the measure of concepts and applications and the composite score, well above the national average, while an average student in the Control Schools would only score at the 44th percentile, well below the national average. If calculations are based on the analysis that used the reduced sample of Control schools, the average student in the DI Schools would score at the 51st percentile on the measure of concepts and applications and at the 50th percentile on the composite measure, while the average student in the control schools would be at the 39th and 44th percentile on these measures.

Table A-20: Simulated Fifth Grade Mathematics Achievement Scores (Norm Equivalent and Percentile of Hypothetical Average Student), DI and Control Schools, Controlling for School Context and First Grade Achievement, BCPSS, 2001-03

	Full Sample of Control Schools		Reduced Sample of Control Schools	
A. Norm Equivalent Scores	DI Schools	Control Schools	DI Schools	Control Schools
Computations	53.5	53.0	50.7	50.5
Concepts and Applications	54.1	47.1	50.7	43.9
Composite	53.8	50.2	50.0	47.0
B. Corresponding Percentiles				
Computations	56.6	55.7	51.3	51.0
Concepts and Applications	57.8	44.5	51.3	38.6
Composite	57.2	50.4	50.0	44.3

Note: Scores were calculated by substituting the average values of first grade achievement and factor scores in the equations for Model 3 in Table A-11. For the sample that includes all the Control Schools the mean of Factor 1 was 1.4 and the mean for Factor 2 was -.3. For the sample with the reduced set of Control Schools the mean of Factor 1 was -.08 and the mean for Factor 2 was .23.

Research Question Three. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

Because the measures of first and fifth grade achievement are statistically dependent, Cohen's d, the most often used measure of effect size, is not appropriate. To control for the dependence of these measures, effect sizes were calculated adjusting for the correlation between scores in the two years (Dunlap et al 1996). It was anticipated that effect sizes would be larger for the students who received Direct Instruction in the first grade than for other students. The effect sizes were calculated using the descriptive statistics from the norm equivalent scores. McLean and associates, 2000, have demonstrated that effect sizes for NCE scores are lower than those that would be obtained for equivalent raw and scaled scores. Thus, the calculations presented here should be seen as conservative estimates.

Table A-21 summarizes effect scores for the change in achievement from first grade to fifth grade on each of the measures. Using the criterion of .25 as an educationally significant effect size, it may be seen that all of the effect sizes associated with the DI schools clearly surpass that criterion as do several for the other groups. Both sets of Control Schools had educationally significant changes on the measure of computations, as did the full sample of Control Schools on the composite measure and the reduced set of Control Schools on the concepts and applications measure. The largest differences between the groups occur with the measure of concepts and applications and the composite measure, where the effect sizes are substantially larger for the DI schools. The very small effect sizes for the concepts and applications measure for the full set of control schools and for the composite measure for the reduced sample of control schools reflect the much lower t-values for those difference scores.

Table A-21: Effect Sizes of Change, First to Fifth Grade, by Group, Mathematics Achievement Norm Equivalent Scores

	Computations	Concepts and Applications	Composite
DI Schools	0.82	0.61	0.82
Control Schools Full Sample	0.55	0.07	0.38
Control Schools Redeuced Sample	0.93	0.32	0.07

Note: All effect sizes are based on norm equivalent scores.

References

Fashola, Olatokunbo S. and Robert E. Slavin. 1997. Promising programs for elementary and middle schools: Evidence of effectiveness and replicability. *Journal of Education for Students Placed at Risk* 2: 251-307.

MacIver and Kempner. 2002. The impact of Direct Instruction on elementary students' reading achievement in an urban school district. *Journal of Education for Students Placed at Risk* 7: 197-220.

McLean, James E., Marcia R. O'Neal, and J. Jackson Barnette. 2000. Are All Effect Sizes Created Equal? Paper presented at the 28th Annual Meeting of the Mid-South Educational Research Association, November, 2000, ED 448 188.

Stockard, Jean and Maralee Mayberry. 1992. *Effective Educational Environments.* Newbury Park, California: Corwin/Sage, 1992.

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

September 2008

Technical Report 2008-03



© 2008 National Institute for Direct Instruction
PO Box 11248 • Eugene, OR 97440
1-877-485-1973

Table of Contents

	Page
Executive Summary	iv
List of Tables	v
List of Figures	vii
Report	1
A: First Grade Mathematics Achievement	1
B. The Relationship of First Grade Direct Instruction and Fifth Grade Mathematics Achievement	8
C. Summary and Discussion	13
Appendix I: Sample and Measures – First Grade Analysis Sample Measuring Achievement Measuring School Context	15
Appendix II: Detailed Statistical Results – First Grade Analysis Methodology Research Question One: Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools? Research Question Two: Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized? Research Question Three: Did the impact of being in a DI School increase over time as reforms became institutionalized within a school? Research Question Four: What is the Magnitude of the Effect of Direct Instruction?	20
Appendix III: Panel Sample: Changes from First to Fifth Grade	33
Appendix IV: Detailed Statistical Results: Fifth Grade Achievement Analysis Methodology Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools? Research Question Two: Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the	37

socio-demographic context of the school and first grade achievement are controlled?
Research Question Three. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

References 50

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Executive Summary

In the late 1990s, in response to very low achievement scores, the Baltimore City Public School System (BCPSS) implemented curricular reforms. Eleven schools in the district introduced Direct Instruction as part of a whole school reform under the guidance of the National Institute for Direct Instruction. They began by implementing the reading curriculum and, one to two years later, added mathematics Direct Instruction, using *Distar Arithmetic* and *Connecting Math Concepts*. The Comprehensive Test of Basic Skills (CTBS) was administered to all first graders and all fifth graders in BCPSS in the spring of 6 school years: 1997-1998 through 2002-2003. Data were obtained from over 40,000 first grade students on mathematics achievement in computations, concepts and applications, and a composite score. Of the students who were in first grade in 1997-98 and 1998-99, more than 4,500 remained in the system and in the same schools in 2001-02 and 2003-03.

This study compares students' mathematics achievement in schools that implemented Direct Instruction and (DI schools), and schools without Direct Instruction (the Control schools). Key findings are listed below.

- Combining data across all years (1998-2003) first grade students in schools that implemented Direct Instruction had significantly higher mathematics achievement than students in other schools, and these differences became even stronger when demographic characteristics of the schools were controlled.
- The impact of being in a DI school became greater over time as the curricular changes became institutionalized within the schools. At the beginning of the study period average scores in the DI schools were significantly lower than those in the Control Schools, but by the end of the period they were significantly higher.
- First grade students in both groups of schools had higher achievement scores in 2003 than in 1998, but the increase was significantly larger for students in NIFDI-supported schools than for students in the other schools. On average, first grade composite mathematics achievement scores in the DI schools increased by 135 percent from 1998 to 2003, but by only 54 percent in the Control Schools.
- The magnitude of the effect on first grade achievement from attending a school that implemented DI mathematics with NIFDI support was statistically significant and substantively large.
- Among the students who remained in the same school from first grade to fifth grade, those in the DI schools had significantly lower achievement in first grade than those in the other schools. By fifth grade, however, these differences had disappeared.

List of Tables

Table Title	Page
A-1: Schools in Treatment Group and Start Dates of Reading and Mathematics Direct Instruction	16
A-2: Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels by Group	17
A-3 Factor Analysis of School Characteristics, Elementary Schools, BCPSS, 1998-2003	18
A-4: Average Mathematics Achievement Scores of Students With and Without Direct Instruction, All Years, First Graders, BCPSS	21
A-5: Variables in Models and Measures of Fit, Analysis of First Grade Mathematics Achievement, BCPSS, 1998-2003	23
A-6: Coefficients Associated with Mixed Model Regressions of First Grade Mathematics Achievement, BCPSS, 1998-2003	26
A-7: Predicted Achievement Scores Assuming Equal School Contexts	27
A-8: First Grade Mathematics Achievement Scores by Group and Year, Adjusting for School Advantage, BCPSS	28
A-9: Mathematics Achievement Scores by Year and Group, First Graders, BCPSS, 1998-2003	30
A-10: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Mathematics Achievement, BCPSS, 1998-2003	32
A-11: Turnover in the BCPSS, 1997-98, 1998-99 to 2001-02, 2002-03	33
A-12: Schools in Treatment Group by Year for Panel Analysis	34
A-13: Persistence in the Same School, by School Group	35
A-14: Comparison of Achievement Scores of Panel Group and Other First Graders, DI and Control Schools	35
A-15: Descriptive statistics and Paired t-test Results, 1st and 5th Grade Mathematics Achievement Norm Equivalent Scores, DI and Control Schools, Full and Reduced Samples, BCPSS, 1998-2003	38

A-16: Independent t-tests, Average Mathematics Achievement, DI and Control Schools (Full and Reduced Samples), 1st and 5th Grade, BCPSS, 1998-2003 40

A-17: Repeated Measures Analysis of Variance, Norm Equivalent Mathematics Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools 41

A-18, Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Mathematics Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction 42

A-19: Coefficients Associated with Mixed Model Results, Fifth Grade Mathematics Achievement Regressed on First Grade Achievement, School Advantage/Disadvantage, and Attending a DI School 45

A-20: Simulated Fifth Grade Mathematics Achievement Scores (Norm Equivalent and Percentile of Hypothetical Average Student), DI and Control Schools, Controlling for School Context and First Grade Achievement, BCPSS, 2001-03 47

A-21: Effect Sizes of Change, First to Fifth Grade, by Group, Mathematics Achievement Norm Equivalent Scores 49

List of Figures

Figure Title	Page
1: Percentile Scores, Mathematics, Average Student, DI and Control Schools	3
2: Percentile Scores, Average Student by Group, Adjusting for School SES	4
3: Computation Achievement, Average Student, BCPSS, 1998-2003 by Group	5
4: Concepts and Applications Achievement, Average Student, BCPSS, 1998-2003 by Group	5
5: Composite Achievement, Average Student, BCPSS, 1998-2003 by Group	5
6: Percentage Increase in Mathematics Achievement, 1998-2003, by Group	6
7: Effect Sizes, Mathematics Achievement, All Years and 2003	7
8: Effect Sizes, Mathematics Achievement, Adjusting for School Context, All Years and 2003	8
9: Computation, Concepts and Applications, and Composite Mathematics Percentile Achievement Scores of an Average Student, First and Fifth Grade, DI and Control Schools, BCPSS, 1998-2003	10
10: Percentage Gain in Achievement Scores, First to Fifth Grade, by Group	10
11: Mathematics Percentile Scores of an Average Fifth Grade Student, Adjusting for School Context and First Grade Achievement, DI and Full Set of Control Schools	11
12: Mathematics Percentile Scores of an Average Fifth Grade Student, Adjusting for School Context and First Grade Achievement, DI and Reduced Set of Control Schools	12
13: Effect Sizes, Change from First to Fifth Grade in Mathematics Achievement Scores, DI and Control Schools	12

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in Baltimore City Public Schools

Low achievement in mathematics has been a major concern of school districts throughout the country. The Baltimore City Public School System (BCPSS) is similar to many other large city school districts that serve students with high levels of poverty and struggle with low levels of achievement. In the late 1990s, curricular reforms were implemented in the BCPSS elementary schools to address this low achievement. Eleven schools in the district introduced Direct Instruction as part of a whole school reform under the guidance of the National Institute for Direct Instruction (NIFDI).[1] The schools began with the reading curriculum and, after one to two years, introduced the mathematics curriculum. Two mathematics curricula were used, depending upon the students' level and needs: *Distar Arithmetic* and *Connecting Math Concepts*. Other schools in the district used a variety of curricular materials.

A large body of literature, including well designed meta-analyses, has demonstrated the effectiveness of Direct Instruction in promoting high achievement, although somewhat fewer studies address the mathematics curricula with data from an extended time period and from a relatively large number of schools. This paper addresses this area by examining data on the mathematics achievement of first and fifth graders in the BCPSS from 1998 through 2003. The first part of this report focuses on the mathematics achievement of first graders in the BCPSS from 1998 through 2003, comparing the achievement of students in schools that used the Direct Instruction to the achievement of other students in the district. The second part focuses on the subset of students who were in the same schools in the BCPSS in first and fifth grade and examines the extent to which experiencing the Direct Instruction curriculum in first grade is associated with higher achievement in fifth grade. In this report results are summarized in graphs and are also expressed as effect sizes. Appendices provide full statistical details.

A. First Grade Mathematics Achievement

The analysis of first grade mathematics achievement uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Achievement measures were obtained each spring using the Comprehensive Test of Basic Skills (CTBS), a widely used, nationally normed standardized test. Subtest measures of "Computations" and "Concepts and Applications," as well as a Composite measure were available. Results in this part of the report are presented as percentiles. Percentiles can be simply interpreted as the percentage of students that would have scores lower than a given student. For instance, if a student has a score at the 60th percentile, 60 percent of all students had scores that were lower.

[1] One additional school was included in the original group that implemented Direct Instruction. However, it closed very soon after the implementation began and thus it is not included in this count or in any of the analyses. In addition, a few other schools in the system used Direct Instruction reading programs. They, however, did not use DI mathematics curricula.

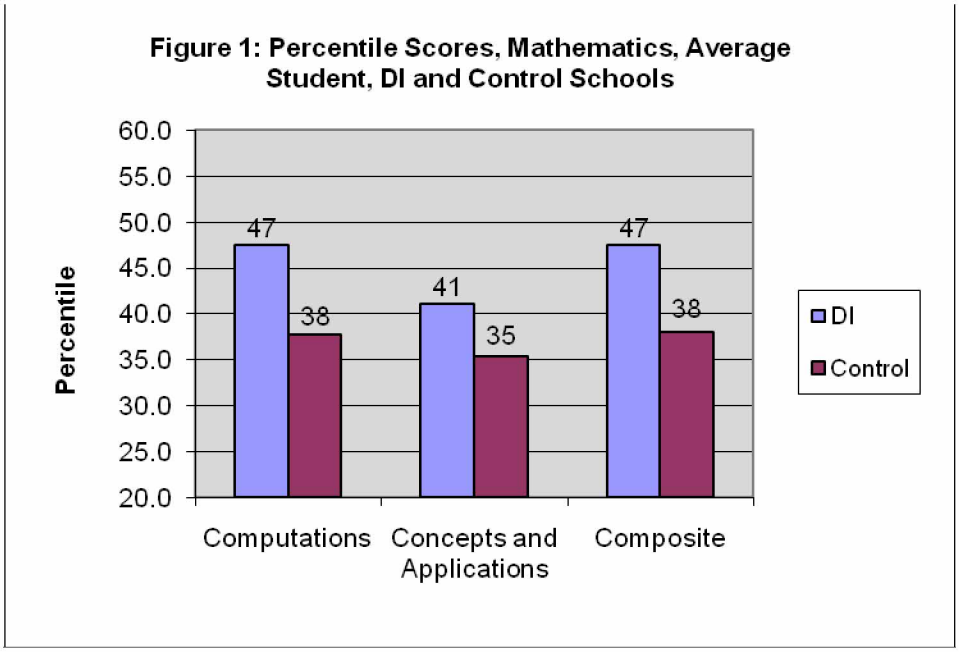
The average school in the BCPSS had large proportions of low income and minority students. On average, 75 percent of the students in a school were on free or reduced lunch. This varied however, from a low of 22 percent to a high of 93 percent. On average 84 percent of the students in average elementary school were African American, although the percentage varied from a minimum of 3 to almost 100 percent. Reflecting this high level of segregation, schools also varied in their representation of non-Hispanic white students, from having virtually no white students to being 94 percent white, with the average school having 14 percent of its students reported as being non-Hispanic white. Asian, Hispanic and Native American students were quite rare in the district, with an average representation of 1 percent in the student bodies. The DI Schools were similar to the Control Schools in racial-ethnic composition. However, the DI schools had proportionately more students who received free or reduced lunch: an average of 83 percent in the DI Schools compared to an average of 74 percent in the Control Schools.

The analysis focused on comparing the achievement of students who received Direct Instruction with those who did not. Full details on the statistical techniques and results are given in Appendix II. Four research questions were addressed:

- Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?
- Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?
- Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?
- What is the magnitude of the effect of have Direct Instruction on mathematics achievement?

1. Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?

Figure 1 summarizes achievement data for students who received Direct Instruction and those who did not on the three measures of mathematics achievement. These results combine the data from all the available years: 1997-98 through 2002-03 and represent the percentile score of the average student in a school. Combining data from all the years in the study, students who received Direct Instruction had significantly higher achievement scores than students in the Control Schools. The average first grader in a DI school scored at the 47th percentile, slightly below the national norm, in both the measure of computations and the composite score, but somewhat lower, at the 41st percentile, on the measure of concepts and applications. The average student in the Control schools had scores that were substantially lower, between the 36th and 38th percentile. These differences were statistically significant.

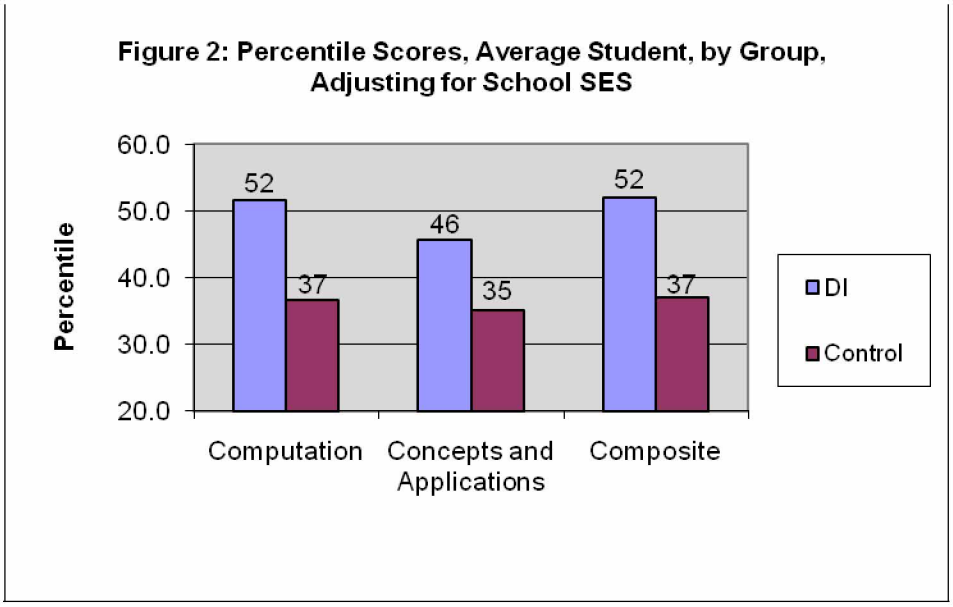


Note: Scores in this graph represent the percentile score of an average student in each type of school, averaging data over all years in the analysis: 1998-2003.

2. Did students in Direct Instruction Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?

Many studies have demonstrated that the socio-economic characteristics of a school have a strong influence on achievement, and thus it is important to control for this factor. As noted above, elementary schools in the Baltimore City Public School System varied substantially in their racial-ethnic composition and poverty rate, and, as in other studies students in BCPSS schools with fewer children from disadvantaged backgrounds had substantially higher average levels of achievement.

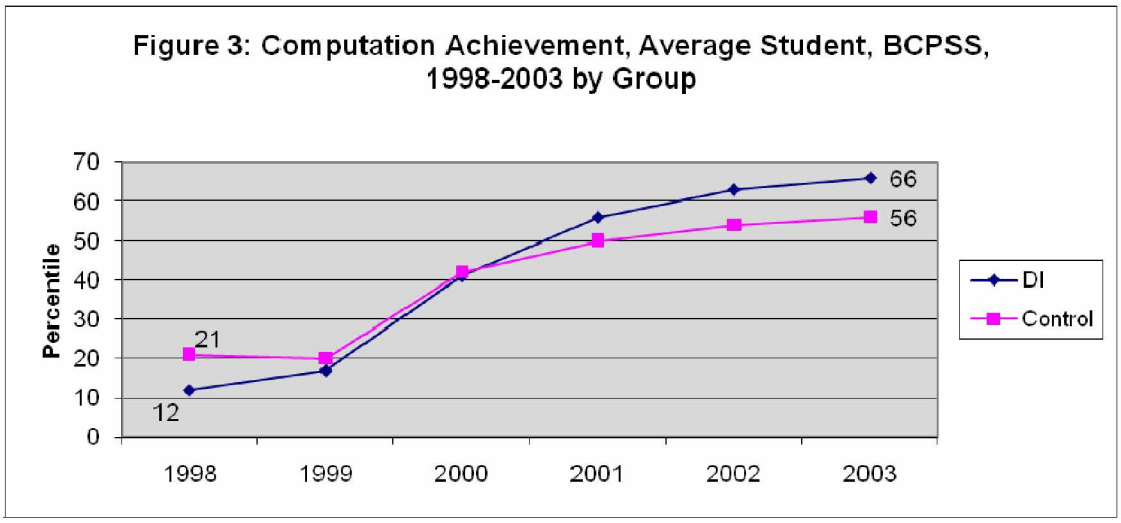
When the socio-economic characteristics of the schools were statistically controlled the differences in mathematics achievement between students in the DI schools and those in other schools became more marked. Figure 2 summarizes these results. The numbers in this graph represent the percentile score for an average student in each group if schools were equal in their racial-ethnic and poverty status. (Details regarding these calculations are in Appendix II.) After adjusting for the socioeconomic characteristics of the schools, the average student in a Direct Instruction School scored at the 52nd percentile on both the measures of computations and the composite score and at the 46th percentile on the measure of concepts and applications. In contrast, the average student at a Control School scored between the 35th and 37th percentile.

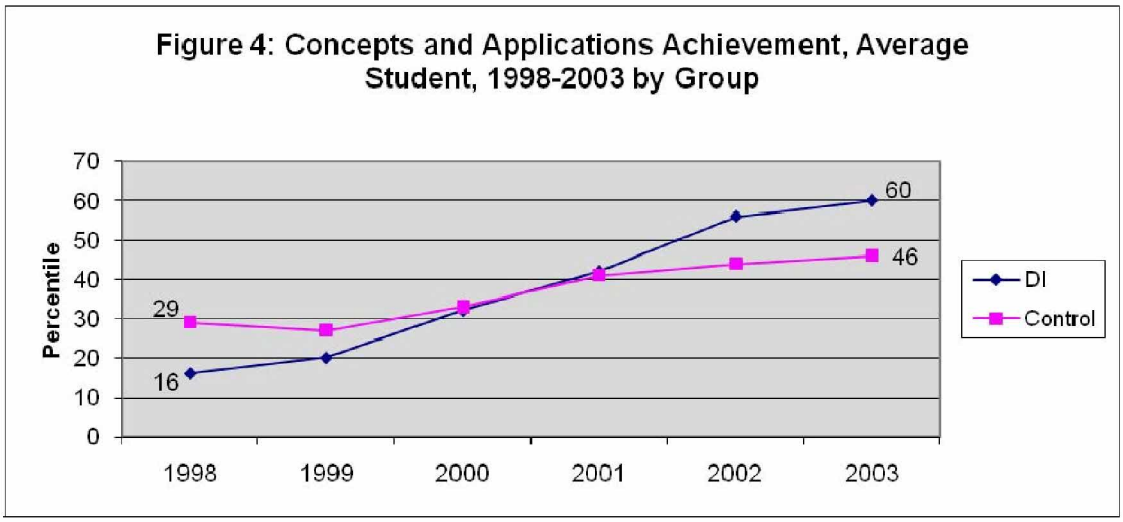


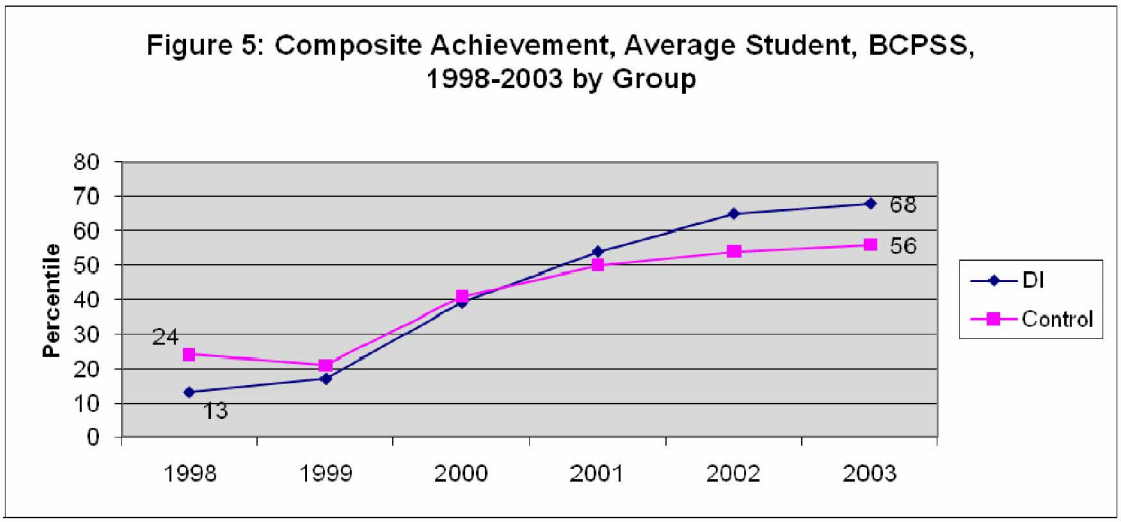
Note: Scores in this graph represent the percentile score of an average student in each type of school if the school contexts were equal in their socio-economic and demographic characteristics, averaging over all the years in the analysis: 1998-2003.

3. Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?

Changing school practices takes time and effort. It is reasonable to expect that the advantage of an effective instructional program would become more apparent as schools have more experience with a new curriculum, teachers gain more practice, and the procedures become part of the institutionalized and accepted practices within a school. During the time period under study the BCPSS paid concerted attention throughout the district to enhancing student achievement. The data indicate that these efforts were fruitful. Across the BCPSS first graders' mathematics achievement was significantly higher in 2003 than 1998. However, the changes over time were significantly larger for the students in the DI Schools than for those in the Control schools. Figures 3, 4, and 5 illustrate these changes. They show the percentile at which an average student in each group scored in each of the 6 years on each measure. Unlike the data in Figure 2, no adjustment was made for the school context of disadvantage.



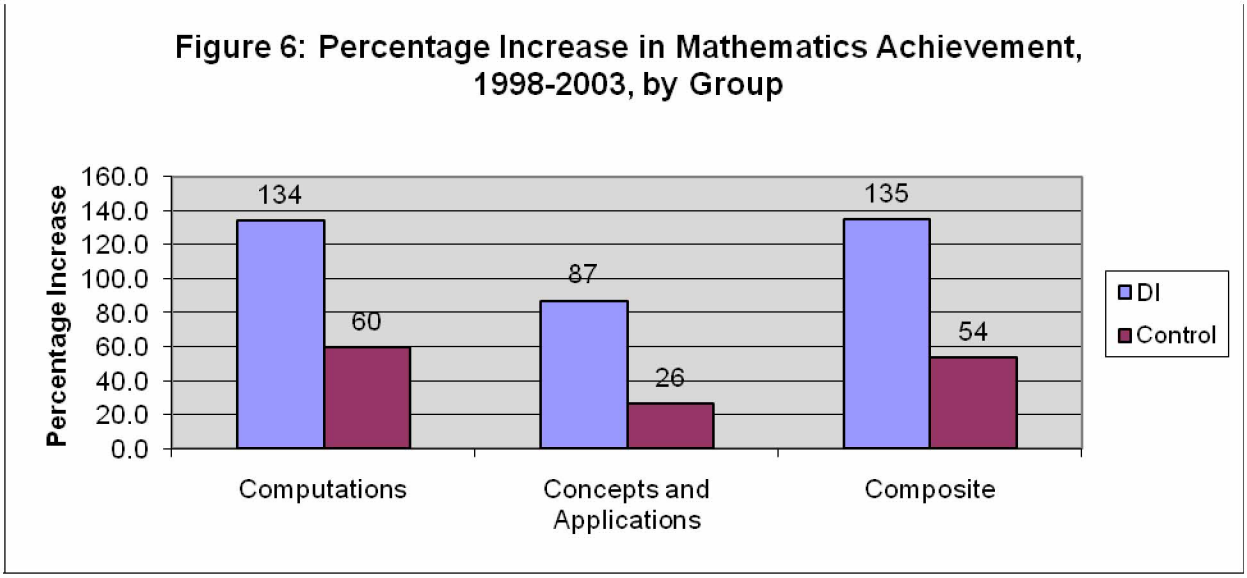




Note: Scores in Figures 3, 4, and 5 represent the percentile score of an average first grade student in each type of school for each year in the analysis.

At the beginning of implementation, in 1998, students in the DI Schools had significantly lower scores, on average, than students in the Control Schools. An average student in a DI school scored from the 12th (the computation and composite scores) to the 16th percentile (the concepts and applications score), while an average student in the Control Schools had scores ranging from the 21st to the 29th percentile. By 2003 the situation had changed markedly, with students in the DI schools having significantly higher scores than students in the Control Schools on all measures. By 2003, an average student in the DI Schools had scores were well above the national average, ranging from the 60th percentile on the measure of concepts and applications to the 68th percentile on the composite measure. In contrast, an average student in the Control Schools had scores ranging from the 46th percentile (on the measure of concepts and applications) to the 56th percentile.

Figure 6 illustrates the magnitude of these changes. Values in Figure 6 represent the difference between the 1998 and 2003 achievement scores as a percentage of the 1998 score, thus showing the percentage increase over time. For each measure, the changes over time were substantially larger for students in the DI schools than for students in the Control Schools. For instance, for the composite scores, first graders in DI schools had composite mathematics achievement scores in 2003 that were 135 percent higher than the scores of first graders in the same schools in 1998. In contrast, the scores of first graders in the Control Schools improved by only 54 percent. Similar differences appear with the other measures.



Note: The figures in this graph represent the percentage change in achievement from 1998 to 2003 for each of the groups of schools and each measure. These change figures were calculated using normal curve equivalent (NCE) scores rather than percentiles. As explained more fully in Appendix I, NCE scores are preferable for statistical procedures and were used for all calculations in the analysis. NCE scores were translated to percentiles for Figures 1-5.

4. What is the Magnitude of the Effect of NIFDI Supported Direct Instruction?

Unless specifically noted, all of the results presented in this report are statistically significant. That is, various statistical tests indicate that the results do not appear by chance. However, it is possible that statistical significance can simply reflect sample size. When data from large samples are subjected to statistical tests relatively small substantive differences can produce significant results. To combat this problem, researchers use measures of effect size. These measures provide descriptions of the magnitude of the results. One of the most commonly used measures is Cohen's *d*. It is simply calculated as the difference between two average values divided by the common standard deviation. Thus, it reports the magnitude of a difference between two groups as a proportion of the standard deviation. A d-value, or effect size, of .25 or larger has traditionally been considered educationally significant, indicating that an intervention has an important impact on students' achievement.

Figures 7 and 8 summarize the magnitude of the various effects reported earlier. Figure 7 summarizes the effects without any adjustment for school characteristics, using the data summarized in Figure 1. When data from all the years are examined together, the effect sizes comparing achievement of students in the DI Schools and those in the Control Schools are slightly lower than the .25 criterion (.13 for the measure of concepts and applications to .21 for the computations and composite scores). However, by 2003 the effect sizes were substantially larger, meeting or exceeding the .25 criterion: .25 for the measure of computations and.31 and .32 for the composite and concepts and applications measures.

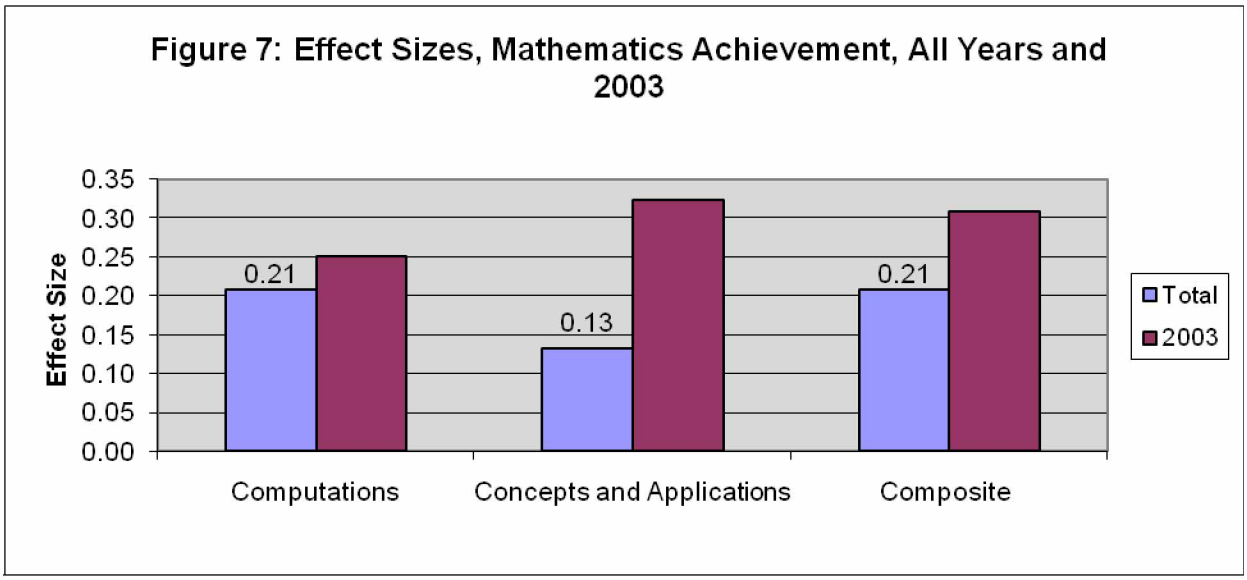
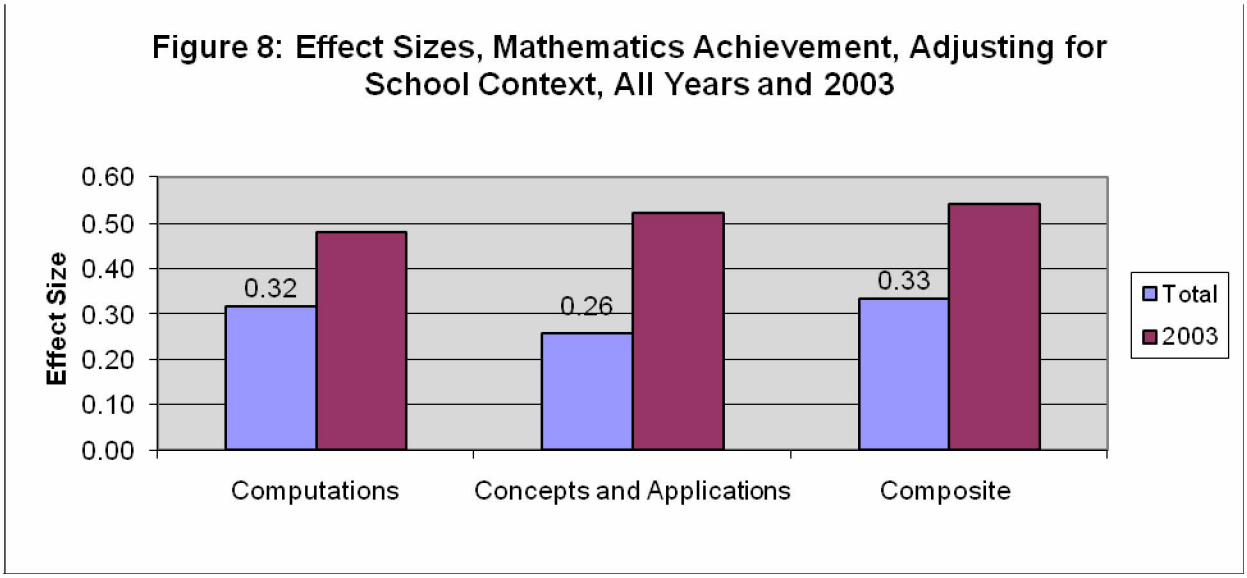


Figure 8 gives the effect sizes for the measure for all years and 2003 after the impact of school socio-economic status has been controlled (the data shown in Figure 2). Averaged across all years, the effect of being in a DI School relative to a Control School ranged from .26 for the measure of concepts and applications to 32 and .33 for computations and the composite score. At the end of the study period, the effect sizes were substantially larger: ranging from .48 to .54.



5. Summary

This analysis of the mathematics achievement of first grade students in the Baltimore City Public School System indicates that students in schools that implemented Direct Instruction had significantly higher achievement scores on all three indicators (computations, concepts and applications, and a composite score) than students in the Control Schools. These differences became even larger when the socio-demographic context of the schools was controlled and after several years of implementation of Direct Instruction. The effects were not just statistically significant, but substantively large and educationally important.

B. The Relationship of First Grade Direct Instruction and Fifth Grade Mathematics Achievement

Data were available for first and fifth graders in 1997-98 through 2002-03. While the first part of this report examined changes in first grade achievement over this time period, this section examines changes in achievement of individual students who began in the BCPSS in 1997-98 or 1998-99 and were still in the same schools five years later, in 2001-2 or 2002-3. As in the first part of the report, comparisons are made between students in the DI Schools and the Control Schools. (See Appendix III for full details on the sample used in the analysis.)

Two factors limited the size of the sample for this analysis: a high rate of turnover of students in the BCPSS and the fact that only six schools within the DI group had implemented Direct Instruction in mathematics by 1998-99. To provide additional insight into the results, two different control groups were used in the analysis: One includes all students in the Control Schools who started first grade in 1997-98 or 1998-99 and were in fifth grade in the same schools in 2001-02 or 2002-03. The second includes only students from schools with the lowest levels of achievement in the BCPSS in first grade in 1997-98 and 1998-99.[2] The DI group included 164 students, the total Control Group included 4,607 students, and the reduced Control Group included 1,820 students.

[2] Students who were retained, or held back a grade, were excluded from the analysis.

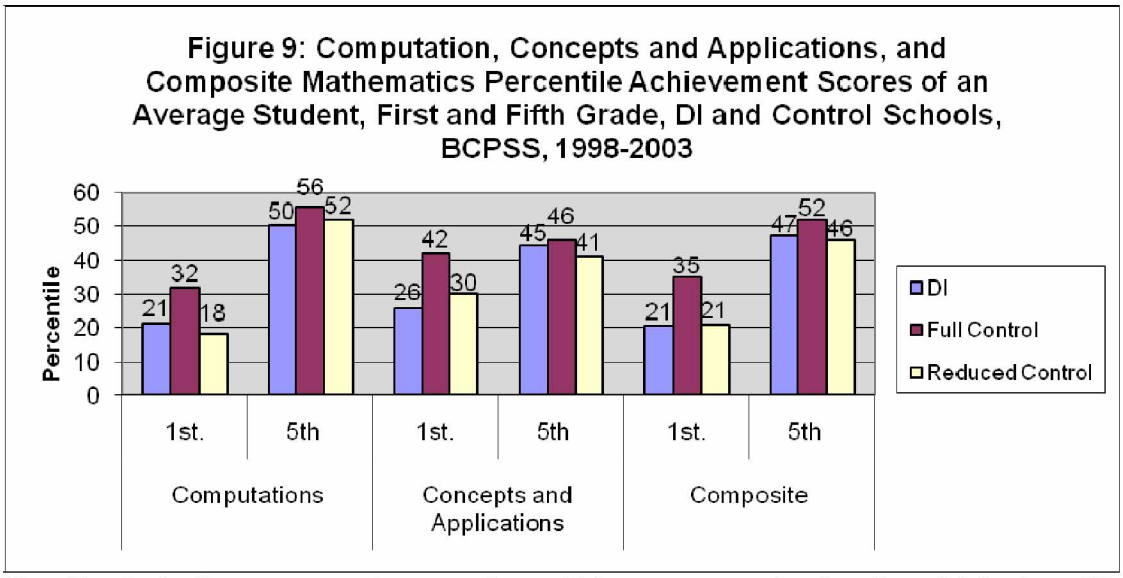
Three research questions were addressed:

- Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools?
- Did students who received Direct Instruction in first grade have higher mathematics achievement in fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?
- What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

1. Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools?

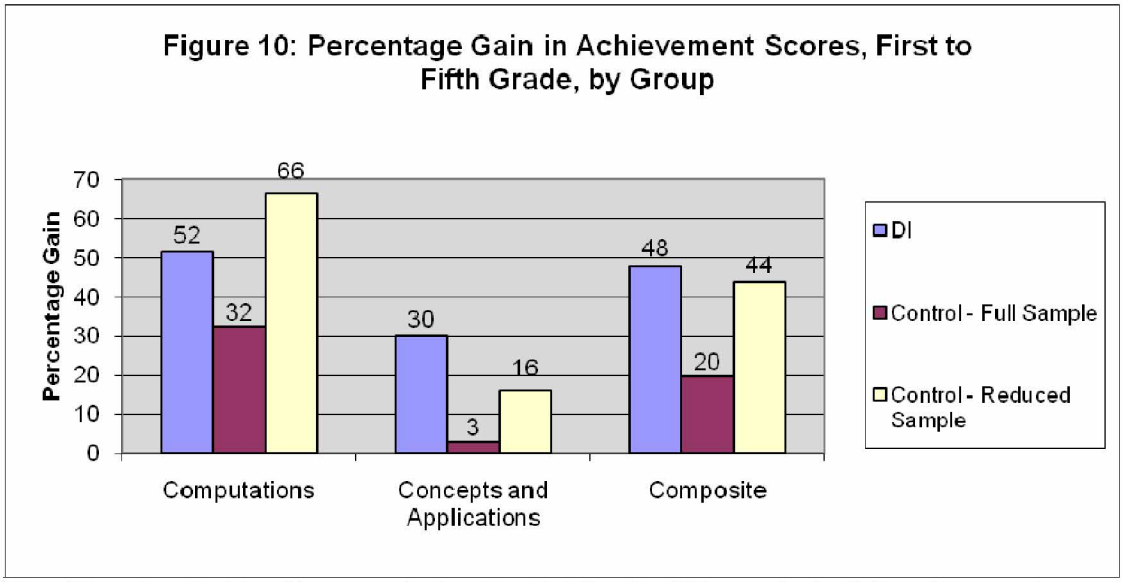
As explained fully in Appendix IV, several different statistical methods were used to examine differences in the changes in achievement from first to fifth grade of students who received DI and those who did not. All the results, whether with the full or reduced set of Control Schools, pointed to the same substantive conclusions: Students in the Direct Instruction schools had significantly greater gains than students in either of the set of Control schools on the measure of concepts and applications. Increases on the measure of computations and the composite measure did not differ significantly between the three groups. Figures 9 and 10 summarize these results.

Figure 9 reports the percentile score of an average student in each group of schools on each measure in both first grade and fifth grade. On average, all students, in both the DI and the Control Groups, had higher average achievement scores in fifth grade than in first grade. In first grade, the average student scored well below the national average, but by fifth grade, the average student, in all three groups, had scores that were much closer to, and in some cases above, the national average.



Note: Data in the figure represent the percentile at which an average student in a Control School or a DI school scored in first grade and in fifth grade. As with the analysis in Part A, all statistical calculations were conducted with Norm Equivalent Scores, which can be compared from one year to another, and the results were translated to percentiles for graphic display.

Figure 10 gives the percentage change in scores from first grade to fifth grade for students in each group on each measure. Students in the DI group had the highest average gain on both the measure of concepts and applications and the composite measure but, as noted above, these differences were only significant with the measure of concepts and applications. On average, students in the DI schools had concepts and applications scores in fifth grade that were 30 percent higher than their first grade scores, compared to an average increase of 3 percent for students in the full set of Control schools and 20 percent for students in the reduced set of Control schools.

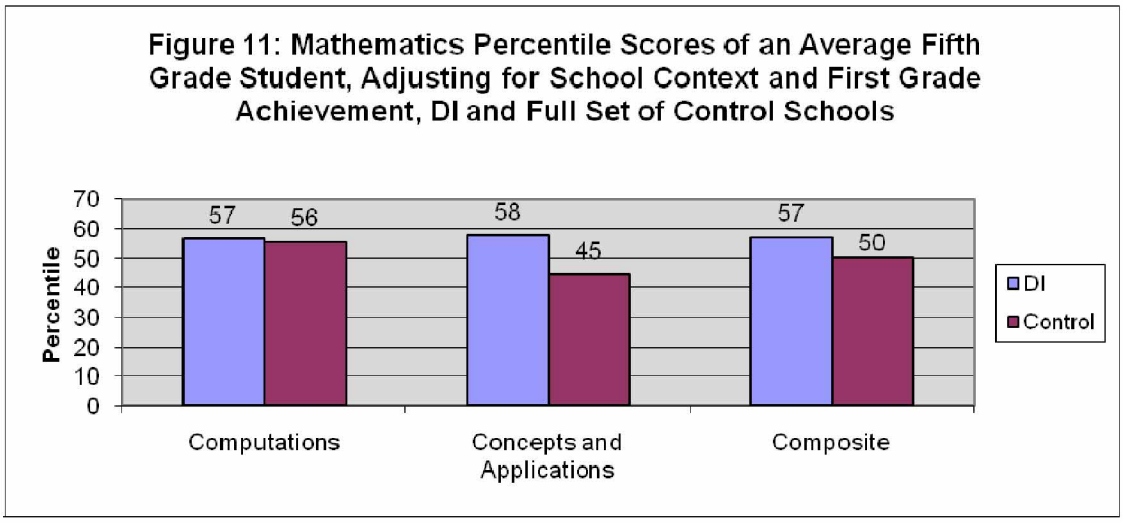


Note: The values in this table were calculated by dividing the difference in the fifth and first grade norm equivalent scores by the first grade score. Details on these calculations are in Appendix IV.

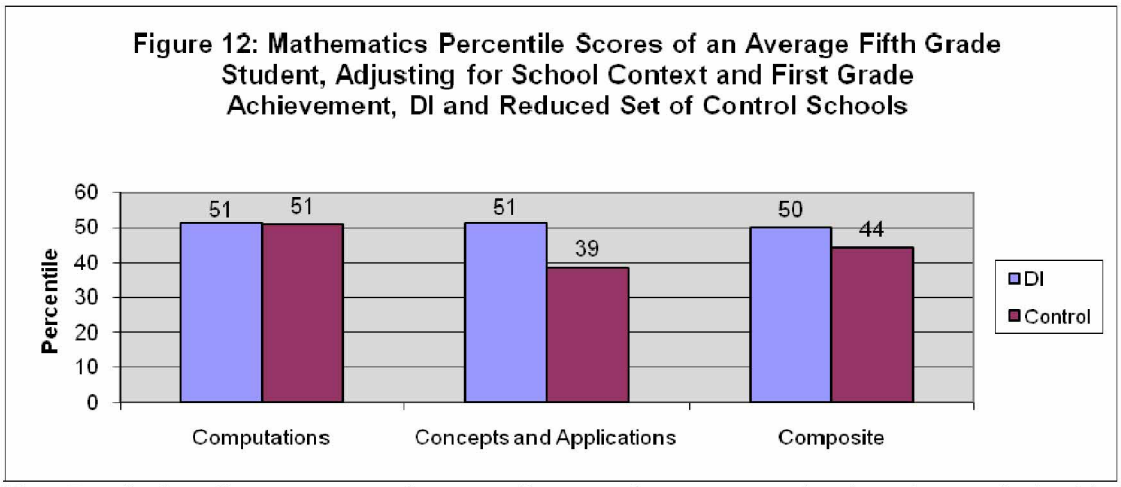
2. Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?

As described in Appendix IV, multivariate statistical techniques were used to address the second research question. This analysis compared the differences between students in the DI schools and those in the two sets of Control schools after statistically equalizing students on both their level of first grade achievement and the socio-demographic characteristics of their schools. The results were similar with analyses with both the full set of Control schools and the reduced set and parallel the results with Research Question 1: Students in Direct Instruction schools had significantly higher scores on the measure of mathematics concepts and applications than students in other schools, even when school context and first grade achievement were controlled. While the DI students also had higher scores on the measure of computations and the composite score, these differences were not statistically significant.

Figures 11 and 12 summarize the results of this analysis. The data in the figures are the percentile score that an average student in each group of schools would be expected to achieve if the schools had equal socio-demographic characteristics and if the students all had equal levels of achievement in first grade. Figure 11 compares the DI schools with the full set of Control schools, while Figure 12 compares the DI schools with the reduced sample of Control schools.[3] In both comparisons, the students in the DI Schools have higher levels of achievement, but the differences are substantially greater with the measure of mathematics concepts and applications.



[3] Different multivariate analyses were conducted with these two separate samples. The lower average values in Figure 12 reflect the fact that the reduced sample of Control Schools had lower levels of first grade achievement as well as a socio-demographic school context that was more high-risk.

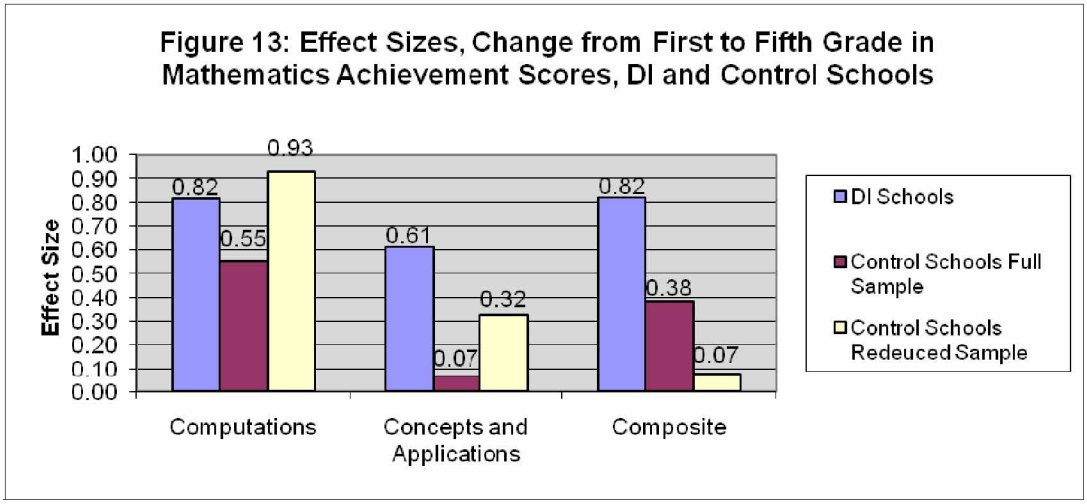


Note: Data in these figures represent the percentile score of an average student in each type of school in fifth grade if they had the same achievement scores in first grade and were in schools with equal socio-demographic characteristics.

3. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

Standard tests of statistical significance are influenced by the size of a sample in the calculations. To avoid this bias researchers sometimes use measures of effect sizes. With data involving change over time these calculations consider both the size of the scores at the two time points as well as the correlation between these two measures. (Details are in Appendix IV.) Traditionally, measures of effect size of .25 or larger are considered educationally important.

Figure 13 gives the effect sizes associated with the changes from first to fifth grade for each of the measures of achievement for students in the DI schools and for those in the two sets of Control schools. All of the effect sizes associated with the DI schools far surpass the .25 criterion. For the full set of Control schools, effects surpass the .25 criterion for both the measure of computations and the composite score. For the reduced set of Control schools, effects surpass the .25 criterion for the measure of computations and the composite score.



C. Summary and Discussion

The first section of the report examined changes in the average achievement of first graders in the BCPSS from 1998 to 2003, comparing students in schools that implemented Direct Instruction and students in other schools in the district. Results indicated that, averaging over all years, first grade students who received Direct Instruction had significantly higher levels of mathematics achievement than students who did not receive Direct Instruction. These differences became stronger when measures of the school context were statistically controlled and also became stronger when trends were examined over time. The students in the DI schools had achievement scores that were significantly lower than those in the Control schools at the beginning of the implementation (1998) but scores that were significantly higher by the end (2003). The percentage change in average first grade achievement was more than twice as high for students in the DI schools than in the Control schools. These differences in achievement were statistically significant and also met established criteria for educationally meaningful results.

The second part of the report looked at a smaller subset of students: those who began first grade in the BCPSS in 1997-98 or 1998-99 and remained in the same schools as fifth graders in 2001-02 or 2002-03. It and examined changes in the students' achievement from first to fifth grade. All students in the sample had higher levels of mathematics achievement in fifth grade than in first grade. However, students in the DI schools had significantly higher increases in their mathematics concepts and applications scores than other students. This result appeared when comparisons were made to students in all Control schools and in a more limited set of low achieving schools. They also continued to appear when strong controls for first grade achievement and school context were applied.

Several possible implications of these results should be noted: First, they provide additional support to the large body of studies that has documented the superiority of Direct Instruction in promoting mathematics achievement. Students in schools with Direct Instruction had significantly higher achievement than students in schools with other curricula. These results appeared in first grade on all measures and in fifth grade on the measure of mathematics concepts and applications. The differences met and, usually, surpassed the criteria of statistical significance and educational importance.

Second, the results show the necessity of examining a relatively long period of time to assess the full implications and results of school reform efforts. The changes in first grade achievement only occurred after several years of implementation within a school. In addition, it should be noted that all of the DI schools had strong technical support from the National Institute of Direct Instruction. Other work on reading achievement has shown that such strong technical support is important for helping to ensure that strong achievement gains occur. This result also may address a common claim that Direct Instruction involves a relatively mindless application of a script with little intellectual engagement or skill required by the teacher. The fact that strong changes in achievement only appeared over an extended period of time suggests that learning to teach Direct Instruction involves a great deal of skill and practice. Those that are most successful have more practice and more skilled, technical guidance.

Finally, it is important to note the impact of Direct Instruction on the measure of concepts and applications, typically assumed to be tapping more of a "higher order" skill than the measure of computations, the other element of the composite score. Some have suggested that Direct Instruction is useful for teaching rote, elementary skills such as computations, but has more limited utility in teaching higher level skills such as those tapped by the mathematics concepts and applications measure. The results in this paper would contradict that view, for it was in the area of concepts and applications that the long-term impact of Direct Instruction was most apparent.

Appendix I
Sample and Measures – First Grade Analysis

This appendix provides details on the sample and the measures used in the analysis of first grade mathematics achievement.

Sample

The analysis uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Eleven schools implemented Direct Instruction with technical assistance provided by NIFDI and are referred to below as the DI schools.[4] The remaining 103 schools are termed the Control schools. Schools in the intervention group are listed in Table A-1[5]

Table A-1 also lists the date at which programs were implemented in each school. All schools that used Direct Instruction began their implementation with the reading programs. They added DI instruction in mathematics in a subsequent year. In all but two schools (City Springs and Collington Square) the Direct Instruction mathematics curriculum was added one year after beginning reading instruction.

In the analysis schools were designated as belonging to a given condition only during the years in which their school was receiving the mathematics implementation. For instance, Collington Square had no mathematics DI instruction in either the first or second year of data collection (1997-98 and 1998-99), but began implementing Direct Instruction mathematics programs in 2000-2001. Thus, first grade students in Collington Square were included in the mathematics intervention group in 2000-2001 and subsequent years, but in the control condition in the two earlier years.

The analysis also considers the amount of time that a school had implemented the program. For instance, Collington Square was considered to be in its first year of implementation of Direct Instruction in 2000-2001, its second in 2001-2002, etc.

Data were available for a total of approximately 45,000 students (n = 44,766 for the composite measure, n = 45,287 for the measure of computation, and n = 45,535 for the concept and applications measure). Enrollment in the BCPSS declined from 1997-98 to 2000-2003, with data available for approximately 9000 students for the first year and 6400 for the last year. On average, data were available for approximately 380 students per school. Only four schools, all in the control group, had fewer than one hundred students in the sample.

---

[4] One additional school, Charles Carroll, was part of the original intervention group but was closed shortly after the start of the study period. Because data are not available throughout the time span of the study, data from that school are not included.

[5] There were two additional schools in the set of control schools, but they, unfortunately, had the same name: Highland Town. The number assigned to the schools was not available for all years, and alternative spellings of the schools' name across years made it impossible to clearly differentiate them. Thus these two schools were eliminated from the analysis. Data were available for one home-schooled student and that was also omitted.

Table A-1: Schools in Treatment Group and Start Dates of Reading and Mathematics Direct Instruction

	Reading Start Year	Math Start Year
Arundel	96 - 97	97 – 98
CC Barrister	97 - 98	98 – 99
City Springs	96 - 97	98 – 99
Collington Square	98 - 99	00 – 01
Dickey Hill	98 - 99	99 – 00
Federal Hill	97 - 98	98 – 99
General Wolfe	96 - 97	97 – 98
Hampstead Hill	96 - 97	97 – 98
Langston Hughes	98 - 99	99 – 00
Margaret Brent	98 - 99	99 – 00
Dr. Rayner Browne	98 - 99	99 – 00

Measuring Achievement

The Comprehensive Test of Basic Skills (CTBS), a widely used standardized achievement test, was administered to all first graders in the spring of each year, from 1997-1998 through 2002-2003. The 4th edition was administered in the spring of 1998 and 1999, and the 5th edition was administered in the spring of 2000 through 2003. Two subtest scores, Mathematics Concepts and Applications and Mathematics Computations, and a Composite Mathematics Achievement score were analyzed.

Normal Curve Equivalent (NCE) scores were used for all statistical analyses. NCE scores are calculated by translating the percentile scores to a distribution that is normally distributed. Like percentile scores, NCE scores range from 1 to 99 with a mean of 50. However, while percentile scores are evenly distributed (the graph of percentile scores would look like a rectangle), NCE scores comprise a normal distribution. The transformation results in scores that can be meaningfully added and subtracted, so that the difference of an NCE score of 50 and 55 (=5) is the same as the difference between 30 and 35. This interval scale allows computations of most common statistics. Percentiles and raw scores do not meet the technical requirements of interval scales, and thus statisticians urge researchers to analyze data based on NCE scores rather than percentiles or raw scores. All statistical calculations for this report were done using NCE scores.

The meaning of NCE scores is, however, not intuitively obvious. For this reason, results in the body of the text and the executive summary have been translated into percentile scores. After completing statistical calculations with NCE scores the resulting descriptive information

was converted into percentiles for display in the graphs in the body of the report. This conversion was done using a standard conversion table.[6]

Measuring School Context

Given the strong influence of school context on student achievement, it was important to develop an efficient, yet strong, measure of the demographic context of the schools in the sample. Preliminary analysis indicated that the demographic characteristics of schools were very highly correlated from one year to another. Thus, one summary measure was developed for each school that would be valid for all the years included in the data set.

There were a few cases in which data were not available for a school for all years. A regression-based method was used to predict values of missing cases from other years (e.g. predicting 1998 levels of proportion African-American from levels in 1999 through 2003).[7] Values for each variable were averaged across the years to produce aggregate measures of the demographic characteristics of the school.

Table A-2 gives means and standard deviations on each of the demographic characteristics for the DI and the Control schools. Results of t-tests indicated no significant differences between the groups of schools in their racial-ethnic composition. However, the DI schools had significantly more students who received free or reduced lunch: 83 percent in the DI schools versus 74 percent in the Control schools.

A factor analysis was conducted to develop summary measures of school context. Six variables were included in this analysis: the average proportion of Asian, Hispanic, African American, Native American, and non-Hispanic White students as well as the average proportion of students receiving free and reduced lunch. A principle components extraction method with varimax rotation was used. Two significant factors (with eigenvalues greater than 1.0) were found.[8] Factor scores were computed for each school and saved.

Results for the factor analysis are given in Table A-3. The commonalities indicate that, except for the measure of proportion of Asian students, there is a great deal of shared variance among the indicators, ranging from .66 for the proportion of students receiving free and reduced lunch to .94 for the proportion of African American students. Eigenvalues are also relatively high, with 73.5 percent of the total variance between schools explained by the two factors.[9]

---

[6] The conversion can also be accomplished manually through using a normal curve table. The NCE scores can be converted to z-scores (z = ((nce-50)/21.06))). The percentile that corresponds to the z score can then be found in the normal curve table.

[7] There were two cases where a predicted value fell outside of the theoretical range. Both of these involved the predictions for proportion white, where the predicted values were less than zero. For these cases (Malcolm X and Mildred Monroe schools), an average of the other years in the data file was used as the predictor. The average values for both cases were .01 or less.

[8] The factor analyses were also conducted with data for each year separately (i.e. the non-aggregated data), and the results were virtually identical to those obtained with the aggregated data.

[9] N = 120 for these analyses, for Carroll School was included.

Table A-2: Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels By Group

	DI		Control			
	Mean	S.D.	Mean	S.D.	t	prob.
Asian	0.01	0.03	0.01	0.01	-0.88	0.40
Black	0.75	0.34	0.85	0.26	1.12	0.27
Hispanic	0.05	0.12	0.01	0.01	-1.10	0.30
Native American	0.01	0.02	0.002	0.004	-1.31	0.22
Non-Hispanic White	0.17	0.25	0.13	0.25	-0.49	0.62
Free and Reduced Lunch	0.83	0.06	0.74	0.15	-4.02	<.001
Factor 1	-0.12	0.96	0.02	1.01	0.44	0.66
Factor 2	1.18	2.82	-0.12	0.46	-1.53	0.16
N	11		108			

The first factor, which accounts for 44% of the total variance, has a strong positive loading of the proportion of non-Hispanic White students, a slightly smaller positive loading for the proportion of Asian students, and strong negative loadings for the proportion of African American students and the proportion receiving free and reduced lunch. Thus, schools with positive scores on this factor would have proportionately more non-Hispanic White and Asian students and many fewer African-American and poor students. As can be seen in Table A-2, there are no significant differences between the two groups on this factor.

Table A-3: Factor Analysis of School Characteristics, Elementary Schools, BCPSS,1998-2003

	Rotated Factor Loadings		Communalities
	Factor 1	Factor 2	
Proportion Asian	0.64	0.09	0.42
Proportion African-American	-0.88	-0.40	0.94
Proportion Hispanic	0.13	0.85	0.74
Proportion Native American	0.08	0.87	0.77
Proportion Non-Hispanic White	0.90	0.27	0.89
Proportion Free/Reduced Lunch Eligible	-0.80	0.14	0.66
Eigenvalues	2.66	1.75	
% of total variance	44.4	29.1	

N = 120 schools

The second factor, which accounts for 29% of the total variance, has positive loadings on the proportion of Hispanics and the proportion of Native Americans in a school, and a negative loading, of a somewhat smaller magnitude, of the proportion of African Americans. Thus, schools with a positive score on this factor would have proportionately more Hispanic and Native American students and somewhat fewer African American students. The results in Table A-2 again indicate no significant differences between the groups on this factor.

Appendix II: Detailed Statistical Results – First Grade Analysis

This appendix expands upon the material presented in the text including an extended discussion of the analysis techniques employed and the results that address each research question.

Methodology

As noted in the text, the data analysis focused on comparing the mathematics achievement of first grade students who received Direct Instruction with those in the Control Schools. Four research questions were addressed.

The first research question was, "Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?" Simple t-tests were used to answer this question, comparing the average achievement scores of students in the DI schools with scores of students in the Control schools, combining data from all years of implementation.

The second research question asked, "Did students who received Direct Instruction have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?" Mixed model regression, including measures of school advantage and disadvantage from the factor analysis as predictors, was used to answer this question. To the extent that Direct Instruction is more effective at promoting student achievement, it would be expected that students in the DI Schools would have higher achievement scores than those in the Control Schools, even when school characteristics were controlled.

Mixed models are particularly appropriate for analyzing multi-level data, such as data regarding students and the schools that they attend. In these models a "random variable" is used to control for differences between schools (often termed the level 2 entity) while calculating regression coefficients regarding the impact of variables from both students and schools on achievement. The random variable is equivalent to having a separate intercept in the regression equation for each school. The coefficients associated with the various individual and school related variables are then calculated while this between school variance is controlled. The analysis also allows one to calculate the amount of variance in the dependent variable that occurs between schools and the extent to which various independent variables can account for this between school variance.

The third research question asked, "Did the impact of being in a DI school increase over time as reforms became institutionalized within a school?" It was expected that the impact of the DI implementation would become stronger over time. This question was analyzed both through examination of mean scores over time, with independent sample t-tests to compare the means, as well as through adding variables regarding length of implementation to the mixed models.

The final research question asked, "What is the magnitude of the effect of Direct Instruction on mathematics achievement." To address this question, the results were translated into effect sizes. Effect sizes are commonly used to summarize the magnitude of differences

between two groups. They are calculated by simply dividing the difference in the means of two groups by their common standard deviation. The resulting figure thus describes the magnitude of a difference between two groups as a percentage of the standard deviation. Unlike tests of significance, which can be heavily influenced by sample size, effect sizes are unaffected by the size of a sample. Effect sizes greater than .25 have traditionally been considered educationally significant.

Research Question 1: Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?

The first research question addressed was, "Do students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?" This was answered through comparing the average achievement scores of students in the DI schools with students in the Control schools. In this analysis data for all years in the sample (1997-98 through 2002-03) were combined.

Table A-4 summarizes the information used for these comparisons. Panel A reports the means and standard deviations of the normal curve equivalent scores for students who received Direct Instruction and students in the Control schools on each of the measures of achievement as well as the results of t-tests examining the hypothesis that any differences occur by chance. Panel B translates the average Normal Curve Equivalent scores in Panel A to the corresponding percentiles. The values in Panel B may be interpreted as the score of an average student in each type of school over all the years in the study and were used to construct Figure 1 in the text.

Table A-4: Average Mathematics Achievement Scores of Students With and Without Direct Instruction, All Years, First Graders, BCPSS

Panel A: Descriptive Statistics, Norm Equivalent Scores, and t-test Results

Computations	Control Schools	Direct Instruction
Mean	43.4	48.6
S.D.	24.9	25.8
N	42838	2449
t-value	9.73	
d.f.	2716	
prob.	<.001	
Concepts and Applications		
Mean	42.3	45.2
S.D.	22.1	22.9
N	43072	2463
t-value	6.40	
d.f.	45533	
prob.	<.001	

Composite

Mean	43.6	48.6
S.D.	24.0	25.4
N	42340	2426
t-value	9.47	
d.f.	2679	
prob.	<.001	

Panel B: Percentile Scores Corresponding to Mean Values

	Control School Students	Direct Instruction Students
Computations	37.7	47.4
Concepts and Applications	35.7	41.0
Composite	38.0	47.4

Note: For the composite measure and the measure of computations, the F-test comparing the variances of the two groups indicated that the hypothesis of equal variances should be rejected. Thus the t-values and degrees of freedom reflect the use of the unequal variance formulas for those two analyses.

The results in Table A-4 show that students in schools with Direct Instruction had significantly higher achievement scores than those in the Control Schools on all three measures: Computations, Concepts and Applications, and the Composite score. The average student in the Control schools, across all the years in the study, had scores ranging from the 35th to the 38th percentile. In contrast, the average student in a DI school had scores ranging from the 41st to the 47th percentile. The lowest scores for both groups occurred with the measure of Concepts and Applications. In neither of the groups did the average student reach the national average of the 50th percentile when scores were combined across all years.

Research Question 2: Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized?

Research Question 2 deals with the extent to which students in Direct Instruction schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized. (See Stockard and Mayberry, 1992, for an extensive discussion of the literature regarding such "school effects.") This question is addressed with mixed model regressions. Table A-5 summarizes the models that were tested and gives summary information on the results with each model. Panel A gives the variables that are included and Panel B gives the model fit statistics.

Table A-5: Variables in Models and Measures of Fit, Analysis of First Grade Mathematics Achievement, BCPSS, 1998-2003

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3	Model 4
Random intercept for Schools	x	x	x	X
Year		x	x	X
Factors 1 and 2		x	x	X
DI School			x	X
DI * Year				X

Panel B: Model Fit Statistics

Computations				
Random Effects Estimate	47.5	37.6846	38.0	37.8873
s.e.	6.5	5.1378	5.2	5.1685
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	582.9	515.3	514.5	514.3
s.e.	3.9	3.4	3.4	3.4
sig.	<.0001	<.0001	<.0001	<.0001
BIC	417320	411740.3	411646	411661.3
- 2 Log Likelihood	417306	411711.6	411660	411623.1
Change in LL		5594	52	37
Df		3	1	1
P		<.0001	<.0001	<.0001
PRE measure		0.12	0.001	0.0005
Correlation ratio	0.08			
Concepts and Applications				
Random Effects Estimate	46.4	34.8	34.5	34.6
s.e.	6.4	4.8	4.8	4.8
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	451.6	434.5	434.1	433.2
s.e.	3.0	2.9	2.9	2.9
sig.	<.0001	<.0001	<.0001	<.0001
BIC	408014	406238.6	406202.5	406112.9
-2 Log Likelihood	407999.7	406209.9	406169	406074.7
Change in LL		1790	41	94
Df		3	1	1
P		<.0001	<.0001	<.0001
PRE measure		0.04	0.001	0.0021
Correlation ratio	0.09			
Composite				

Random Effects Estimate	53.4	40.5	40.5	40.4
s.e.	7.3	5.5	5.5	5.5
sig.	<.0001	<.0001	<.0001	
Residual Estimate	533.2	474.8	474.0	473.47
s.e.	3.6	3.2	3.2	3.2
sig.	<.0001	<.0001	<.0001	
BIC	408554	403360	403293	403245
-2 Log Likelihood	508539	403331	403259	403207
Change in LL		105209	71	53
Df		3	1	1
sig.		<.0001	<.0001	<.0001
PRE measure		0.11	0.002	0.0012
Correlation ratio	0.09			

Note: The correlation ratio is computed by dividing the random effects estimate for schools by the sum of the estimate for schools and the residual. Thus, the ratio represents the proportion of total variation that is between schools. The PRE measure is the ratio of the difference of residual estimates of two models divided by the estimate from the less complex model. Thus it tells the proportionate change in the variance that occurs by adding more variables to a model.

Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average reading achievement. The correlation ratio attached to Model 1 is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that from eight to nine percent of the variance in achievement is between schools rather than simply between students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model and school differences are controlled. These null hypotheses can be easily rejected with Model 1 and with all subsequent models. There is significant variation between schools and also between students in all models that we examine. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those available to in this analysis.

The models become incrementally more complex, with each subsequent model adding one or more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the year in which data were collected and the two factor scores to the baseline model. Year was included to test the hypothesis that test scores vary over the years in the study (1997-98 through 2002-03). This is important to control for any general changes within the district. Including the two factor scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African

American students and more non-Hispanic white students, would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient).

Model 3 then adds the grouping variable – the indicator of attending first grade in a DI school or a Control School. It was expected that, net of the year data were gathered and the measures of school context, students in DI schools would have higher levels of achievement (a positive coefficient).

The -2 log likelihood measures and the BIC values in Panel B of Table A-5 can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the comparisons between these values are in the bottom part of each section of Panel B of Table A-7. For example, Model 2 provides a significantly better fit to the data than Model 1 for all measures. For computations, the change in the -2 Log Likelihood = 5594 (417,306 -41712). With three degrees of freedom (because three new variables were added to Model 2 compared to Model 1), this result is highly significant. The comparisons of Model 3 with Model 2 and of Model 4 with Model 3 also indicate that adding the variables in the more complex model significantly improves the fit. (See the significance associated with the change in the -2 log likelihood.)

The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. Looking at all the models in Table A-4 it may be seen that the lowest BIC values appear for Model 4. The Proportionate Reduction of Error (PRE) measures are another descriptive measure of the incremental changes in fit of the models and simply reflect the proportionate changes in the residual variance from one model to the next. The greatest changes occur from Model 1 to Model 2 and the proportionate reduction of error in prediction is substantially less for each of the more complex models. Again, this could be expected given the large number of factors that affect student achievement.[10]

Table A-6 gives the coefficients associated with each of the models in Table A-5, beginning with Model 2. As expected, the coefficients associated with Factor 1 (where higher scores indicate schools with fewer students on free and reduced lunch and more non-Hispanic whites) are positive and highly significant and the coefficients associated with Factor 2 (where higher scores indicate more Hispanic and Native American children) are negative and also statistically significant. The coefficient for year is positive and significant in all models, indicating that, over time, students' achievement scores increased.

The coefficients in Model 3 directly test the second research question by including both a dummy variable for treatment group and the controls for the measures of school context. It should be recalled that students were only coded as having the treatment if they were in a school in a year when the school was in a treatment condition. The coefficients associated with being in a DI school are positive and highly significant ($p < .0001$) in analyses of all three measures of

[10] It should be noted that the values of BIC, the changes in the Log-Likelihood Ratio, and the PRE measures would alter if the order in which variables were introduced were changed.

achievement. In other words, once the demographic context of the school is controlled, students who received Direct Instruction in mathematics had significantly higher achievement than students in the Control Schools.

Table A-6: Coefficients Associated with Mixed Model Regressions of First Grade Mathematics Achievement, BCPSS, 1998-2003

	Model 2			Model 3			Model 4		
Computations	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-9754.1	127.1	<.0001	-9568.6	129.1	<.0001	-9442.3	131.7	<.0001
Year	4.9	0.1	<.0001	4.8	0.1	<.0001	4.7	0.1	<.0001
Factor 1	1.6	0.6	0.005	1.6	0.6	0.005	1.7	0.6	0.004
Factor 2	-0.8	0.6	0.16	-1.7	0.6	0.004	-1.5	0.6	0.009
DI	-----	-----	-----	7.9	1.0	<.0001	-3090.9	653.0	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	1.5	0.3	<.0001
Concepts and Applications	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-4902.4	116.6	<.0001	-4768.7	118.4	<.0001	-4532.7	120.7	<.0001
Year	2.5	0.1	<.0001	2.4	0.1	<.0001	2.3	0.1	<.0001
Factor 1	2.6	0.6	<.0001	2.7	0.6	<.0001	2.7	0.6	<.0001
Factor 2	-1.0	0.6	0.06	-1.7	0.6	0.003	-1.4	0.6	0.01
DI	-----	-----	-----	5.7	0.9	<.0001	-5824.1	599.9	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	2.9	0.3	<.0001
Composite	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-9060.7	122.7	<.0001	-8873.5	124.6	<.0001	-8687.6	127.2	<.0001
Year	4.6	0.1	<.0001	4.5	0.1	<.0001	4.4	0.1	<.0001
Factor 1	2.2	0.6	0.0002	2.3	0.6	0.0001	2.3	0.6	0.0001
Factor 2	-1.0	0.6	0.09	-1.9	0.6	0.002	-1.7	0.6	0.005
DI	-----	-----	-----	8.1	1.0	<.0001	-4568.5	631.2	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	2.3	0.32	<.0001

Table A-7 shows the predicted achievement scores for students in the DI and Control Schools based on the results in Model 3. These values were calculated by substituting average values for the Factor Scores and Year in the models and calculating predicted values for students in each group. Thus, the resulting values give the achievement score that would be predicted if students were equalized on the factor scores and time. Panel A gives the Normal Curve Equivalent scores and Panel B gives the percentile equivalent of these scores. Values in Panel B may be interpreted as the percentile at which an average student in each of group would be expected to achieve if schools were equal in their demographic characteristics. The values in Panel B were used in Figure 2 in the body of the text. The results in Table A-7 indicate the

magnitude of the advantage that accrued to students in the DI Schools for they have higher average scores on all three measures and surpass the national mean on both the measure of computations and the composite score.

Table A-7: Predicted Achievement Scores Assuming Equal School Contexts

Panel A: Norm Equivalent Scores

	DI	Control
Computation	50.8	42.8
Concepts and Applications	47.7	42.0
Composite	51.1	43.0
Panel B: Percentile Scores		
Computation	51.5	36.6
Concepts and Applications	45.6	35.2
Composite	52.0	37.0

Note: Scores were computed by substituting average values for the entire sample for Year and Factors 1 and 2 into the equations associated with Model 3 in Table A-6. The average value for year was 2000.3. The average value for Factor 1 was -.03203, the average value for Factor 2 was -.04252. The value for year does not equal 2000.5 (the average of 1998-2003) because there were fewer students in the BCPSS in later years than in earlier years. The average values for Factors 1 and 2 do not equal zero because the unit of analysis for computing the average values was students rather than schools. If the alternative values were used in computations, the substantive results would not differ.

Research Question 3: Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?

Research question three tests the hypothesis that the advantage to students in a DI School would become stronger as practices and procedures were institutionalized within a school. As noted in the text, it takes time for teachers and other school staff to fully adjust to and incorporate the nuances of a new curriculum. Thus, it would be expected that the advantages accruing to attending a DI School would become greater over time.

Model 4 in Tables A-5 and A-6 addresses this hypothesis by adding the interaction of attending a DI school and year. It was expected that the coefficient associated with the interaction of year and DI would be positive, indicating stronger effects of Direct Instruction as years of treatment increased. This hypothesis was confirmed. In each analysis the interaction term is positive and highly significant.

The coefficients associated with Model 4 in Table A-7 can be used to calculate the average change expected in the three groups of schools from one year to the next. For instance, first grade students in the control schools had an average gain from one year to the next of 4.4 points in their normal curve equivalent scores on the composite measure of mathematics achievement (the coefficient associated with year), controlling for differences in the demographic context of the schools. In contrast, the scores of first grade students in the DI Schools had an average gain of 6.7 normal curve equivalent points with each additional year of implementation of Direct Instruction (summing the impact of year and the impact of years of implementation: 4.4 + 2.3).

Table A-8 gives the scores that would be predicted for students in the two groups of schools if they attended schools with similar demographic characteristics in 1998, the first year for which data are available, and 2003, the last year. These scores were obtained by using the regression coefficients in Table A-6 for Model 4 and substituting the average factor scores, as was done for the calculations reported in Table A-9, and either 1998 or 2003. Panel A gives the resulting normal curve equivalent scores and Panel B translates these scores into the corresponding percentiles. The scores in Panel B can be interpreted as the percentile that an average student in a group of schools would be expected to achieve at a given year.

Table A-8: First Grade Mathematics Achievement Scores by Group and Year, Adjusting for School Advantage, BCPSS

Panel A: Norm Equivalent Scores

	Year	DI	Control
Computations	1998	35.0	32.0
	2003	66.5	55.7
Concepts and Applications	1998	33.2	36.7
	2003	59.2	48.2
Composite	1998	33.8	33.1
	2003	67.1	54.9

Panel B: Expected Percentile Score of Average Student

	Year	DI	Control
Computations	1998	24	20
	2003	78	61
Concepts and Applications	1998	21	26
	2003	67	47
Composite	1998	22	21
	2003	79	59

Even after just one year of implementation, in 1998, the average scores of students in the DI schools, after adjustments for the socio-economic status of the schools, are higher than those of students in the other schools on the computation and composite measures. By 2003 these differences become much stronger and more dramatic. Assuming that they attended schools with similar socio-economic contexts, by 2003 the average student in a DI School would be expected to score between the 67thth and 79th percentile on the various measures, while the average student at Control School would score from the 47th to the 61st percentile, depending on the measure of achievement.

Table A-9 presents the data that were used to construct Figures 3, 4, and 5 in the text. Unlike data in Tables A-7 and A-8, these data are simply the average achievement scores in each year on each measure for each group. No adjustments were included for school advantage/disadvantage. They simply compare the average scores in the schools in the two groups from 1998 to 2003. The t-tests in each row test the null hypothesis that the mean NCE scores of students in the two groups are equal.

The results show that students in the DI schools had significantly lower scores than students in the control schools in the first two years for which data are available (1998 and 1999) for the concepts and applications and the composite measures and the first year (1998) for the computation measure. However, by the end of the study period the situation had totally reversed. Students in the DI Schools had significantly higher scores than students in the Control Schools on both the computation and composite measure from 2001 to 2003 and on the concepts and applications measure in 2003.

Another way to examine the change in scores is to calculate the percentage change in average NCE scores from 1998 to 2003. The last row of each section of Panel A gives these figures for each group and measure. The percentage change in average scores was much larger for students in the DI Schools than in the Control Schools. For instance, with the composite achievement measure, average normal curve equivalent scores for students in the DI Schools increased by 135 percent, while scores for students in the Control Schools increased by only 54 percent, less than half as much. Differences between the three groups were most marked in the measure of Concepts and Applications. Scores of students in the DI Schools increased by 87% over the 6 year period, while those of students in the Control Schools increased by only 26 percent. (These data were used to create Figure 6 in the text.)

Panel B of Table A-9 translates the Normal Curve Equivalent scores in Panel A into the corresponding percentiles. (These percentiles were used to create Figures 3 through 5 in the text.) The percentile scores confirm the pattern that appears in the other analyses. Students scores increased in both groups, but the increases were substantially stronger for students in the DI Schools. For instance, on the composite measure of mathematics achievement the average student in the DI Schools scored at the 13th percentile in 1998, but by 2003 the average student in these schools scored at the 68th percentile. In contrast, the average student in the Control Schools went from the 24th percentile in 1998 to the 56th percentile in 2003. In general, the raw, unadjusted scores confirm the more efficient and parsimonious results obtained with the mixed models, which incorporate the statistical controls for school context.

Table A-9: Mathematics Achievement Scores by Year and Group, First Graders, BCPSS, 1998-2003

Panel A: Means and t-test Results

	Group				
	DI	Control	t-value	df	prob.
Computations					
1998	25	33	-4.76	176	<.001
1999	30	32	-1.27	343	0.21
2000	45	46	-0.56	7700	0.58
2001	53	50	2.26	608	0.02
2002	57	52	4.57	570	<.001
2003	59	53	5.01	530	<.001
Percent Increase	134	60			
Concepts and Applications					
1998	29	38	-5.64	176	<.001
1999	32	37	-4.42	342	<.001
2000	40	41	-0.10	7851	0.92
2001	46	45	0.91	7352	0.36
2002	53	47	5.97	6805	<.001
2003	55	48	5.98	518	<.001
Percent Increase	87	26			
Composite					
1998	26	35	-5.71	174	<.001
1999	30	33	-3.09	334	0.002
2000	44	45	-0.51	7610	0.61
2001	52	50	1.93	7248	0.05
2002	58	52	5.58	568	<.001
2003	60	53	5.89	523	<.001
Percent Increase	135	54			

Panel B: Percentile Equivalents of Norm Equivalent Scores

	DI	Control
Computations		
1998	12	21
1999	17	20
2000	41	42
2001	56	50
2002	63	54
2003	66	56

Concepts and Applications		
1998	16	29
1999	20	27
2000	32	33
2001	42	41
2002	56	44
2003	60	46
Composite		
1998	13	24
1999	17	21
2000	39	41
2001	54	50
2002	65	54
2003	68	56

Note: In Panel A, the smaller degrees of freedom reflect cases where F tests revealed that the variances of the two groups were unequal and the alternative formula for calculating t and degrees of freedom was used.

Research Question 4: What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?

The fourth and final question asked, "What was the magnitude of the effects that were found?" Cohen's d, a standard measure of effect size, was calculated to provide a descriptive measure of the extent to which students in the NIFDI supported schools had achievement scores that were superior to students in the other schools. Cohen's d is simply the difference between the average score of two groups divided by the common standard deviation. Traditionally, d scores of .25 or larger have been deemed educationally significant, although McLean and associates (2000) caution that effect sizes calculated with normal curve equivalent scores are conservative in nature.

Table A-10 gives the d scores comparing the average achievement of students in the DI and Control Schools**.** Effect sizes are given in Panel A for the unadjusted scores. Effect sizes in Panel B were calculated with school context controlled, based on the results in Tables A-7 and A-8. The standard deviations for the total group were used for all calculations, and all calculations were made with the norm equivalent scores.

With the unadjusted scores combined over all years, the effect sizes are positive, but below the .25 threshold. The effect sizes for 1998 are negative and surpass the .25 criterion, indicating that the students in the DI Schools had substantially lower achievement than students in the Control Schools at the beginning of the implementation. However, after 2000 all of the

effect sizes are positive and by 2003 they all surpass the .25 threshold. In other words, while the d score at the beginning of the study period indicates an educationally significant disadvantage for students in the Direct Instruction Schools, this had totally reversed by the end of the study period and the students in the DI Schools had an educationally significant advantage.

Table A-10: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Mathematics Achievement, BCPSS, 1998-2003

Panel A: Gross Effect Sizes

	Computations	Concepts and Applications	Composite
Total	0.21	0.13	0.21
1998	-0.33	-0.38	-0.39
1999	-0.07	-0.22	-0.16
2000	-0.03	0.00	-0.02
2001	0.11	0.04	0.09
2002	0.23	0.28	0.27
2003	0.25	0.32	0.31

Panel B: Effect Sizes Independent of School Context

	Computations	Concepts and Applications	Composite
Total	0.32	0.26	0.33
1998	0.12	-0.15	0.03
2003	0.48	0.52	0.54

Similar, but somewhat stronger, results occur with the measures derived from the results where school context was statistically controlled (see Panel B). These effect sizes far surpass the threshold of educational significance by the 7th year of implementation, ranging from .48 to .54.

Thus, the results with the effect sizes parallel the results with the other statistical analyses. Students in DI schools had higher achievement than students in the Control Schools, and the magnitude of this advantage increased substantially over the years of implementation.

Appendix III
Panel Sample: Changes from First to Fifth Grade

This appendix provides details on the sample used in the analysis of changes in achievement from first to fifth grade. The potential pool included only students who were in the BCPSS in 1997-98 or 1998-99 and thus could still be in the system in fifth grade in 2001-02 or 2002-03, the last years for which data were available. Thus, data for this study included student cohorts who began first grade in the BCPSS in 1997-98 or 1998-99 and were in the same schools in fifth grade in either 2001-02 (for the first graders in 1997-98) or 2002-03 (for the first graders in 1998-99). Students who were held back were not included.

There was substantial student turnover during the study period, as summarized in Table A-4. There were almost 18,000 students in first grade in the BCPSS in 1997-98 and 1998-99. Five years later over half of these students (52.8%) had left the BCPSS, and an additional 20 percent had transferred to another school within the system. Only 27 percent of the students who were in first grade in the system in 1997-98 and 1998-99 were still in the same school in 5th grade in 2001-2 or 2002-3. The panel sample included a total of 4,771 students. Of these, 164 were in DI Schools and the remaining 4,607 were in the Control Schools.

Table A-11: Turnover in the BCPSS, 1997-98, 1998-99 to 2001-02, 2002-03

	N	Percent
Left BCPSS (In first grade, but not in fifth)	9362	52.80%
Transferred within BCPSS (In BCPSS in fifth grade, but in a different school)	3595	20.30%
Panel Sample (Same School, first and fifth grade)	4771	26.90%
Total Number of Students In First Grade 1997-98 or 1998-99	17,728	100.00%

Students were considered to have received an intervention only if they could be considered to have begun receiving Direct Instruction in first grade.[11] As summarized in Table A-12, this limitation of the sample resulted in students in several of the schools listed in Table A-1 not being included within the intervention group. For example, students who were fifth graders

[11] Analyses were also conducted that considered students who began DI instruction in second grade as part of the implementation group as well as at any time in their elementary career. The results with these comparisons were substantively similar to those reported here. It was decided, however, to limit the intervention sample only to those who had received DI throughout their first 5 elementary years to provide the least contaminated test possible of the effect of the curriculum.

in 2001-02 or 2002-03 at Collington Square would not have had Direct Instruction in first grade (the 1997-98 or 1998-99 school year) and thus were included in the control condition. Students who were fifth graders in 2001-02 in City Springs, Barrister, and Federal Hill did not have Direct Instruction in mathematics in first grade and were included in the control group. But, students who were fifth graders in 2002-03 in these schools did have first grade Direct Instruction and were included in the treatment group.

Table A-12: Schools in Treatment Group by Year for Panel Analysis

School	Year Started DI in Math	Fifth Graders in 2001-02	Fifth Graders in 2002-03	N in Panel Sample
Arundel	97 – 98	yes	yes	37
General Wolfe	97 – 98	yes	yes	23
Hampstead Hill	97 – 98	yes	yes	55
CC Barrister	98 – 99	no	yes	23
City Springs	98 – 99	no	yes	16
Federal Hill	98 – 99	no	yes	10
Collington Square	00 – 01	no	no	
Dickey Hill	99 – 00	no	no	
Langston Hughes	99 – 00	no	no	
Margaret Brent	99 – 00	no	no	
Dr. Rayner Browne	99 – 00	no	no	

The pattern of persistence within the same school differed somewhat between the groups of schools, as shown in Table A-13, with the Direct Instruction Schools having a higher proportion of their students remaining within the same schools from first to fifth grade[12] About a third of the students in first grade remained in the same schools in fifth grade, compared to slightly over one-fourth of those in the Control Schools. These differences were statistically significant.

As would be expected the students in the panel group (those that were in the same schools in 1st and 5th grade) differed from other students in their achievement levels in first grade. As shown in Table A-14, students who persisted in the schools in which they attended first grade had significantly higher first grade mathematics achievement on all three scores. These differences appeared for both the Direct Instruction Schools and the Control Schools, but were slightly stronger for the Control Schools than for the DI Schools (see the marginally significant interaction effects for concepts and applications and composite scores).

[12] There was no way to control for the possibility that a student had attended multiple other schools between first and fifth grade. It is possible that some of the students in the panel sample were in the targeted schools in both first and fifth grade, but had attended other schools in the interim. Assuming that such children would have less of the "full treatment," this would bias results in a conservative direction.

Table A-13: Persistence in the Same School, by School Group

	DI	Control Schools	Total
Not in Panel Sample	66.8	73.3	73.1
In Panel Sample	33.2	26.7	26.9
Total (%)	100	100	100
N (first grade)	494	17234	17728
N (panel)	164	4607	4771

chi-square = 10.209, df = 2, p = .001

The results in Table A-14 also indicate that the students in the DI Schools had substantially lower average levels of first grade achievement on each of the measures than students in the Control Schools. These differences were highly significant.

Table A-14: Comparison of Achievement Scores of Panel Group and Other First Graders, DI and Control Schools

	Computations		Concepts and Applications		Composite	
DI Schools	Mean	s.d.	Mean	s.d.	Mean	s.d.
Non-Panel Sample	26.2	22.9	28.8	19.5	25.7	20.1
Panel Sample	33.3	21.9	36.2	19.0	33.0	20.1
Control Schools						
Non-Panel Sample	29.9	24.1	34.7	22.8	30.9	22.7
Panel Sample	40.2	25.1	46.2	22.2	42.2	23.0
ANOVA Results	F	p	F	p	F	p
Panel vs. Non-Panel	53.16	<.001	71.66	<.001	68.34	<.001
Treatment Group	20.14	<.001	50.89	<.001	40.49	<.001
Interaction	1.79	0.18	3.25	0.07	3.21	0.07

Note: Mean and standard deviation values in the table are computed from Normal Curve Equivalent (NCE) Scores.

In order to minimize the differences between the Control Schools and the DI schools the sample of schools included within the control group was also limited to produce a subsample that had a mean as close as possible to that of the DI schools. This was accomplished by rank

ordering all schools by their average composite mathematics achievement score and then selecting the subsample of schools (beginning with the lowest) that produced an average as close as possible to that of the Direct Instruction schools. The resulting sample includes the 164 students in the DI Schools, as noted in Tables A-12 through A-14, and 1,820 students in the Control Schools. The sample includes 50 different schools. The minimum number of cases from a school is 16, and the maximum is 101. There are no significant differences in first grade achievement between these two groups. Results with both groups of Control Schools are presented in Appendix IV and in the body of the text.

Appendix IV: Detailed Statistical Results – Fifth Grade Achievement

This appendix expands upon the material presented in the text, including a more extended discussion of the analysis techniques employed and the results related to each of the three research questions regarding the impact of DI instruction in first grade on mathematics achievement in fifth grade.

Analysis Methodology

The analysis focused on changes in students' mathematics achievement from first to fifth grade and the extent to which students in schools with Direct Instruction had greater gains over time. Three research questions were addressed. The first was, "Do students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools?" Several comparisons were used to examine this question. Simple paired t-tests were used to examine changes from first to fifth grade in both sets of schools. Second, independent sample t-tests were conducted to compare the average achievement scores of students in the DI and Control schools in both first and fifth grade. Third, a repeated measures analysis of variance was conducted with the first and fifth grade scores as repeated measures and the school as a factor. If gains were greater in the DI schools, the interaction between the repeated measures would be expected to be significant.

The second research question was, "Do students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school is controlled?" This question was addressed by employing a mixed model regression, having schools as a random variable and regressing students' fifth grade scores on first grade scores, the measures of a school context of advantage/ disadvantage, and school type (DI or Control). Successively more complex models were used and differences between these models were examined.

The third research question asked, "What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?" To answer this question, the results were translated into effect sizes. Unlike tests of significance, effect sizes are not influenced by the size of a sample. Because the issue of interest in this report is the change in scores from one year to another the standard method of calculating effect sizes (dividing the difference between two means by the common standard deviation) is not appropriate. Instead, the correction suggested by Dunlap and associates (1996) is used. The calculations involve an adjustment for the correlation between the two scores. Results may be interpreted in the same manner as Cohen's d. Effect sizes greater than .25 have traditionally been considered educationally significant (Fashola and Slavin 1990). However, McLean and associates (2000) caution that effect sizes calculated with norm equivalent scores, as with this analysis, are inherently smaller than with other metrics. Thus the effect sizes presented are a conservative estimate of the results.

For all analyses results are presented comparing the students in the Direct Instruction Schools with students in all the Control Schools as well as those in the schools with more similar levels of achievement in first grade. The latter group is referred to as the "reduced sample" of control schools.

Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools?

Three different approaches were used to address the first research question: paired t-tests, independent sample t-tests, and repeated measures analysis of variance.

Difference Scores and Paired t-tests

Table A-15 summarizes the data used to examine differences in achievement gains from first to fifth grade for students in the Direct Instruction and Control Schools using paired t-tests. Panel A gives descriptive statistics for each measure of mathematics achievement using norm equivalent scores. Panel B gives the average change from first to fifth grade, the correlations of first and fifth grade scores, and results of paired t-tests that examine the null hypothesis that there was no change, on average, in achievement from one grade to the other. Panel C translates the averages in Panel A into percentile scores. The data in Panel C can be interpreted as the percentile score that the average student obtained in the two groups of schools in first grade and fifth grade. The data in Panel C were used to construct Figure 9 in the text.

Table A-15: Descriptive statistics and Paired t-test Results, 1st and 5th Grade Mathematics Achievement Norm Equivalent Scores, DI and Control Schools, Full and Reduced Samples, BCPSS, 1998-2003

A: Descriptive Statistics (Norm Equivalent Scores)

		DI Schools Full Sample		**Control Schools Full Sample**		**Control Schools Reduced Sample**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Computations	1st.	33.1	22.1	40.3	25.1	30.8	21.7
	5th	50.2	19.5	53.4	22.1	51.2	22.0
	% ch.	0.52		0.32		0.66	
Concepts and Applications	1st.	36.2	19.0	46.2	22.2	38.7	20.3
	5th	47.1	16.2	47.5	17.4	44.8	16.9
	% ch.	0.30		0.03		0.16	
Composite	1st.	32.8	20.2	42.3	22.9	33.3	19.9
	5th	48.5	17.8	50.6	20.3	47.9	19.9
	% ch.	0.48		0.20		0.44	

B: Average Differences, Correlations, paired (dependent sample) t-scores

	Ave. Dif.	r	t	p	N
DI Schools					
Computations	17.1	0.45	9.85	<.001	159
Concepts and Applications	10.9	0.46	7.44	<.001	159
Composite	15.7	0.54	10.58	<.001	155

Control Schools Full Sample					
Computations	13.1	0.38	33.26	<.001	4495
Concepts and Applications	1.3	0.53	4.59	<.001	4498
Composite	8.3	0.52	25.80	<.001	4437
Reduced Sample					
Control Schools Reduced Sample					
Computations	20.4	0.34	33.94	<.001	1755
Concepts and Applications	6.1	0.53	14.06	<.001	1756
Composite	14.6	0.49	3.02	<.001	1726

Results indicate that students in both the DI and the Control Schools had significantly higher achievement from in fifth grade than in first grade. For both groups and each measure the average change in norm equivalent scores from first to fifth grade was statistically significant. Changes on all three scores were larger for students in the DI group than for those in the full sample of Control Schools. Changes for students in the DI group were larger than those for students in the reduced sample of Control Schools on both the measure of Concepts and Applications and the Composite Score. This can be seen in the average differences reported in Panel B as well as the percentage gain reported in Panel A. The largest differences between students in the DI Schools and those in the Control Schools occurred with the measure of concepts and applications. On average students in the DI schools had concepts and application scores that were 30 percent higher in fifth grade than in first grade compared to an increase of only 3 percent for students in the total group of Control Schools and 16 percent for those in the reduced sample of Control Schools. These data were used to create Figures 9 and 10 in the text.

Independent t-tests

Table A-16 provides an alternative way to examine the data by looking at the difference in average achievement scores of students in the DI schools and Control schools in first grade and in fifth grade. While the computations reported in Table A-15 compared the average difference between students' first and fifth grade scores, the computations in Table A-16 compare average scores of students in DI and Control Schools in first grade and in fifth grade.

Panel A includes comparisons of students in the DI Schools with those in the full sample of Control Schools. Results indicate that scores of those in the Control Schools were significantly higher than those in the DI schools on all measures in first grade but only on the measure of computations in the fifth grade. Panel B includes comparisons with the reduced sample of Control Schools. None of the comparisons in Panel B are statistically significant, although the results for Concepts and Applications in fifth grade approaches significance ($p = .12$).

Table A-16: Independent t-tests, Average Mathematics Achievement, DI and Control Schools (Full and Reduced Samples), 1st and 5th Grade, BCPSS, 1998-2003

Panel A: DI Schools vs. Full Sample Control Schools

First Grade	t	df	p
Computations	3.94	177	<.001
Concepts & Apps.	6.46	174	<.001
Composite	5.625	172	<.001
Fifth Grade			
Computations	2.01	174	0.05
Concepts & Apps.	0.37	4756	0.71
Composite	1.42	174	0.16

Panel B: DI Schools vs. Reduced Sample Control Schools

First Grade	t	df	p
Computations	-1.44	1934	0.15
Concepts & Apps.	1.46	1920	0.15
Composite	0.09	186	0.93
Fifth Grade			
Computations	0.53	197	0.60
Concepts & Apps.	-1.56	1974	0.12
Composite	-0.42	1957	0.68

Note: F tests indicated that the t-value assuming unequal variances between the groups should be used for all comparisons with the total group of control schools but the one for concepts and applications for fifth grade. The t-value assuming unequal variances was used with comparisons with the reduced sample for the composite score for first grade and the computations score for fifth grade. Negative t values indicate that students in the DI schools had higher average scores than students in the Control schools. Positive values indicate that the Control school students had higher scores.

Repeated Measures ANOVA

Table A-17 gives the results of a repeated measures analysis of variance with the first and fifth grade scores as a repeated measure and the type of school (DI or Control) as a factor. This approach is more parsimonious than the separate t-tests reported in Tables A-15 and A-16. Panel A gives results with comparisons with the total set of Control Schools and Panel B gives results with the reduced sample of Control Schools. If exposure to Direct Instruction in the first grade is related to higher achievement in fifth grade, a significant interaction effect would be expected. With comparisons with the total sample of Control Schools (Panel A) the results are as expected with the measure of concepts and applications and the composite score. With comparisons with the reduced sample, the results are as expected only with the measure of concepts and applications.

Table A-17: Repeated Measures Analysis of Variance, Norm Equivalent Mathematics Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools

Panel A: DI Schools versus Full Set of Control Schools

	Computations	Concepts and Applications	Composite
F-Mathematics	203.16	60.02	189.09
P	<.001	<.001	<.001
F- Group	10.76	13.94	14.32
P	0.001	<.001	<.001
Interaction	3.59	36.56	17.87
P	0.06	<.001	<.001
Df	1,4652	1, 4655	1, 4590

Panel B: DI Schools versus Reduced Set of Control Schools

	Computations	Concepts and Applications	Composite
F-Mathematics	329.69	126.76	327.31
P	<.001	<.001	<.001
F- Group	0.20	0.004	0.003
P	0.66	0.95	0.95
Interaction	2.59	10.03	0.43
P	0.11	0.002	0.51
Df	1, 1912	1,1913	1,1879

Note: The F test associated with mathematics tests the null hypothesis that the change from first grade to fifth grade equals zero. The F test associated with group tests the null hypothesis that the average mathematics score of students in the DI schools. The F test associated with the interaction tests the null hypothesis that the differences between groups are the same for each grade or, alternatively, that the difference between the two grades is the same for each group.

Research Question Two: Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?

Tables A-18 and A-19 report the results of the mixed model regressions. Panel A of Table A-18 summarizes the models that were used and Panels B and C give summary statistics associated with each model for the full set of Control Schools (Panel B) and the reduced set (Panel C). The models become increasingly more complex. Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average mathematics achievement in fifth grade. The correlation ratio attached to Model 1 is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that between 9 and 12 percent of the variance in fifth grade achievement is between schools rather than simply between

students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model and school differences are controlled. These null hypotheses can be easily rejected with all three models. There is significant variation between schools and also between students even when the explanatory variables are included. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those in this analysis.

The models become incrementally more complex, with each subsequent model adding more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the norm equivalent reading achievement score for year one and the two factor scores measuring the advantage/disadvantage of the schools the students attended. It was expected that students' first grade achievement would be positively and significantly associated with their achievement in fifth grade. Including the two factor scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African American students and more non-Hispanic white students, would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient).

Model 3 adds the treatment group, testing the hypothesis that students in DI Schools have higher achievement than students in the Control Schools when first grade achievement and school advantage/disadvantage are controlled. If having Direct Instruction in first grade influences students' fifth grade achievement, independent of their first grade achievement and school context, we would expect positive associations with this variable. It should be recalled that students were only coded as having Direct Instruction if they were in a school in a year when the school was in a treatment condition.

Table A-18, Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Mathematics Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3
Random intercept for Schools	x	x	x
First Grade Achievement		x	x
Factors 1 and 2		x	x
DI in First Grade			x

Panel B: Model Fit Statistics Comparing DI Schools and Full Sample of Control Schools

		Model 1			Model 2			Model 3	
Computations	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects Estimate	54.8	9.0	<.0001	51.4	8.4	<.0001	51.4	8.4	<.0001
Residual Estimate	434.6	9.0	<.0001	363.3	7.6	<.0001	363.2	7.6	<.0001
BIC	42452			40873			40877		
- 2 Log Likelihood	42438			40845			40845		
Change in LL				1593			0.0		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.11								
Concepts and Applications									
Random Effects Estimate	35.1	5.8	<.0001	34.5	5.5	<.0001	34.7	5.5	<.0001
Residual Estimate	267.7	5.6	<.0001	181.3	3.8	<.0001	181.0	3.8	<.0001
BIC	40303			37690			37686		
- 2 Log Likelihood	40297			37662			37653		
Change in LL				2634			9.0		
df				3			1		
p				<.001			<.01		
Correlation ratio	0.12								
Composite									
Random Effects Estimate	49.8	8.1	<.0001	53.2	8.3	<.0001	53.2	8.3	<.0001
Residual Estimate	361.6	7.5	<.0001	247.3	5.2	<.0001	247.2	5.2	<.0001
BIC	41535			38603			38606		
- 2 Log Likelihood	41522			38575			38573		
Change in LL				2947			1.6		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.12								

Panel C: Model Fit Statistics Comparing DI Schools and Reduced Sample of Control Schools

		Model 1			Model 2			Model 3	
Computations	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects Estimate	53.2	12.8	<.0001	57.4	13.5	<.0001	57.4	13.5	<.0001
Residual Estimate	422.0	13.6	<.0001	361.6	11.8	<.0001	361.6	11.8	<.0001
BIC	17527			16825			16829		
- 2 Log Likelihood	17515			16801			16801		
Change in LL				714			0		
df				3			1		
p				<.001			n.s.		

Correlation ratio	0.11								
Concepts and Applications									
Random Effects Estimate	26.4	6.7	<.0001	36.7	8.3	<.0001	38.3	8.7	<.0001
Residual Estimate	262.5	8.5	<.0001	168.8	5.5	<.0001	167.9	5.5	<.0001
BIC	16704			15389			15384		
- 2 Log Likelihood	16692			15365			15357		
Change in LL				1327			8.5		
df				3			1		
p				<.001			<.01		
Correlation ratio	0.09								
Composite									
Random Effects Estimate	42.7	10.4	<.0001	57.9	13.0	<.0001	58.6	13.2	<.0001
Residual Estimate	350.4	11.3	<.0001	237.6	7.9	<.0001	237.4	7.8	<.0001
BIC	17134			15764			15767		
- 2 Log Likelihood	17123			15741			15740		
Change in LL				1382			1.1		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.11								

The -2 log likelihood measures and the BIC values in Panels B and C of Table A-18 can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the comparisons between these values are also included. For both sets of comparisons the results indicate that Model 2 is the best fitting model for the analyses of computations and the composite scores, but Model 3 is the best fitting for the analysis of concepts and applications.

Table A-19 gives the coefficients associated with variables in Models 2 and Model 3 for each of the measures of achievement and for both comparisons. The coefficients associated with first grade achievement are positive and highly significant in all models for all three dependent variables, indicating that, as expected, students with higher mathematics achievement scores in first grade also have higher mathematics achievement scores in fifth grade. As expected, the coefficients associated with factor 1 are positive and statistically significant with all analyses for the total group of Control Schools, indicating that students have higher achievement scores in schools with more non-Hispanic white children and fewer African American children and fewer children receiving free or reduced lunch. Associations with Factor 2 are insignificant in all analyses. Coefficients associated with Factor 1 are not significant for the analyses in Panel B, perhaps because the variation in the factor is somewhat smaller in this group.

Table A-19: Coefficients Associated with Mixed Model Results, Fifth Grade Mathematics Achievement Regressed on First Grade Achievement, School Advantage/Disadvantage, and Attending a DI School

Panel A: Coefficients for Comparisons of DI Schools and Total Sample of Control Schools

	Model 2			Model 3		
Computations	B	s.e.	p.	b	s.e.	p.
Intercept	38.5	0.9	<.0001	38.5	0.9	<.0001
First grade score	0.4	0.01	<.0001	0.4	0.01	<.0001
Factor 1	1.4	0.7	0.055	1.4	0.7	0.06
Factor 2	-0.2	0.7	0.795	-0.3	0.8	0.76
Had DI in First Grade	------	------	------	0.5	3.2	0.87
Concepts and Applications						
Intercept	27.1	0.7	<.0001	26.9	0.8	<.0001
First grade score	0.4	0.01	<.0001	0.44	0.01	<.0001
Factor 1	1.7	0.6	0.004	1.7	0.6	0.004
Factor 2	0.1	0.6	0.90	-0.7	0.6	0.25
Had DI in First Grade	------	------	------	7.0	2.3	0.003
Composite						
Intercept	29.5	0.9	<.0001	29.4	0.9	<.0001
First grade score	0.5	0.01	<.0001	0.5	0.01	<.0001
Factor 1	1.5	0.7	0.03	1.5	0.7	0.03
Factor 2	0.1	0.7	0.89	-0.31	0.8	0.70
Had DI in First Grade	------	------	------	3.6	2.8	0.20

Panel B: Coefficients for Comparisons of DI Schools and Reduced Sample of Control Schools

	Model 2			Model 3		
Computations	b	s.e.	p.	b	s.e.	p.
Intercept	39.4	1.4	<.0001	39.4	1.4	<.0001
First grade score	0.4	0.02	<.0001	0.4	0.02	<.0001
Factor 1	-0.04	1.3	0.97	-0.04	1.3	0.97
Factor 2	-0.2	0.8	0.79	-0.2	0.9	0.80
Had DI in First Grade	------	------	------	0.1	3.3	0.97
Concepts and Applications						
Intercept	26.5	1.1	<.0001	26.1	1.1	<.0001
First grade score	0.5	0.02	<.0001	0.47	0.02	<.0001
Factor 1	1.0	1.0	0.32	0.9	1.0	0.35
Factor 2	0.3	0.7	0.69	-0.6	0.7	0.43
Had DI in First Grade	------	------	------	6.8	2.3	0.004
Composite						
Intercept	29.9	1.3	<.0001	29.7	1.3	<.0001
First grade score	0.5	0.02	<.0001	0.5	0.02	<.0001
Factor 1	0.4	1.2	0.76	0.4	1.2	0.77
Factor 2	0.1	0.8	0.91	-0.3	0.9	0.75
Had DI in First Grade	------	------	------	3.0	2.8	0.29

The coefficients associated with attending a DI school in first grade are positive in all analyses, indicating that students in the DI schools have higher fifth grade achievement when other variables are equalized. However, the coefficients are significant only in the analyses of the measure of concepts and applications. These results appear both in the comparison with the total group of Control schools and with the reduced sample. As expected, students who attended a DI school in first grade had significantly higher scores in the measure of concepts and applications than other children, even when the socio-demographic context of the school and their first grade achievement scores were controlled.

Table A-20 summarizes the impact of these effects assuming that students had the same first grade reading score and that all schools were equivalent in their socio-economic characteristics. Average scores for the total group were substituted into the regression equations for Model 3 in Table A-19 for these calculations.[13] Results are given both in the norm equivalent

[13] The average value for the total group for computations was 39.97; for concepts and applications, 45.81; and for the composite score, 41.9. The average value for factor 1 was .1495 and for factor 2 was -.0497. These values do not equal zero for two reasons. The unit of analysis for these computations is students, rather than schools; and the population of students in the panel sample, as described above, is more advantaged than the general population and thus comes from schools with higher scores on Factor 1 and lower scores on Factor 2 than did the original sample of students in first grade.

scores and as the corresponding percentile score for a student scoring at this average. It can be seen, with both sets of comparisons, that, if they were to have equal scores at first grade and to attend schools with similar socio-demographic characteristics, students in the Direct Instruction schools and the Control Schools would be expected to have almost identical computation scores. However, students in the Direct Instruction school would have scores on the composite and concepts and applications measure that were higher than those of students in the Control Schools. The difference in the concepts and applications measure is the largest. With calculations based on the entire population of students, an average student in a DI school would score at the 58th percentile on the measure of concepts and applications and the composite score, well above the national average, while an average student in the Control Schools would only score at the 44th percentile, well below the national average. If calculations are based on the analysis that used the reduced sample of Control schools, the average student in the DI Schools would score at the 51st percentile on the measure of concepts and applications and at the 50th percentile on the composite measure, while the average student in the control schools would be at the 39th and 44th percentile on these measures.

Table A-20: Simulated Fifth Grade Mathematics Achievement Scores (Norm Equivalent and Percentile of Hypothetical Average Student), DI and Control Schools, Controlling for School Context and First Grade Achievement, BCPSS, 2001-03

	Full Sample of Control Schools		Reduced Sample of Control Schools	
A. Norm Equivalent Scores	DI Schools	Control Schools	DI Schools	Control Schools
Computations	53.5	53.0	50.7	50.5
Concepts and Applications	54.1	47.1	50.7	43.9
Composite	53.8	50.2	50.0	47.0
B. Corresponding Percentiles				
Computations	56.6	55.7	51.3	51.0
Concepts and Applications	57.8	44.5	51.3	38.6
Composite	57.2	50.4	50.0	44.3

Note: Scores were calculated by substituting the average values of first grade achievement and factor scores in the equations for Model 3 in Table A-11. For the sample that includes all the Control Schools the mean of Factor 1 was 1.4 and the mean for Factor 2 was -.3. For the sample with the reduced set of Control Schools the mean of Factor 1 was -.08 and the mean for Factor 2 was .23.

Research Question Three. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

Because the measures of first and fifth grade achievement are statistically dependent, Cohen's d, the most often used measure of effect size, is not appropriate. To control for the dependence of these measures, effect sizes were calculated adjusting for the correlation between scores in the two years (Dunlap et al 1996). It was anticipated that effect sizes would be larger for the students who received Direct Instruction in the first grade than for other students. The effect sizes were calculated using the descriptive statistics from the norm equivalent scores. McLean and associates, 2000, have demonstrated that effect sizes for NCE scores are lower than those that would be obtained for equivalent raw and scaled scores. Thus, the calculations presented here should be seen as conservative estimates.

Table A-21 summarizes effect scores for the change in achievement from first grade to fifth grade on each of the measures. Using the criterion of .25 as an educationally significant effect size, it may be seen that all of the effect sizes associated with the DI schools clearly surpass that criterion as do several for the other groups. Both sets of Control Schools had educationally significant changes on the measure of computations, as did the full sample of Control Schools on the composite measure and the reduced set of Control Schools on the concepts and applications measure. The largest differences between the groups occur with the measure of concepts and applications and the composite measure, where the effect sizes are substantially larger for the DI schools. The very small effect sizes for the concepts and applications measure for the full set of control schools and for the composite measure for the reduced sample of control schools reflect the much lower t-values for those difference scores.

Table A-21: Effect Sizes of Change, First to Fifth Grade, by Group, Mathematics Achievement Norm Equivalent Scores

	Computations	Concepts and Applications	Composite
DI Schools	0.82	0.61	0.82
Control Schools Full Sample	0.55	0.07	0.38
Control Schools Redeuced Sample	0.93	0.32	0.07

Note: All effect sizes are based on norm equivalent scores.

References

Fashola, Olatokunbo S. and Robert E. Slavin. 1997. Promising programs for elementary and middle schools: Evidence of effectiveness and replicability. *Journal of Education for Students Placed at Risk* 2: 251-307.

MacIver and Kempner. 2002. The impact of Direct Instruction on elementary students' reading achievement in an urban school district. *Journal of Education for Students Placed at Risk* 7: 197-220.

McLean, James E., Marcia R. O'Neal, and J. Jackson Barnette. 2000. Are All Effect Sizes Created Equal? Paper presented at the 28th Annual Meeting of the Mid-South Educational Research Association, November, 2000, ED 448 188.

Stockard, Jean and Maralee Mayberry. 1992. *Effective Educational Environments.* Newbury Park, California: Corwin/Sage, 1992.

From: jstockard@nifdi.org
Sent: Tuesday, September 30, 2008 4:19 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number:
1104699251

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: I would like to submit this study for the elementary school math area. Thank you for considering it.
Please let me know if the PDF comes through.
thank you.

From: jstockard@nifdi.org
Sent: Monday, October 06, 2008 2:27 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number:
29487267

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: Re: submission for elementary mathematics:
reference # 1104699251
Please substitute this copy for the report submitted earlier.
The executive summary was wrong in the previous report. Nothing else has changed.
Thank you.
Jean Stockard
Director of Research
National Institute for Direct Instruction

From: WhatWorks
Sent: Tuesday, September 30, 2008 4:11 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse: Submit a Study to Review

Dear Dr. Stockard,

Thank you for your submission. We received your attached PDF of the study.

Please note that studies submitted to the WWC could be considered for review under the next round of reviews as determined by IES. We are unable to provide a timeline for the review process at this time.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

```
-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Tuesday, September 30, 2008 4:19 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID
Number: 1104699251

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: I would like to submit this study for the elementary school math
area. Thank you for considering it.
Please let me know if the PDF comes through.
thank you.
```

Prologue

On December 7, 2007, Jean Stockard wrote the first of a series of communications to What Works Clearinghouse about its unreasonable evaluation of Reading Mastery. WWC had asserted that Reading Mastery had no studies that documented its effectiveness as a "beginning reading" program. Jean listed several Reading Mastery studies that should be used as evidential support. She also pointed out some serious problems with WWC's acceptance of studies that purportedly provided evidence of effectiveness for Reading Recovery.

On June 25, 2008, Jean sent a more detailed critique to WWC and Mathematica, the organization that operates WWC under a federal contract. This communication was followed by what is best described as hemming and hawing. We became cautiously optimistic when we received a letter from Mathematica indicating that it was very concerned with the allegations Jean had made and was providing a careful appraisal of WWC's practices. Jean didn't receive a response that addressed the various issues she had raised until September 8. The response consisted of a detailed account from WWC and a letter from Mathematica stating that it found no irregularities in any WWC practice or judgment.

The response from WWC appears as Appendix A of "Machinations of What Works Clearinghouse." The justifications toggle between sophomoric and casuistic. Some are contradictory.

The response that I wrote was not addressed to WWC or Mathematica. It deals with the issues in the Sept. 8 letter from a logical but less empirical standpoint than Jean's analysis provides. The basic argument in my paper is that what WWC did could not have occurred by chance and was not the product of applying rigorous standards. Instead the outcome was the product of rubber standards, logically unsound arguments, and possibly confusion.

I consider WWC a very dangerous organization. It is not fulfilling its role of providing the field with honest information about what works, but rather seems bent on finding evidence for programs it would like to believe are effective (like Reading Recovery and Everyday Mathematics).

Machinations of What Works Clearinghouse

by Siegfried Engelmann

The following critique of What Works Clearinghouse (WWC) is based largely on a letter sent to Jean Stockard from Mathematica on September 8, 2008. The letter appears as Appendix A.

The conclusion of this critique is that What Works Clearinghouse is so irreparably biased that it would have to be thoroughly reoriented and reorganized under different management rules to perform the function of providing reliable, accurate information about what works.

This conclusion derives from two facts:

1. There are over 90 studies that examine the effectiveness of Reading Mastery (and its predecessor, DISTAR Reading). Most of these studies have appeared in refereed journals.

2. The WWC has concluded that "No studies of Reading Mastery that fell within the scope of the Beginning Reading review meet WWC evidence standards" (http://ies.ed.gov/ncee/wwc/reports/beginning_reading/rdgmastery/).

In other words, there is complete discontinuity between two groups that are in the domain of evaluating whether studies document effectiveness. The Reading Mastery group is composed of over 150 professional researchers who conducted the studies and at least the same number of reviewers for refereed journals who judged that the studies provide evidence of effectiveness for Reading Mastery. This group also includes authors of several meta-analyses that summarized studies and those who reported the extensive research base of Reading Mastery, such as the American Institutes for Research.

The second group consists of those who judge effectiveness of studies for WWC.

The discontinuity in judgments between these two groups provides prima facie evidence that WWC reviewers use procedures, standards, and evaluation criteria that are not in agreement with criteria used by any of the professionals who judged that Reading Mastery studies provide evidence of effectiveness. If WWC had found and reported on all the studies, the likelihood that the rejection of all would have happened by chance is 1 in 1,237,940,000,000,000,000,000,000,000 trials.

In fact, however, WWC used a questionable ploy to reduce the number of studies it deemed to be reviewable, and it failed to locate a fairly large percentage of remaining studies.

The ploy was to disallow any studies that were reported earlier than 1985. The ploy eliminated at least 38 of the studies, dropping the total of reviewable studies to 54. If all of these studies had been reviewed by WWC, the odds of rejecting all of them by chance would be 1 in 18,014,398,000,000,000 trials. WWC reported that it located 61 of the studies, but all but 15 of these were not really studies worthy of review. (They were "success stories" and other types of anecdotal material.) So WWC located only about 27 percent of the reviewable studies. The probability that the rejection of all 15 legitimate studies was a chance occurrence is 1 in 32,768, but this number, coupled with the rationales that WWC used to reject these studies, leaves little doubt that the stripping of Reading Mastery's evidence base was the result of intent (possibly tainted by ineptitude).

Search Procedures

How could a serious search of the literature reveal only 27 percent of the studies? The WWC protocol for Beginning Reading[1] (WWC Evidence Review Protocol for Beginning Reading Interventions) provides an elaborate description of its search procedures (pp. 11–16). It lists sources such as ERIC thesaurus,

[1] http://ies.ed.gov/ncee/WWC/PDF/BR_protocol.pdf

PyschINFO thesaurus, and dissertation abstracts. It has a list of 29 “hand searched” journals and a list of gray-area sources, including associations, such as the American Educational Research Association. One of the gray areas searched was “prior reviews and research syntheses (i.e., using the reference lists of prior reviews and research syntheses to make sure we have not omitted key studies)” (p. 16).

Possibly the discrepancy between the number of studies conducted and those found by WWC hinges on the WWC’s definition of “key studies.” Using this reference-list search technique, Jean Stockard identified 54 studies that occurred no earlier than 1985, and 38 earlier studies. (See appendices.)

At best, there seems to be a striking contrast between what the Beginning Reading protocol indicates WWC does and the performance results of the search for studies involving Reading Mastery.

The Analysis of Legitimate Studies WWC Found

In addition to the procedural inadequacy is the discrepancy between the judgments of WWC and those of the authors and reviewers of the 15 studies that WWC found. Can the discrepancy in judgment be explained as a conspiracy, or is it the effect of scrupulously applying WWC’s rigorous standards?

The outcome is not the product of rigor. Some of the rejected studies have raw scores that show huge outcome differences between matched controls and the experimental group. These studies adhere to basic experimental-design-and- reporting procedures that have been in place since long before 1985. This paper discusses one of the studies in some detail.

The numbers and the discontinuity between WWC’s judgment and those of others who evaluated the studies favorably strongly suggest that WWC

intentionally applied selection criteria that were specifically designed to reject Reading Mastery studies.

Distortion Techniques

WWC describes its criteria for accepting studies in its protocol for Beginning Reading. Although this protocol does not have total concordance with reasonable scientific standards, rigorous application of its standards would result in at least some of the rejected studies being accepted. Therefore, the ultimate cause of at least some rejections has to be that WWC created distortions where they were necessary to achieve rejection of specific studies.

Following is a list of "distortion techniques" that were used by WWC.

1. The use of various standards and criteria that are not commonly recognized by the scientific community.

2. The use of justifications that are largely argumentative (based on correlations, not data about causations) and that have limited or no empirical data to support the position argued.

3. The use of floating standards so that experiments with similar design and results could be viewed variously as evidential support for an approach or as lack of evidence. A subtype of this category would be a specific cut point (for example, 1985) that could be ignored according to the "discretion" of the project manager or person who decides whether a given study is a nay or a yea.

WWC's letter of September 8 provides evidence of the three techniques.

Cut Date of 1985

An example of 1, 2, and 3 ("uncommon" standards, argumentative justifications, and floating standards) is the WWC limitation that no studies reported earlier than 1985 are accepted unless the WWC principal investigator deems the study important enough to report. This criterion uniquely affects Reading Mastery, which had at least 38 studies that had been generally

recognized as providing evidence of effectiveness (Appendix B). No other extant program had more than one or two studies. So given that this cut date affects only one model, but affects it in a serious way, why was the date established in the first place?

This kind of cut date has no precedence in science. Studies are recognized by their quality and the extent to which the conclusions drawn are consistent with current evidence. Since 1985, at least 54 studies have examined the effectiveness of Reading Mastery (Appendix C). These outcomes are perfectly consistent with the outcomes of the 38 earlier studies. Therefore, there is no scientific basis for applying the cut date of 1985 to Reading Mastery.

The justification that WWC provides for this cut date in its letter of September 8 is strictly correlational and is presented as modal conditionals (this may happen and that may happen), but it provides not one bit of evidence about whether the premise WWC espouses is based in fact.

> . . . the fact that preschool enrollment has increased, combined with the fact that more preschool and kindergarten programs run full-day, means that students in the early grades may be better prepared to receive reading instruction today than students 25 years ago. Moreover, it is possible that any changes in reading readiness over this period may not have been evenly distributed, since differences in reading ability by socioeconomic status and race are apparent at the kindergarten level . . . Any of these changes could have implications for the effectiveness of an intervention. If school readiness has increased, then an intervention that was effective 25 years ago may not be effective in more recent years. (p. 2, Appendix A)

This type of argument is categorized in logic as an argument from ignorance. Its basic form is: We don't know if the true condition is A or B. Therefore, we conclude that the true condition is A.

In this case, we don't know how many of these postulated possible causal relations are true; therefore, we'll assume that all of them are true (or could be true). In contrast, the logical conclusion to this situation would be either to state

“We don’t know so let’s not change it,” or ”Let’s do some pointed research to obtain information about these counterfactual conditions.”

The recitation of “possibilities” provides evidence of the instructional naiveté of the author. The assertion that the children are better prepared now and therefore what was effective 25 years ago might not be effective now is logically impossible. Lower performers make all the mistakes that higher performers make. They make additional mistakes that higher performers don’t make and their mistakes are more persistent, more difficult to correct. Therefore, if the program is easier for them now because of their higher degree of undefined “readiness,” they will make fewer mistakes and progress through the program sequence faster.

The justification WWC provides for the cut date in its letter of September 8 is specious:

> . . . [The date] is used for two reasons. First, by limiting the reviews to research to this time period, WWC reviews reflect reasonably current research. . . . Second, the timeframe ensures that the research reviewed is examining versions of interventions that are most likely to be available to practitioners today. (p. 2, Appendix A)

Unless there is current data to show that Reading Mastery was effective but is not effective now, there would be no need to remove its strong data base established before 1985. From an argumentative perspective, consider the difficulties that would have been created in other fields if the history of what works was erased every 25 years or so and had to be reestablished.

Nature of Beginning Reading

WWC’s cut date is also highly insensitive to the nature of beginning reading. Unlike history, paleontology, biology, and other areas that that are subject to change as the world and knowledge of the world change, beginning reading for grades K–3 is stable because nothing of significance has changed in

the last 40 years. The instructional goal is the same—to teach children strategies and information that would permit them to read material that could be easily covered with a vocabulary of 4,000 words. The frequency of these words has not changed. The syntax of the language has not changed significantly. For these reasons, the content of the first four levels of Reading Mastery has not changed over the years. If the 1972 edition of DISTAR were used with the training that is used today, the results would logically have to correlate .9 or better with the performance of children who went through the current edition of Reading Mastery.

Amazon.com provides evidence of this strong correlation. The program Teach Your Child to Read in 100 Easy Lessons is an abridged version of Reading Mastery, designed for parents and based on the 1972 edition of Reading Mastery. It has more than 450 reviews by parents. Two points about the reviews are important:

1. No other beginning reading program has anything approximating the number of positive reviews that Teach Your Child has.

2. There is no tendency that documents a greater percentage of negative reports in more current years. The program continues to have an average sales ranking of around 400th of all books sold by Amazon.com and more than 90% of the reviews rate the program with the highest positive rating. The book is ranked #1 in Family Activities; #3 in School-Age Children; and #8 in Education.

Project Follow Through

A floating standard associated with this cut date is expressed in the September 8 letter.

> WWC principal investigators have the option to expand the period for which studies can be reviewed, if they believe that important research will be excluded. (p. 2, Appendix A)

The principal investigator did not reinstate any Reading Mastery studies even though one of these studies, Project Follow Through, was the largest educational experiment ever conducted, involving 200,000 children, 22 models of instruction (many of which are around today), and 180 communities that spanned the full demographic range of at-risk students in grades K through 3. WWC presents a great deal of rhetoric about causal validity. A strong argument can be presented that Follow Through's procedures and design achieved more causal validity than any other effectiveness study of what works ever conducted.

As pharmaceutical effectiveness trials show, one of the greatest threats to internal validity is whether subjects take the medication on schedule and whether they provide accurate reports. The same problem occurs in education. Few effectiveness studies provide for reasonable monitoring; however, Follow Through had two levels of monitoring to assure that the participating sites implemented the adopted model according to the sponsors' specification. Most of the studies that WWC endorses as showing evidence of effectiveness have no provision for monitoring classrooms to determine the extent to which reports are accurate.

Other details of the Follow Through design and evaluation are as sophisticated as those of current studies. For each Follow Through model there were two comparison groups, one local and one national. The outcome differences of models were measured by units of "educational significance," which were defined as outcomes that were at least ¼ standard deviation higher than the comparison groups and that were statistically significant. How different is that from the criteria that WWC employs?

Fidelity of Implementation

In any case, the WWC principal investigator obviously didn't believe that the Follow Through study was in the class of "important research" that should be

excluded by the cut date. One possible reason is that WWC is not concerned with validation of whether there was fidelity of treatment (even though lack of validation is possibly the greatest threat to studies that require teachers to do things in a new way).

The September 8 letter makes WWC's position quite clear.

> . . . many studies include little information to gauge fidelity, especially information about whether an intervention has been implemented within normal operating regimes of districts, schools, and teachers, not under specialized laboratory conditions. (p. 3, Appendix A)

The technique that WWC uses as a sop for information about fidelity of implementation is to rely on replicated findings "which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies." (p. 3, Appendix A)

According to this logic, if four studies of the same method are conducted, and only two are conducted with fidelity of implementation, the fidelity is somehow inferred from the average: If the scores were 80, 83, 49, and 52, the average is around 66. What does that tell, either about what works or what is required to implement an approach that does work? Also, looking at the individual scores provides no information about which studies were implemented with fidelity and which weren't. A pessimist could look at the data and conclude that the lower two scores were implemented with fidelity. An optimistic interpretation is that the higher-performing outcomes were the result of high fidelity. Unless the fidelity level is indicated operationally or through observations, there is no clear information about what works and what is required to achieve an implementation that works.

Meta-Analyses

WWC rejected the interpretations of other meta-analyses with indefinite justification. The September 8 letter states,

> It is not surprising that WWC findings may differ from those of other analyses. Some meta-analyses may not have such rigorous standards for including studies; others may have standards that differ from those used

> by WWC. Because of these differences, the WWC cannot judge the results of its systematic reviews by how they compare to other analyses. (p. 1, Appendix A)

Possibly not, but WWC should be able to be quite a bit more detailed on the specific basis for each rejection.

Bias for Reading Recovery

The "rigorous standards" of the WWC analyses are not apparent in studies that support Reading Recovery. In contrast with Reading Mastery, the bias seems to be strongly in the direction of Reading Recovery. In two instances, WWC interpreted outcomes as being successful when they weren't.

A study by Baenen et al. (1997) showed that although there was some positive effect following the Reading Recovery intervention after the first year, there was no positive effect two years later. In other words, the comparison students had caught up to the RR students. Considering that the purpose of WWC is to identify what works and also that the WWC provides rationales based on the needs of "the practitioner," it would seem very important to identify this outcome. It would seem reasonable to point out to practitioners that if they installed Reading Recovery—a one-on-one program that is very expensive—the intervention (according to the study) will produce only transitory positive results.

The WWC interpretation of the Baenan et al. study in its September 8 letter contradicts the assertion that WWC "works to ensure that its standards appropriately identify studies with strong causal validity and applies those standards consistently to each study reviewed" (p. 1, Appendix A). WWC's justification for reporting outcomes that are misleading is based on a surprising WWC interpretation of "beginning reading." Although the WWC website describes beginning reading as encompassing grades K through 3, the letter states,

> . . . it is important to note that the beginning reading protocol prioritized one-year results. In effect, the Beginning Reading review is only intended to examine whether beginning reading interventions have an effect within one year. This one-year period is applied consistently to each study reviewed to ensure the results can be compared across studies and interventions. (p. 4, Appendix A)

This "one year" limitation or priority is not articulated anywhere in the document, WWC Evidence Review Protocol for Beginning Reading Interventions. The protocol document does not suggest that the grades are to be disaggregated or even that the outcome measures are limited to K through 3. In fact, it states the opposite.

> - **Sample relevance.** The sample must include students in grades K, 1, 2, or 3 learning to read English.
>
> – The intervention must have taken place in grades K, 1, 2, or 3; **outcome may be measured in grades K–3 or later.**
>
> (p. 4, Protocol)

According to this description, the intervention could be provided in grade K and results compared in grade 4. Furthermore, the use of the word include and the inclusive or indicates that the sample is not limited to one year. The usage does not rule out the possibility of a two-year study or a four-year study (K–3). No language within the Beginning Reading protocol contradicts this interpretation or asserts that studies with a duration of one year or less have priority status.

Reading Recovery reviews also contradict the claim that WWC works to identify studies that have strong causal validity and applies those standards consistently to each study reviewed. A 1993 study conducted by Iverson and Tunmer compared the length of time students in the two treatments required to "catch up" to students in the regular classroom. Both groups caught up to the children in the regular classroom, but Reading Recovery required substantially more time than the comparison approach. WWC reported only on the Reading

Recovery group, even though the point of the study and statements by the authors indicated that Reading Recovery was not as effective as the comparison intervention.

The letter from WWC explains its position as something of the WWC's right.

> . . . the WWC examined the results most relevant to the question of whether Reading Recovery improves reading proficiency compared to a reasonable counterfactual . . . As with any study it reviews, the WWC does not base the findings of its review on the conclusions drawn by the authors. (p. 4, Appendix A)

The letter further argues that the comparison of the more-successful treatment was mentioned in an appendix. The appendix, however, does not seriously reduce the degree of guile in the WWC report. Not only will very few readers bother with the technical trivia of an appendix; WWC frames this study as providing evidence that Reading Recovery is effective. No award of effectiveness went to the superior program in the study. Is that an example of consistent application of rigorous standards?

Finally, WWC accepted the Baenen Reading Recovery study even though it had a serious confound. The comparison group had no intervention program and no time scheduled for intervention. Therefore, the Reading Recovery group devoted considerably more time each day to reading than the comparison group did.

The letter explains why, in the judgment of WWC, these conditions do not confound the results or make it impossible for one to judge the effectiveness of Reading Recovery.

> Many practitioners are interested in knowing whether an intervention is effective relative to customary classroom practices. The WWC reviews studies comparing treatments to no-treatment as well as studies comparing one treatment to another. In each case, the counterfactual is clearly documented in the review. (p. 4, Appendix A)

Stated differently, practitioners are required to scour appendices to discover the major outcome of one study and accept outcomes of another even though total instructional time is not controlled.

The lack of internal validity of the latter study is revealed by the simple assertion, “If the comparison group used the regular reading program for the same amount of daily time that the Reading Recovery had, the achievement of both groups would be the same.” Given that the study provides no data to refute this assertion, the study has serious internal-validity problems.

Elastic Standards?

For the examples of Reading Recovery cited above, there was a strong bias in favor of Reading Recovery. In contrast, an unassailable study showing the effectiveness of Reading Mastery was rejected by WWC on the grounds that it had a confound. The RITE study (Carlson and Francis, 2002) involved 9300 students and 277 teachers, which made it probably the second largest instructional study ever conducted. Students in the study went from K through grade 2. A carefully matched comparison group was identified in schools geographically proximal to the RITE schools. The only reading program used in the experimental treatment was Reading Mastery.

All of the outcome measures favored the RITE students with differences between the groups growing progressively from K through 2. The effect sizes were large and most of the probability levels were < 0.0001.

WWC rejected the study on the grounds of what it identified as a confound.

> A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program. (p. 5, Appendix A)

This support consisted of summer training, less than two hours of monitoring during the year, and help from a designated trainer. Nearly half of the teachers (137) were in their first year of teaching Reading Mastery. The training focused on how to provide positive reinforcement, how to correct specific errors, how to organize and manage the classroom so that one small group is in reading instruction while the other two groups are engaged in independent work and are not disrupting the instruction.

The WWC letter implies that the support teachers received may have caused them to do things that are not part of Reading Mastery and that are not clearly specified in the Teacher's Guide. This is not the case. The teachers were trained to teach Reading Mastery exactly the way the Guide describes it, with all the technical details in place.

Is the training a confound? The WWC Beginning Reading protocol provides answers. The definitions and practices outlined on page 6 describe "branded" interventions and indicate "Branded interventions are commercial programs and products that may possess any of the following characteristics."

One of the characteristics is:

> Have an external developer who: Provides technical assistance (e.g., provides instructions/guidance on the implementation of the intervention). (p. 6, Protocol)

Not only was the training perfectly legitimate according to a careful reading of the WWC standards; it was also completely aligned with the WWC rhetoric about replication.

Under the heading "Elements of intervention replicability" is an appeal for replicability:

> The important characteristics of an intervention that must be documented in a study to reliably replicate the intervention with different participants, in other settings, at other times include: . . . .**The approach to enhancing the skill(s)**. (p. 6, Protocol)

The characteristics of the RITE intervention were explicit and the approach RITE used for enhancing skills is clearly specified in the description of the intervention. The procedures are fairly standard and are used with only slight variation by any of the external developers who work with Reading Mastery sites.

1. The training had no effect on the amount of contact time students received and no effect on the nature or the specific content presented, the sequence of activities, or procedures for teaching the program.

2. The description of the training alerts educators to the specific requirements of preparing teachers to teach the program (especially first-year teachers).

3. The justification for retaining the study as an example of what works is that no other study cited by WWC even approximates the magnitude or quality of this study.

4. The gains in reading skills are progressive from grade to grade, which is an indication that the intervention is not transitory or weak on any of the levels of the program that were used in K through 2.

5. Most important, the study provides prima facie evidence of replicability. It involved 277 teachers. The implementation itself therefore required replication in 277 classrooms over several years. This degree of replication is the equivalent in scope to 186 studies, each with 50 students.

What is particularly disturbing about the WWC dismissal of this study is that WWC treats it as an equal to any other study. The letter of September 8 makes reference to the fact that WWC may contact authors to obtain clarification on specific issues. The principal investigator who had a realistic perspective of the RITE study would certainly have contacted the authors to clarify whether any of the training distorted the content or procedures used in the classroom.

Instead, WWC dismissed this study with no more than its judgment of “a careful reading” of the study.

Action

There is no possibility that WWC achieved its dismemberment of Reading Mastery’s evidence base by fair or objective means. The WWC framework and review practices need to be vigorously challenged, both the cut date and the procedures for finding and reviewing studies.

We need to engage in a full-fledged assault on WWC. We need to involve the scientific community and get its sense of how to respond to WWC’s contrived conception of What Works and what kind of sanctions seem appropriate for WWC’s sophistry.

Appendix A (follows)

RESPONSES TO CONCERNS IN 6/25/08 LETTER FROM JEAN STOCKARD

RESPONSES TO CONCERNS IN 6/25/08 LETTER FROM JEAN STOCKARD

Contrasting conclusions between WWC and extant literature

The letter states that WWC conclusions differ from extant literature including meta-analyses and literature reviews. It is important to note that the WWC is designed to produce a systematic review of literature. A sound definition of a "systematic review" is in the recent publication *Knowing What Works in Health* Care by the Institute of Medicine (IOM):[1] *A systematic review is a scientific investigation that focuses on a specific question and uses explicit, preplanned scientific methods to identify, select, assess, and summarize similar but separate studies.* (pg. 82)

Consistent with the recommendations in the IOM report, the WWC applies evidence-based methodological standards consistently to each study it reviews. These standards, which are available on the WWC website, were developed by leading education research methodologists.

The systematic review conducted by the WWC goes beyond the procedures performed in meta-analyses. Again using the IOM definition, a meta-analysis "quantitatively combines the results of similar studies in an attempt to allow inference from the sample of studies included to the population of interest" (pg. 82). However, as Robert Slavin has noted, meta-analyses rarely describe even one study in any detail.[2] The WWC uses meta-analytic techniques to summarize the results of studies that meet WWC standards. However, the important distinction is that the WWC uses a rigorous set of standards, applied consistently, to determine *which* studies are included in the meta-analytic computations. This ensures that WWC summary measures are based only on studies with causal validity.

It is not surprising that WWC findings may differ from those of other analyses. Some meta-analyses may not have such rigorous standards for including studies; others may have standards that differ from those used by the WWC. Because of these differences, the WWC cannot judge the results of its systematic reviews by how they compare to other analyses. Rather, it works to ensure that its standards appropriately identify studies with strong causal validity and applies those standards consistently to each study reviewed.

---

[1] Institute of Medicine (IOM). *Knowing What Works in Health Care: A Roadmap for the Nation.* Washington, DC: The National Academies Press. 2008.

[2] Slavin, R.E. (1995). Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, Vol.48, No.1 pp.9-18.

Inclusion and exclusion procedures

Limiting studies to 1985 or later

The WWC's default time period for reviews is a study publication date of 1985 or later. This timeframe, which was established in 2005, is used for two reasons. First, by limiting reviews to research to this time period, WWC reviews reflect reasonably current research. In particular, the time period range ensures that effect sizes and improvement indices are based on a counterfactual condition that reflects classrooms as they operate within a recent time period. Second, the timeframe ensures that the research reviewed is examining versions of interventions that are most likely to be available to practitioners today.

WWC principal investigators have the option to expand the period for which studies can be reviewed, if they believe that important research will be excluded. The principal investigators for the Beginning Reading area chose to maintain the default period in large part to maintain currency with the classroom context for beginning readers. For instance, the fact that preschool enrollment has increased,[1] combined with the fact that more preschool and kindergarten programs run full-day,[2] means that students in the early grades may be better prepared to receive reading instruction today than students 25 years ago. Moreover, it is possible that any changes in reading readiness over this period have not been evenly distributed, since differences in reading ability by socioeconomic status and race are apparent at the kindergarten level.[3] Other contextual factors have changed over the past 20 years, including advances in teacher training, increases in home literacy activities, and changes in the content of and variety of curricula used in classrooms.

Any of these changes could have implications for the effectiveness of an intervention. If school readiness has increased, than an intervention that was effective 25 years ago may not be effective in more recent years. If teachers are receiving stronger training and using newer curricula, the counterfactual condition against which interventions were measured 25 years ago have changed, and possibly with it the magnitude of its effects. The Beginning Reading principal investigators judged that they had an inadequate basis for

[1] The proportion of 3 and 4 year olds enrolled in school increased from 37 percent in 1980 to 56 percent in 2006. See Snyder, T.D., Dillow, S.A., and Hoffman, C.M. *Digest of Education Statistics 2007*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2008.

[2] The proportion of pre-kindergartners and kindergartners attending full-day pre-kindergartners or kindergarten increased from 32 percent in 1980 to 59 percent in 2006. See Snyder, Thomas D. *Mini-Digest of Education Statistics, 2007*. Washington, DC: National Center for Education Statistics, Institute of Educational Sciences, U.S. Department of Education, 2008.

[3] For example, the Early Childhood Longitudinal Study of Kindergarteners (ECLS-K) found gaps in the reading knowledge and skills of kindergarteners by race: black and Hispanic children scored just under half of a standard deviation below whites on a test of reading knowledge and skills. Analysis of the ECLS-K and other surveys (i.e., Children of the National Longitudinal Study of Youth and the Infant Health and Development Program) show that socioeconomic status accounts for about half of the standard deviation of racial differences in reading test scores. See Duncan, G. and K. Magnuson. "Can Family Socioeconomic Resources Account for Racial and Ethnic Test Score Gaps?" *Future of Children*, Vol. 15, No. 1, Spring 2005.

assuming that effects of interventions measured more than 20 years ago would be experienced if schools adopted those interventions today.

Restrictive grade range for Beginning Reading studies

The letter raised concerns that the WWC's exclusion of research conducted on children outside of the kindergarten to 3rd grade age range was too restrictive. Because the reviews are focused on assessing interventions for beginning reading, the principal investigators, in coordination with the Institute of Education Sciences (IES), concluded that the review should focus on intervention effects for children in kindergarten through third grade (roughly ages 5 to 8). Some studies examine effects of interventions for students within the specified grade range and also students in higher grades. These studies are included in the review when the WWC is able to isolate the effects for the students who fall within the Beginning Reading grade range. The studies are excluded from the review if the results cannot be disaggregated to isolate effects for relevant grade range.

The grade range criterion is important for the integrity of the review process. The WWC-computed improvement indices and effect sizes are intended to reflect the effect of the intervention on the population in question. Including students above the topic area age range could lead to misstatements about intervention effects on children within the grade range.

It should be noted that the WWC attempts to determine a study's effects for the relevant grade range. When authors present findings aggregated across a broader age range and indicate that findings were analyzed for the relevant age range, it is standard WWC procedure to contact authors to request findings disaggregated for the grade range. Sometimes design limitations or other factors preclude authors from providing disaggregated results. In such cases, the WWC excludes the study.

Fidelity of treatment implementation

The letter notes that the WWC review process may downplay implementation fidelity. Definitions of implementation fidelity vary and many studies include little information to gauge fidelity, especially information about whether an intervention has been implemented within normal operating regimes of districts, schools, and teachers, not under specialized laboratory conditions. Moreover, there is no standard metric with which to rate and assess fidelity across studies that assures comparability.

The WWC's approach emphasizes the importance of replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies. Intervention reports include an "extent of evidence" classification that allows practitioners to place more weight if they choose on interventions for which the extent of evidence is large, meaning the results are drawn from multiple studies and a large number of classrooms and students.

Concerns about interventions and studies reviewed by the WWC

Reading Recovery

The letter expressed concern that Reading Recovery is an intervention outside of the Beginning Reading protocol. The Beginning Reading protocol states that interventions that target specific populations (for example, readers below grade level, and at-risk students) are eligible for the review. Reading Recovery is a short-term tutoring intervention program intended to serve the lowest achieving first-grade students (i.e., those in the bottom 20 percent). As such, it falls within the Beginning Reading protocol.

The letter expressed concerns that the reviews of studies of Reading Recovery mischaracterized the findings from those studies. For both the Baenen et al. (1997) and Iversen and Tunmer (1993) studies, the results presented by the WWC review represent the findings when the WWC standards and procedures are applied to these studies.

With respect to the Baenen et al. study, it is important to note that the beginning reading protocol prioritized one-year results. In effect, the Beginning Reading review is only intended to examine whether beginning reading interventions have an effect within one year. This one-year period is applied consistently to each study reviewed to ensure the results can be compared across studies and interventions. The Baenen et al. study's two- and three-year general reading achievement measures are not ignored, however; they are presented in Appendix A4.4 of the Reading Recovery Technical Appendices (http://ies.ed.gov/ncee/wwc/pdf/techappendix01_209.pdf).

With respect to the Iversen and Tunmer study, and consistent with the protocol, the WWC examined the results most relevant to the question of whether Reading Recovery improves reading proficiency compared to a reasonable counterfactual. That the study examined other comparisons is not ignored however. Appendices A4.1, A4.2 and A4.3 present results from other comparison groups. As with any study it reviews, the WWC does not base the findings of its review on the conclusions drawn by the authors.

We disagree that studies that compare an intervention to a no-treatment condition (as was done in Baenen et al.) provides a "built-in advantage," as the letter suggests. Many practitioners are interested in knowing whether an intervention is effective relative to customary classroom practices. The WWC reviews studies comparing treatments to no-treatment as well as studies comparing one treatment to another. In each case, the counterfactual is clearly documented in the review.

Exclusion of Reading Mastery Program

The letter expresses concern that the Reading Mastery Program is excluded from review by the WWC. The WWC has reviewed studies of Reading Mastery. A report summarizing the results of those reviews was published on the WWC website on August 12, 2008.

Inclusion of unpublished manuscripts for review

The letter suggests that by reviewing unpublished manuscripts, the WWC reviews studies of lower quality. Reviewing only studies that have passed a peer reviewing process could neglect unpublished information that may contain important and different findings. Mark Lipsey and David Wilson caution against this type of publication bias in *Practical Meta-Analysis* (2001). The damage such publication bias can cause in a systematic review is substantial. In *Knowing What Works in Healthcare*, the IOM notes that publication bias is well-established and that systematic reviews need to take steps to counter it, because otherwise "harmful interventions may appear to be worthwhile and beneficial interventions may appear to be useless" (page 97). For this reason WWC procedures are explicitly designed to review evidence from sources other than those published in journals. However, regardless of publication status, studies must meet the same evidence standards.

The letter asserts that the WWC should make available to the public all the research it reviewed. The more important issue is that the WWC provides an explicit citation to the study and thereby enables readers to obtain studies they wish to review. Unpublished information obtained from authors in response to queries that arise in reviews also are made available to requestors, and authors sign a form indicating that the information they provide to the WWC is available to the public.

Exclusion of Carlson and Francis (2002) study

The letter expressed concern over the WWC review's conclusion that there was a confound with the Direct Instruction intervention in Carlson and Francis (2002).[1] The WWC standard regarding intervention confounds was established to ensure that the results in a WWC review reflect what educators can expect if they implement the intervention being reviewed. A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program.

Exclusion of Waldron-Soler et al. (2002) study

As described in the protocols for the WWC Early Childhood Education and Beginning Reading reviews, to establish baseline equivalence, treatment and comparison groups must differ on the pretest measures by less than half a standard deviation, or the differences must be insignificant in an adequately powered statistical test. The Waldron-Soler (2002) study reported pretest differences exceeding half a standard deviation on at least two measures.

---

[1] The letter references Carlson and Freeman (2002). We have no record of a Carlson and Freeman (2002) study, and have assumed the letter refers to Carlson, C.D., & Francis, D.J. (2002). "Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE)." *Journal of Direct Instruction*, 3(1), 29-50.

The same standards were applied to the three Success for All studies you cite (Dianda and Flaherty, 1995; Ross, Alberg and McNelis, 1997; and Ross, McNelis, Lewis and Loomis, 1998). Unlike the Waldron-Soler study, pretest differences between the treatment and comparison groups were less than one half of a standard deviation.

Exclusion of Tobin (2003, 2004) studies

The Tobin (2003) study indicated that treatment and comparison groups were substantively different on pretest measures, with the differences exceeding half a standard deviation. Because the Tobin (2004) study follows the same students for additional time, it did not meet standards for the same reason.

The same standards were applied to the Success for All study cited in the letter (Smith, Ross, Faulks et al., 1993). Unlike the Tobin studies, pretest differences between the treatment and comparison groups were not statistically significantly different, and they were less than one half of a standard deviation. In keeping with WWC procedures, the WWC obtained information on pretest differences through communication with the study authors.

The "Create your own summary" feature

The letter expresses concern about the "create your own summary" feature of the WWC website because the results sort interventions by the magnitude of the improvement index. Because the improvement index calculations accurately reflect the application of the WWC evidence standards and effect size computations to the studies of those interventions, the WWC believes it is informative and useful to practitioners to sort results by improvement index.

The improvement index is one summary measure of the effect of beginning reading interventions. Users can also sort interventions alphabetically (by intervention name), by evidence rating, and by extent of evidence indices.

Appendix B
A Partial List of Studies of Reading Mastery and Its Precursors Completed before 1985

from Jean Stockard's 2008 Technical Report to NIFDI

Note: The letters in parentheses at the end of a citation refer to the source from which it was obtained. These sources are listed at the end of this appendix.

Apffel, J.A., J. Kelleher, M.S. Lilly, and R. Richardson. 1980. Developmental reading for moderately retarded children. Education and Training of the Mentally Retarded 10: 229-235. (b) (g)

Beck, I. L. and E. S. McCaslin. 1978. An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. LRDC Report No. 1978/6. Pittsburgh: University of Pittsburgh Learning Research and Development Center. (c)

Becker, W. C. and D. W. Carnine. 1980. Direct instruction: An effective approach to educational intervention with the disadvantaged and low performers. Pp. 429-473 in B.B. Lahey and A. E. Kazdin (eds.), Advances in Clinical Child Psychology, Volume 3, New York: Plenum. (a)

Becker, W. C. and S. Engelmann. 1976. Analysis of achievement data on six cohorts of low-income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model. Eric Document Reproduction Service No. ED 145922. (a)

Bock, G. and L. B. Stebbins. 1977. Education as experimentation: A planned variation model, Volume IV-B effects of Follow Through models. Cambridge, MA.: Abt Associates (ERIC Document Reproduction Service No. ED 148491). (a)

Booth, A., D. Hewitt, W. Jenkins, and A. Maggs. 1979. Making retarded children literate: A five-year study. The Australian Journal of Mental Retardation 5: 257-260. (d) (i)

Bowers, W. M. 1972. An evaluation of a pilot program in reading for culturally disadvantaged first grade students. Doctoral dissertation, University of Tulsa, 1972. (Eric Document Reproduction Service No. ED 073439). (a) (b)

Bracey, S., A. Maggs, and P. Morath. 1975. Effects of a direct phonic apporach in teaching reading with six moderately retarded children: Acquisition and mastery learning stages. Slow Learning Child 22: 83-90. (b) (d) (i)

Branwhite, A. B. 1983. Boosting reading skills by Direct Instruction. British Journal of Educational Psychology 53: 291-298. (b) (c) (d) (g) (i)

Carnine, D. 1977. Phonics versus look-say: Transfer to new words. The Reading Teacher 30: 636-640. (h)

Carnine, D. 1980. Phonic versus whole-word correction procedures following phonic instruction. Education and Treatment of Children 3: 323-329. (h)

Carnine, D. and R. Gersten. 1983. Effectiveness of Direct Instruction in teaching selected reading comprehension skills: Preliminary Draft. Paper presented at American Educational Research Association (April 11-15, 1983). (b)

DuPree, T. J. 1976. Brief history of Cherokee schools, 1804-1976. BIA Education Research Bulletin 4: 3-11. (ERIC Document Retrieval Service No. ED 127051). (a)

Gersten, R., W.C. Becker, T. J. Heiry, and W. A. T. White. 1984. Entry IQ and yearly academic growth of children in Direct Instruction programs: A longitudinal study of low SES children. Educational Evaluation and Policy Analysis 6: 109-121. (e)

Gersten, R. and A. Maggs. 1982. Teaching the general case to moderately retarded children: Evaluation of a five-year project. Analysis and Intervention in Developmental Disabilities 2: 329-343. (i)

Gersten, R., A. Brockway, and N. Henares. 1983. The Monterey DI program for students with limited English (ESL). Direct Instruction News, 2(4): 8-9. (a)

Gordan, M.B. (ed.) 1971. DISTAR instructional system: Summaries of case studies on the effectiveness of the DISTAR instructional system. Chicago: Science Research Associates. (a)

Gregory, R. P. and B. G. Warburton (1983). DISTAR Reading and remedial children in an infant school. School Psychology International 4: 169-172. (i)

Haring, N.G. and D. A. Krug. 1975. Evaluation of a program of systematic instructional procedures for extremely poor retarded children. American Journal of Mental Deficiency 79: 627-631. (b) (i)

Kastner, S. and M. Hollingshead. 1973. An evaluation of ESEA Title I Programs, Community School District 15. New York: New York University, Center for Educational Research and Field Services, School of Education. (ERIC Document Reproduction Service No. ED 087842). (a)

Kaufman, M. 1973. The effect of the DISTAR Instructional System: An evaluation of the 1972-1973 Title I program of Winthrop, Massachusetts. (ERIC Document Reproduction Service No. ED 110171). (a)

Kaufman, M. 1974. The effect of the DISTAR Instructional System: An evaluation of the 1973-1974 Title I program of Winthrop, Massachusetts. (ERIC Document Reproduction Service No. ED 110170). (a)

Lewis, A. 1982. An experimental evaluation of direct instruction programme with remedial readers in a comprehensive school. Educational Psychology 2: 121-135. (g)

Lloyd, J., D. Cullinan, E.D. Heins, and M.H. Epstein. 1980. Direct instruction: Effects on oral and written language comprehension. Learning Disabilities Quarterly 3: 70-76. (g)

McCabe, T.A. 1974. The DISTAR Reading and Language Program: Study of its effectiveness as a method for the initial teaching of reading. Doctoral dissertation, University of Massachusetts, 1974. (ERIC Document Retrieval Service No. ED 102498).

Meyer, L. A. 1983. Long-term academic effects of Direct Instruction Follow Through. Technical report No. 299. Champaign: University of Illinois at Urbana-Champaign, Center for the Study of Reading. (ERIC Document Retrieval Service No. 237932). (a)

Newark (NJ) Board of Education. 1974. Program to improve the informational processoing of children with reading and learning problems. ERIC 106 826. (b)

Ogletree, E. J. 1976. A comparative study of the effectiveness of DISTAR and eclectic reading methods for inner-city children. Chicago: Chicago State University. (ERIC Document Reproduction Service No. ED 146544). (a) (b)

Ogletree, E.J. 1977. Does DISTAR meet the reading needs of inner-city kindergarten pupils? ERIC 146 303. (b)

Rawl, R. K. and F. S. O'Tuel. 1982. A comparison of three prereading approaches for kindergarten students. Reading Improvement 19: 205-211. (a) (b)

Richardson, E., B. Dibenedetto, A. Christ, M. Press, and B. Winsbert. 1978. An assessment of two methods for remediating reading deficiencies. Reading Improvement 15: 82-95. (a) (b) (d) (g) (i)

Sassenrath, J. M. and R. E. Maddux. 1974. Language instruction, background, and development of disadvantaged kindergarten children. California Journal of Educational Research 25: 61-68. (a)

Serwer, B. L., B. J. Shapiro, and P. P. Shapiro. 1973. Comparative effectiveness of four methods of instruction on the achievement of children with specific learning disabilities. Journal of Special Education 7: 241-249. (b) (d)

Singer, B. 1973. The effects of structured instruction on kindergarten pupils: Final report. Washington, D.C.: U.S. Office of Education. (ERIC Document Reproduction Service No. ED 087564). (a)

Stebbins, L. B., R.G. Pierre, E. C. Proper, R. B. Anderson, and T. R. Cerva. 1976. Education as experimentation: A planned variation model (Vols. 3A-3B). Cambridge, MA: Abt Associates. (ERIC Document Reproduction Service No. ED 148489). (a)

Stein, C. L. and J. Goldman. 1980. Beginning reading instruction for children with minimal brain dysfunction. Journal of Learning Disabilities 13: 52-55. (b) (g) (i)

Summerell, S. and G. G. Brannigan. 1977. Comparison of reading programs for children with low levels of reading readiness. Perceptual and Motor Skills 44: 743-746. (a) (b) (d) (g)

Williamson, F. 1970. DISTAR reading: Research and experiment. (ERIC Document Reproduction Service No. ED 045318). (a) (b)

Note: Letters in parentheses refer to the source(s) providing the citation.

a. Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.
b. Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.
c. McGraw-Hill, 2002. "Appendix: Research supporting instruction in Reading Mastery: A Selected Annotated Bibliography." Results with Reading Mastery. ERIC ED 469 523.
d. Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.
e. Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.
f. NIFDI research staff found this citation independently.
g. Grossen, Bonnie. n.d. "References for Reading Mastery Reading Review," The Research Base for Reading Mastery, SRA, Eugene, Oregon: University of Oregon. (http://darkwing.uoregon.edu/~adiep/rmref.htm).
h. National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washing ton, D.C.: National Institute of Child Health and Human Development.
i. Kinder, D., R. Kubina, N.E. Marchand-Martella. 2005. Special education and Direct Instruction: An effective combination. Journal of Direct Instruction 5: 1-36.

Appendix C
Studies of Reading Mastery Published 1985 or Later But Not Included in the WWC Review of Beginning Reading

from Jean Stockard's 2008 Technical Report to NIFDI

Note: The letters in parentheses at the end of a citation refer to the source from which it was obtained. These sources are listed at the end of this appendix.

Ball, E. and B. Blachman. 1991. Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? Reading Research Quarterly 26: 49-66. (d)

Bateman, B. 1991. Teaching word recognition to slow-learning children. Reading, Writing, and Learning Disabilities 7: 1-16. (b)

Brent, G. and N. DiObilda. 1993. Effects of curriculum alignment versus direct instruction on urban children. Journal of Educational Research 86: 333-338. (a)

Brett, A., L. Rothlein, and M. Hurley. 1996. Vocabulary acquisition from listening to stories and explanation of target words. Elementary School Journal 96: 415-422. (d)

Bryne, B. and R. Fielding-Barnsley. 1991. Evaluation of a program to teach phonemic awareness to young children. Journal of Educational Psychology 83: 451-455. (d)

Butler, P.A. 2003. Achievement outcomes in Baltimore City Schools. Journal of Education for Students Placed at Risk 8: 33-60. (e)

Chamberlain, L.A. 1987. Using DI in a Victoria, B.C. resource room. ADI News 7 (1): 7-8. (i)

Cole, K.N. Dale, P.S. and P. E. Mills. 1991. Individual differences in language delayed children's responses to direct and interactive preschool instruction. Topics in Early Childhood Special Education 11: 99-124. (i)

Cole, K. N., P.S. Dale, P.E. Mills, and J. R. Jenkins. 1993. Interaction between early intervention curricula and student characteristics. Exceptional Children 60: 17-28. (i)

Cooke, N.L., S.L. Gibbs, M. L. Campbell, and S. L. Shalvis. 2004. A comparison of Reading Mastery Fast Cycle and Horizons Fast Track A-B on the reading achievement of students with mild disabilities. Journal of Direct Instruction 4: 139-151. (i)

Cunningham, A. 1990. Explicit versus implicit instruction in phonemic awareness. Journal of Experimental Child Psychology 50 :249-444. (d)

Dale, P.S. and K.N. Cole. 1988. Comparison of academic and cognitive programs for young handicapped children. Exceptional Children 54: 439-447. (i)

Darch, C. and E. Kameenui. 1987. Teaching critical reading skills to learning disabled children. Learning Disability Quarterly 10: 82-92. (g)

Diobilda, N. and G. Brent. 1986. Direct Instruction in an urban school system. Reading Instruction Journal 29: 2-5. (e)

Dowdell, T. 1996. The effectiveness of Direct Instruction on the reading achievement of sixth graders. (ERIC Document Retrieval Service No. ED 396268). (a) (d)

Fredrick, L.D., M.C. Keel, and J. H. Neel. 2002. Making the most of instructional time: Teaching reading at an accelerated rate to students at risk. Journal of Direct Instruction 2: 57-63. (f)

Gersten, R. and D. Carnine. 1986. Direct Instruction in reading comprehension. Educational Leadership 44: 68-78. (b)

Gersten, R., C. Darch, and M. Gleason. 1988. Effectiveness of a Direct Instruction academic kindergarten for low-income students. The Elementary School Journal 89: 227-240. (e) (h)

Gersten, R., T. Keating, and W. Becker. 1988. The continued impact of the direct instruction model: Longitudinal studies of follow through students. Education and Treatment of Children 11: 318-327. (h)

Gunn, B., A. Biglan, K. Smolkowski, and D. Ary. 2000. The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. Journal of Special Education 2: 90-103. (c) (d)

Haskell, D. W., B. R. Foorman, and P. R. Swank. 1992. Effects of three orthographic/phonological units on first-grade reading. Remedial and Special Education 13: 40-49. (c) (d)

Herrera, J. A., C. H. Logan, P. G. Cooker, D. P. Morris, and D. E. Lyman. 1997. Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children. Learning disability or learning difference? Reading Improvement 34: 71-89. (d)

Jenkins, J. R., B. Matlock, and T. A. Slocum. 1989. Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. Reading Research Quarterly 24: 215-235. (d)

Juel, C. 1988. Learning to read and write: A longitudinal study of 54 children from first through fourth grade. Journal of Educational Psychology 80: 437-447. (d)

Juel, C. and C. Minden-Cupp. 2000. Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly 35: 458-492. (d)

Kamps, D.M., H.P. Wills, C.R. Greenwood, S. Thorne, J.F. Lazo, J.L. Crockett, J.M. Akers, and B. OL. Swaggart. 2003. Curriculum influences on growth in early reading fluency for students with academic and behavioral risk. Journal of Emotional and Behavioral Disorders 11: 211-224. (f)

Kaiser, S. K. Palumbo, R.C. Bialozor, and T. F. McLaughlin. 1989. Effects of Direct Instruction with rural remedial students: A brief report. Reading Improvement 26: 88-93. (b) (c) (g)

Keel, M.C., L. D. Frederick, T. A. Hughes, and S.H. Owens. 1999. Using paraprofessionals to deliver Direct Instruction reading programs. Effective School Practices 18: 16-22. (e)

Kuder, S.J. 1990. Effectiveness of the DISTAR reading program for children with learning disabilities. Journal of Learning Disabilities 23: 69-71. (b) (d) (i)

Kuder, S. J. 1991. Language abilities and progress in a Direct Instruction reading program for students with learning disabilities. Journal of Learning Disabilities 24: 124-127. (b) (d)

Leach, D. and S. Siddall. 1990. Parental involvement in the teaching of reading: A comparison of hearing reading, paired reading, pause, prompt, praise, and direct instruction methods. British Journal of Educational Psychology 60: 349-355. (h)

Ligas, M. R. and D. W. Vaughan. 1999. Alliance of quality schools: 1998-99 evaluation report. Broward , Florida: Broward County Schools. (a)

MacIver, M.A., E. Kedmper, and S. Stringfield. 2000. The Baltimore Curriculum Project: Fourth year report. Report for the Abell Foundation. Baltimore: Johns Hopkins University, Center for Social Organization of Schools. (a)

Marston, D., S. Deno, D. Kim, K. Diment, and D. Rogers. 1995. Comparison of reading intervention approaches for students with mild disabilities. Exceptional Children 62: 20-37. (h) (i)

Mathes, P.G. and T.J. Proctor. 1988. Direct Instruction for teaching “hard to teach” students. Reading Improvement 25: 92-97. (b)

Neely, M. 1995. The multiple effects of whole language, precision teaching, and Direct Instruction on first-grade story-reading. Effective School Practices 14: 33-42. (d)

O’Connor, R. E. and J. R. Jenkins. 1995. Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. The Journal of Special Education 29: 255-275. (i)

O'Connor, R., J. Jenkins, K. N. Cole, and P.E. Mills. 1993. Two approaches to reading instruction with children with disabilities: Does program design make a difference? Exceptional Children 59: 312-323. (b) (d) (i)

Rawl, R. K. and F. O'Tuel. 1982. A comparison of three prereading approaches for kindergarten students. Reading Improvement 19: 205-211. (d)

Reutzel, D. R. and P. M. Hollingsworthy. 1993. Effects of fluency training on second graders' reading comprehension. Journal of Educational Research 86: 325-331. (d)

Ross, S., J. Nunnery, E. Goldfeder, A. J. McDonald, R. Rachor, M. Hornbeck, et al. 2004. Using school reform models to improve reading achievement: A longitudinal study of Direct Instruction and Success for All in an urban district. Journal of Education for Students Placed at Risk 9: 357-389. (e)

Salerno, C. 1992. A comparison of classrooms using a meaning-centered approach and a code-centered approach. ADI News 11 (2): 26-29. (d)

Sexton, C. W. 1989. Effectiveness of the DISTAR Reading I Program in developing first graders' language skills. Journal of Educational Research 82: 289-293. (a) (b) (d) (g)

Smith, S., D. Simmons, M. Gleason, E. Kameenui, S. Baker, M. Sprick, et al. 2001. An analysis of phonological awareness instruction in four kindergarten basal reading programs. Reading and Writing Quarterly 17: 25-50. (d)

Snider, V. E. 1990. Direct Instruction reading with average first-graders. Reading Improvement 27: 143-148. (b) (g) (h)

Stein, M. 1990. *Reading research ... and Reading Mastery*. Chicago: SRA. (b)

Tobin, K. 2000. A comparison between Horizons Fast Track A-B and Silver, Burdett, and Ginn reading curricula in first grade: June 2000 final report. Oregon: National Institute for Direct Instruction. (a) (e)

Torgesen, J., S. Morgan, and C. Davis. 1992. Effects of two types of phonological awareness training on word learning in kindergarten children. Journal of Educational Psychology 84: 364-370. (d)

Traweek, D. and Berninger, V. 1997. Comparisons of beginning literacy programs: Alternative paths to the same learning outcome. Learning Disability Quarterly 20: 160-168. (a) (d) (h)

Umbach, B. T., C. Darch, and G. Halpin. 1987. Teaching reading to low performing first graders: A comparison of two instructional approaches. ERIC 290 130. (b)

Umbach, B., C. Darch, and G. Halpin. 1989. Teaching reading to low performing first graders in rural schools: A comparison of two instructional approaches. Journal of Instructional Psychology 16: 23-30. (h)

Urdegar, S. M. 1998. Evaluation of the Success for All Program, 1997-98. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research. (a)

Urdegar,k S. M. 2000. Evaluation of the Success for All Program, 1998-99. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research. (a)

Weinstein, G. and N. L. Cooke. 1992. The effects of two repeated reading interventions on generalization of fluency. Learning Disability Quarterly 15: 21-28. (d)

Varela-Russo, C., K. A. Blasik, and M. Ligas. 1998. Alliance of quality schools evaluation report. Ft. Lauderdale, FL: School Board of Broward County. (a)

Yu, L. and R. Rachor. 2000. The two-year evaluation of the three-year Direct Instruction program, in an urban public school system. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 441831). (a) (e)

Note: Letters in parentheses refer to the source(s) providing the citation.
a. Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.
b. Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.
c. McGraw-Hill, 2002. "Appendix: Research supporting instruction in Reading Mastery: A Selected Annotated Bibliography." Results with Reading Mastery. ERIC ED 469 523.
d. Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.
e. Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.
f. NIFDI research staff found this citation independently.
g. Grossen, Bonnie. n.d. "References for Reading Mastery Reading Review," The Research Base for Reading Mastery, SRA, Eugene, Oregon: University of Oregon. (http://darkwing.uoregon.edu/~adiep/rmref.htm).

h. National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washing ton, D.C.: National Institute of Child Health and Human Development.
i. Kinder, D., R. Kubina, N.E. Marchand-Martella. 2005. Special education and Direct Instruction: An effective combination. Journal of Direct Instruction 5: 1-36.

Subject: RE: Machinations of What Works Clearinghouse

-----Original Message-----
From: zig@nifdi.org [mailto:zig@nifdi.org]
Sent: Tuesday, October 07, 2008 7:46 PM
To: Paul Decker
Cc: ZZ HSC Anita Summers
Subject: Machinations of What Works Clearinghouse

The attached paper documents the incredible inadequacies of WWC (both its logic and its execution of what it claims it does). I don't engage in political games, but trust me, I am going to go after What Works with concerted energy, and I am going to take steps to see that the measurement
community and the extended scientific community respond to the kind of casuistry you defend in your letter of September 8, 2008.

Siegfried Engelmann

P.S. The paper is on my website: zigsite.com. Hopefully it will be on others
soon.

The What Works Clearinghouse Beginning Reading Reports and Rating of Reading Mastery: An Evaluation and Comment

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

September 2008

Technical Report 2008-04



© 2008 National Institute for Direct Instruction
PO Box 11248 • Eugene, OR 97440
1-877-485-1973

The What Works Clearinghouse Beginning Reading Reports and Rating of Reading Mastery: An Evaluation and Comment

NIFDI Technical Report 2008-4
September, 2008
Executive Summary

The What Works Clearinghouse is a federally funded program established in 2002 that evaluates educational interventions on the basis of the "rigor of research evidence" and provides summary ratings on its website. Ratings have appeared only since 2007, and some of the work of the Clearinghouse has already been subjected to harsh criticism. This paper adds to the body of critiques.

Direct Instruction's Reading Mastery curriculum is a comprehensive core reading program that has been found, in decades of research, to be highly effective. Yet, the WWC's recent review of Reading Mastery concluded that no studies met their evidence standards and they were "unable to draw any conclusions based on research" regarding its effectiveness.

A careful examination of the research cited by the WWC found that this conclusion was not well founded. The WWC failed to examine close to 100 research studies that were cited in other well-known reviews of the literature. In addition, reviews of studies that were included were flawed. For instance, one large, well-regarded study was rejected because training of teachers to use the curriculum was considered a "confound" to the use of the curriculum itself. Other studies were dismissed because the treatment (Reading Mastery) and control groups were unequal at baseline. Yet, these differences were statistically controlled and, in two cases, the treatment groups had lower pretest scores but equal and higher posttest scores.

Examination of reviews of other curricula indicates that at least some of the material posted on the website is misleading. Summaries do not reflect the conclusions of the articles and, in at least one case, are exactly opposite to the published conclusions.

Attempts have been made to communicate these concerns and recommendations for change to the WWC and Mathematica Policy Research, Inc., which administers the Clearinghouse. These concerns and suggestions, which parallel those of other researchers, have been rejected.

Given the very serious problems in the review procedures and policies, the NIFDI research staff suggests extreme caution in the use of any ratings or conclusions from the WWC.

Table of Contents

	Page
Executive Summary	2
The What Works Clearinghouse	4
Direct Instruction, Reading Mastery, and the WWC	5
The WWC Ignored a Significant Body of Research on Reading Mastery	6
The WWC Failed to Review Many Existing Studies of Reading Mastery	8
Inaccuracies in the Reviews of Reading Mastery Studies	10
Summary	12
Inaccuracies in Other WWC Reports	12
Recommendations	14
References	16
Appendix A: Communications with the What Works Clearinghouse and Mathematica	19
Appendix B: A Partial List of Studies of Reading Mastery and its Precursors Completed Before 1985	
Appendix C: Studies of Reading Mastery published 1985 or Later But Not Included In the WWC Review	
Appendix D: The Impact on Summary Ratings of Ignoring Fidelity of Implementation	

The What Works Clearinghouse Beginning Reading Reports and Rating of Reading Mastery: An Evaluation and Comment

This document discusses serious concerns with the policies, procedures and judgments of the federally funded What Works Clearinghouse, focusing both on a recent Clearinghouse review of the Direct Instruction program Reading Mastery as well as more general issues regarding their procedures and conclusions. The report begins with a brief description of the stated mission of the WWC and criticisms of their procedures. The second section provides a detailed analysis of the WWC's recent report on Reading Mastery, describing the extensive extant body of scientific research on the program and serious errors within the WWC report. The third section describes more general concerns with the work of the WWC and misrepresentations of research on one other curricular program, and the final section provides recommendations for changes in WWC procedures. Appendices provide supporting material.

The What Works Clearinghouse

The What Works Clearinghouse (WWC) was established in 2002 through a grant from the U.S. Department of Education's Institute of Education Sciences. Billing itself as "a central and trusted source of scientific evidence for what works in education," the Clearinghouse reviews studies of educational interventions, assessing the "rigor of research evidence" and "giving educators the tools to make informed decisions" (http://ies.ed.gov/ncee/wwc/aboutus/). The WWC has focused its work on analyzing individual curricula, giving its highest rankings to generally small, but tightly controlled, experimental studies on the impact of different curricula.[1]

After a very slow start, the Clearinghouse began issuing reviews in 2007, periodically posting summary judgments of individual curricular programs on its website. The assessments appear to be internally developed by WWC staff with no external oversight of the process or results. Unlike most of the scientific literature, the assessments are not exposed to peer review or other independent judgment by experts in the field. Not surprisingly, the work of the WWC has faced extensive criticism for its procedures and methodology including the ways in which studies are identified, the manner in which summary rankings are constructed, the accuracy of the reviews that have been produced, and the absence of any outside or independent oversight (e.g. McArthur 2008, Slavin, 2008).

The author of this document queried the WWC regarding its practices in e-mail communications from December, 2007 through April, 2008. Based on the responses received to those queries and further examination of WWC materials, a detailed letter was sent to Mathematica, the organization responsible for administering the Clearinginghouse on June 25, 2008. This letter summarized concerns with the WWC's procedures and policies. A reply was received from Mathematica on September 8, with a cover letter from the CEO of Mathematica and a more detailed report, apparently prepared by Mark Dynarski, the project director of the

[1] This focus is somewhat surprising given the emphasis within the educational community and research literature on the importance of comprehensive changes within schools to produce long-lasting changes in achievement. In addition, the WWC's "Statement of work" claims that the goal of the clearinghouse was to develop reviews of "educational interventions," parenthetically defined as "programs, products, practices, and policies" (WWC Statement of Work, Section I.B., paragraph 1).

WWC. All of these communications are included in Appendix A and are occasionally referenced in this document.

Unfortunately, none of the information that has been provided by the WWC or Mathematica has eased the concerns with the WWC's procedures that were outlined in the original queries. The material below summarizes some of these concerns and indicates why the conclusions of the WWC, both with respect to Reading Mastery and to other instructional programs, should be seriously questioned.

Direct Instruction, Reading Mastery and the WWC

Numerous studies have demonstrated the effectiveness of Direct Instruction with a wide variety of students in very different settings. Meta-analyses that examine large bodies of research literature also consistently point to the superiority of Direct Instruction. For instance, a recent meta-analysis of the achievement effects of different models (Borman et al 2003) found that Direct Instruction was the most researched of any of those studied and had the most research conducted by those not affiliated with the developers. As they put it, "the research base for Direct Instruction is very extensive and of very good quality" (p. 187). Among the many different models used in their analysis, Direct Instruction had the strongest evidence of effects on student achievement. Numerous other analyses have reached similar conclusions (e.g. Adams and Engelmann 1996, AFT 1998, Beck and McCaslin 1978, Herman, et al 1999, Kennedy 1978, Schieffer et al 2002). Additional studies expound upon the importance of the principles that underlie Direct Instruction, especially systematic and explicit instruction (see Adams 1990, Anderson et al 1985, Baker 1994, Bond and Dykstra 1967, Chall 1967, Foorman 1995, Fukkink and deGlopper 1998, Grossen 1997, ICHHD 2000, Juel and Minden-Cupp 2000, National Reading Panel 2000, NICHHD 1996, Pflaum, et al 1980, Smith et al 2001, Snider 1990, Snow et al 1998, Stanovich 1994). Finally, Direct Instruction has been lauded as not just the purveyor of effective curricular materials, but also as an important model of comprehensive school reform, guiding the transformation of schools into effective organizations in which all students can have high achievement (e.g. Borman 2002, Buechler 2002, CSRQ 2006).

Direct Instruction's Reading Mastery curriculum is a comprehensive core reading program with multiple levels and is the most widely used program within the DI curricular library. It has long been acclaimed as highly effective, with strong research accumulated over decades of work, supporting its development and documenting its success.

Despite the large body of research on Direct Instruction and the various associated curricular programs, Reading Mastery was not included in the initial WWC review process for Beginning Reading. Through mid-summer, 2008, the WWC website simply stated, without further explanation, that they chose not to include Reading Mastery in their review of Beginning Reading material.[2] In mid-August, 2008, however, the WWC issued a report on Reading Mastery with the following conclusion:

[2] In July, 2007, the WWC posted a review of Corrective Reading on its Beginning Reading page. Corrective Reading is a Direct Instruction program designed for children in the upper elementary grades who are reading below grade level. No explanation was given for why this program was included in the Beginning Reading category when its target audience did not match either the intended grade level or the general classroom focus of the review. While the review accepted one study of Corrective Reading as meeting its criteria and determined that it showed

> No studies of Reading Mastery that fell within the scope of the Beginning Reading review meet WWC evidence standards. The lack of studies meeting WWC evidence standards means that, at this time, the WWC is unable to draw any conclusions based on research about the effectiveness or ineffectiveness of Reading Mastery.[3]

To support their conclusion the WWC report lists 61 citations on Reading Mastery that were examined and a summary of the reasons they did not meet the WWC's standards of evidence.

Because the WWC conclusion regarding Reading Mastery contrasts so starkly with previous findings of the scientific community, the research staff of the National Institute for Direct Instruction (NIFDI) carefully examined the listed works and compared them to the extant literature and other reviews. As detailed below, our examination uncovered numerous examples of faulty judgments by the WWC. These involve three major points, all of which are elaborated below: 1) the WWC report ignored a large body of the extant research through seemingly arbitrary review decisions, 2) the WWC report did not find or review a significant proportion of the research on Reading Mastery that should have met their review criteria, and 3) the few reviews that were conducted often misconstrued and misinterpreted the research evidence.

Based on our review, others' critiques of the work of the WWC, and the very extensive research base that has documented the effectiveness of Direct Instruction and Reading Mastery, we conclude that the WWC statement regarding Reading Mastery is inaccurate and misleading. A large body of literature points to the very strong effect of Direct Instruction and Reading Mastery on children's reading achievement.

The WWC Ignored a Significant Body of Research on Reading Mastery

The WWC limited its review of studies for the Beginning Reading category to those that involve students in general education in grades K-3 and that were published between 1985 and 2007. When NIFDI's Director of Research questioned the decision to ignore studies published before 1985, the WWC responded that the "cut-off date of 20 years is a parameter set to ensure the research reviewed is most relevant to classrooms as they operate today" (e-mail correspondence, April 9, 2008, see Appendix A). The report included with the September 8 letter from Mathematica expands upon this reasoning by suggesting that numerous other changes over time such as increased levels of reading readiness (supposedly as a result of increased preschool enrollment), "stronger training" of teachers and the use of "newer curricula" "could have implications for the effectiveness of an intervention" (Dynarski, page 2). The exposition in the

---

"potentially positive effects," the issues raised in this document regarding WWC's review of Reading Mastery also apply to its review of Corrective Reading. One study of Reading Mastery was also included in the ELL portion of the WWC webpage, but again this placement does not reflect the target audience of Reading Mastery. It should be noted, however, that the high ratings these studies of both CR and RM received, even with these alternative audiences, should be seen as further evidence of the quality of the programs.

[3] The WWC Reading Mastery report, including the studies they reviewed, can be found at http://ies.ed.gov/ncee/wwc/pdf/WWC_ReadingMastery_081208.pdf. Interestingly, the September, 2008 letter from Mathematica (Dynarski report, page 4) notes the addition of the RM review to the website on August 12, 2008. It is unclear to what extent the review was prompted by our letter of June 25, 2008.

correspondence pointedly excludes any mention of empirical data to support the causal linkage that is suggested.[4]

We have not been able to find any studies cited on the WWC website, or elsewhere, that document that children's learning styles, classrooms, teachers' actions, or schools' organizational structures have drastically changed since 1985. As noted above, an extensive body of research cites the importance of systemic and explicit instruction in promoting achievement. This research spans the last 4 decades and has produced consistent results, no matter when the research was conducted (see Adams 1990, Anderson et al 1985, Baker 1994, Bond and Dykstra 1967, Chall 1967, Foorman 1995, Fukkink and deGlopper 1998, Grossen 1997, ICHHD 2000, Juel and Minden-Cupp 2000, National Reading Panel 2000, NICHHD 1996, Pflaum, et al 1980, Smith et al 2001, Snider 1990, Snow et al 1998, Stanovich 1994). In all the other areas of scientific study with which we are familiar, such as medicine, psychology, or the other social sciences, the full range of research articles are used to formulate policy, and there is no arbitrary date for inclusion or exclusion of results. If changes over time are hypothesized, the possibility of such changes is subjected to empirical test. Many analyses of the research literature systematically examine variations in results by the time when the study occurred.

This limitation of the time period for review had a disproportionately large impact on the number of studies of Direct Instruction included in the WWC's scope of study compared to other interventions precisely because DI has such a long and strong research base. A strong body of seminal research on DI was conducted before 1985, including Project Follow Through, the largest educational experiment in the history of the United States.[5] In general, research on interventions tends to be most active soon after the creation of the program. Because Reading Mastery was developed in the 1960s and high quality research conducted at that time demonstrated its effectiveness with a wide range of students, researchers have had little incentive to conduct additional studies. Given the consistently positive results from earlier decades, the current recent research focus has moved to populations of students with more extensive needs, such as older children who had not yet learned to read.[6]

A significant amount of research on Reading Mastery was ignored by the WWC because of the seemingly arbitrary choice of 1985 for a cut-off. Appendix B of this report includes a partial list of studies of Reading Mastery and its precursors that were completed before 1985.[7] The list was compiled by reviewing published literature reviews

---

[4] The citations provided are to macro-level data regarding increased preschool and kindergarten enrollment as well as to a study regarding the relationship of race and socio-economic status to achievement. None relate, either directly or indirectly, to variations over time in the impact of curricula or to changes over time in children's learning styles or achievement.

[5] Project Follow Through employed a strong experimental design and an extremely large sample to compare a wide variety of elementary curricula. The results provided very strong evidence of the superiority of Direct Instruction in raising academic achievement as well as students' self image.

[6] Studies of children in special education should have been within the purview of the review. In fact, the description of "populations to be included" in the WWC's Beginning Reading protocol states that "because students with learning disabilities…lag behind the population as a whole in reading achievement, studies involving these groups are of particular interest." No mention is made of excluding studies that involve students in special education.

[7] The cut-off date of 1985 used for the Reading Mastery report differs from the general WWC protocol for the Beginning Reading studies, which uses 1983 as the cut-off date. We have found no explanation for this discrepancy

and meta-analyses, all of which should have been available to the WWC staff. The list includes 38 items.[8] It should be stressed that this list is not exhaustive, but simply relied on easily available meta-analyses and reviews that should have been accessible to the WWC staff.

In short, unless the WWC can provide scientific evidence to the contrary, it would appear reasonable to suggest that high quality studies conducted before 1985 are still relevant today and should be included in any evaluation of Reading Mastery. Any conclusion regarding the effectiveness of Reading Mastery that does not consider this part of the literature should be challenged as incomplete and inaccurate.

The WWC Failed to Review Many Existing Studies of Reading Mastery

In addition to excluding studies published prior to 1985, the WWC review omitted a large number of studies published after 1985 that could have been included. Appendix C provides a partial list of these studies and includes 56 different citations. Like Appendix B, this list was largely generated by examining earlier reviews that should have been available to the WWC staff. Although the WWC lists a methodology for retrieving studies, the fact that they could miss so many citations that were in already published reviews suggests serious difficulties with their selection criteria.

As with the vast majority of work in this area, the studies listed in Appendix C generally found that Reading Mastery was superior to other approaches. For instance, an article by Kamps and associates (2003) uses multi-level analytic techniques to examine data from schools that implemented different reading curricula and found that students in Reading Mastery had significantly stronger achievement gains than in any of the other programs, both for those with and without social behavioral risks. Similarly, Gunn and associates (2000) report a study in which students were randomly assigned to interventions and found that both Hispanic and non-Hispanic students who received instruction in Reading Mastery had significantly more improvement over time than other students.

It should also be noted that the WWC list of writings on Reading Mastery includes material that would generally not be included in summaries of the scientific literature. Many items on the list are not research reports and were never intended as such. A large proportion is reports by school principals, teachers, and/or superintendents of the success their schools have experienced with Direct Instruction. These testimonials appeared in publications of the Association for Supervision and Curriculum Development and Council of Chief State School Officers (accounting for 8 of the items on the list of 61 studies) and on the website of SRA, the publisher of DI materials (26 of the items).[9] Two other studies on the list clearly targeted students outside the stated age range, with the titles indicating that they looked at "ninth and tenth graders" (Airhart 2005) or at "middle

[8] A large number of these studies examined DISTAR, the name by which the content now in Reading Mastery, was once known.

[9] Most of the testimonials include results from norm referenced achievement tests, often involving data from several years. These could have been used, if the WWC desired, to compare the achievement in these schools with national norms.

school students" (Shippen et al 2006). An additional study (Flores and Ganz 2007) did not look at Reading Mastery, but at Corrective Reading, a different Direct Instruction curricular program. Three studies in the list did not have reading achievement as a dependent measure, but instead were content analyses of textbook stories (Jordan 2005), a qualitative study of three students' understandings of what they read (Wilson 2005), and an examination of students' social behaviors (Smolkowski et al 2005). Finally, a few others appeared to have never been subjected to the usual scientific standards of peer review, such as articles posted on websites, papers given at professional meetings, and unpublished dissertations (6 studies in all).[10, 11]

When these 46 studies are omitted from the list, the WWC actually examined only 15 studies of Reading Mastery that would generally be considered part of the scientific research base. This number is less than one-fourth of the number given in the report. The listing of materials in the WWC's report on Reading Mastery implies that the research base that was reviewed was much larger than it, in fact, was.

In general, it appears that the WWC list of materials that were reviewed suffers from errors of both exclusion and inclusion. A large number of studies appear to have been ignored. Our review of already existing reviews of the literature revealed almost one hundred articles that could have been consulted if a full analysis were to occur. These citations were easily found through simply searching the reference list of already published literature reviews and thus are no doubt an undercount of the available resources. The WWC review missed a large proportion of the research base, not just through the time parameters that were used but also, apparently, through a search of the literature that was far from thorough.[12] At the same time, much of the material listed in their report should have fallen outside the domain of a comprehensive review of the

[10]Meta-analyses often include such unpublished works in their reviews, and that has been the case with several meta-analyses of beginning reading. Given that the WWC review is not subject to peer review, omitting these studies from their work could be one way to add additional controls. If the WWC's work were subject to peer review, including unpublished work would be less problematic. Dynarski's report included with the September, 2008, letter addresses the issue of peer review, citing a study of publication bias in medical research. Whether or not the result with medical research applies to educational research is an empirical question and a citation to the education or more general social science literature would be more appropriate. We know of no study that has found that high quality social science work remains unpublished. In any case, the standard practice in meta-analyses and other reviews is to treat the issue of the relationship of publication status to results as an empirical question and compare the unpublished and published articles.

[11] The WWC categorized one of the studies that had not been subjected to peer review (CSRQ 2006) as "ineligible…because it does not examine the effectiveness of an intervention." This study, prepared by the highly regarded Comprehensive School Reform Quality Center, reviewed and summarized the evidence on the effectiveness of 22 different school reform models, including Direct Instruction. As with other such reviews, there were more studies available regarding the effectiveness of Direct Instruction than for any other model and the results indicated that Direct Instruction was consistently one of the most effective, if not the most effective, in each of the categories examined. Thus, while the CSRQ document did not address the effectiveness of a single intervention, to dismiss it because it examined numerous models could be seen as misleading.

[12] Ironically, one of the literature reviews used to generate the listings in Appendices B and C was the 2006 report of the Comprehensive School Reform Quality Center (source e in the listings) and discussed in footnote 10 above. The WWC lists this report as one that they consulted and found "ineligible for review," but apparently did not use its references for further examination.

scientific literature, providing a misleading representation of the amount of literature that was examined.

Inaccuracies in the Reviews of Reading Mastery Studies

The WWC report provides one sentence summaries of the reason that a study was judged not to meet their established "evidence standards." These reasons involved either questions about the comparability of the control and experimental groups or the existence of a possible "confound" to the treatment. Examples are given below of the studies rejected for these reasons and explanations of the faulty logic involved in the conclusions. The errors in these examples are so serious that they should raise serious concerns regarding the WWC's evaluation procedures.

Equivalence of Experimental and Control Groups – Three of the published studies examined by the WWC that addressed student achievement were rejected because, although they used a "quasi-experimental design" with both an intervention and a control group, they did "not establish that the comparison group was comparable to the treatment group prior to the start of the intervention."[13] The NIFDI research staff carefully examined these three articles and found that, in each case, the conclusion of the WWC could be questioned. In one of the studies (Brent et al, 1986) the presentation of the results clearly noted that adjustments had been made, using standard multivariate procedures, for any possible pretest differences between the treatment and control conditions. In the two other studies (O'Brien and Ware 2002, Thomson 1991), the pretest scores of the treatment group (those receiving Reading Mastery) were lower than the scores of students in the control group. Such a situation is routinely considered as an indication of a conservative test, for it biases results against the treatment, and is not used as a reason to automatically discount the results. Again, supporting the large body of extant evidence, these three studies all showed positive effects of Reading Mastery on students' achievement.

Confounding Factors – Nine studies were rejected from the WWC's review because, although they used a quasi-experimental design, the reviewers determined that there was "a confounding factor, such as combining with other interventions, which makes it impossible to attribute the observed effect solely to Reading Mastery." Unfortunately, the WWC provided no specific details in the report published on the web regarding how they reached this conclusion for each study. Again, however, the review by the NIFDI Research staff suggests that this judgment may not be appropriate for at least some of the works.

For instance, one of the studies rejected in the Reading Mastery review for having a "confounding factor" was Carlson and Francis's (2002) evaluation of the RITE program. For some reason this study had mistakenly been included in WWC's earlier review of the Direct Instruction Corrective Reading program and had also been rejected in that review for a supposed confound. This rejection was questioned in the letter written to Mathematica on June 25, 2008, and the response to this query in the September 8 correspondence sheds greater light on the WWC's definition of a "confound." This explanation also, however, raises serious questions about the validity of the WWC review and its implications for educators.

[13] The WWC report lists 6 studies that were rejected for this reason, but three of these were unpublished works such as those described above.

The Carlson and Francis study involved thousands of students, hundreds of teachers, and dozens of schools and used sophisticated multivariate techniques to analyze the data. The only intervention in the experimental schools was Direct Instruction's Reading Mastery, and as with the other studies, the results strongly favored this curriculum. After scouring the article, the NIFDI research staff found one possible explanation for the WWC's determination. On page 143, the authors described the intervention: "In addition to the teaching of skills directly related to the RM curricula, the … program also strives to provide teachers with strong classroom management techniques" (p. 143).[14] The September 8, 2008, communication from Mathematica confirms this interpretation, with the statement that, "A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program" (Dynarski, p. 5).

In reality, strong classroom management is part and parcel of the Direct Instruction approach. Training teachers in such management is part of the in-service training that teachers receive in learning how to implement the curriculum as well as prominently included within the teacher's guide to the program. Part of the reason that DI is so successful is that it provides not just well designed curricular materials but well developed, research-based guidance on how the curriculum should be administered. Because these guidelines are well documented in the various guides to the programs, they should have been part of the knowledge base of a competent reviewer. The classroom management was not a confounding element of the intervention, but was an integral part of the curriculum and its appropriate delivery.

More importantly, virtually all curricular programs include elements of teacher training and discussions of classroom management. It is reasonable to argue that if studies of Reading Mastery that include training for teachers are to be excluded, all other studies that include training for teachers should also be excluded. We know of virtually no legitimate curriculum that does not include some type of instructional overview for teachers.

In addition, complete training in a program is vital to ensuring that it is implemented with fidelity. Disallowing such elements from a design could produce a very serious threat to internal validity of a study. Surprisingly, the WWC's approach seems to discount the importance of the fidelity of treatment implementation, suggesting that "there is no standard metric" with which to rate and assess fidelity" and that a better approach is their reliance on "replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies" (Dynarski 2008, p. 3).

At least two very serious problems are immediately apparent with these statements. First, those familiar with any number of structured learning and behavioral approaches know that there is in fact a well developed literature of fidelity measures and that such measurements are very important both in helping to train practitioners and to assess the extent to which programs are being adequately implemented. To suggest that such measures do not exist is to demonstrate an unfortunate unfamiliarity with the literature. Second, to rely on replications as a substitute for

[14] The WWC Reading Mastery bibliography lists the Carlson and Francis article as in the Journal of Direct Instruction. The article also appeared in the Journal of Education for Students Placed at Risk in 2002, volume 2, issue 2, pp. 141-166. The page number in the text refers to this citation.

measures of fidelity could well produce very misleading results. Numerous replications of a poorly implemented program can, in no stretch of the imagination, provide an adequate test of a program. To assume that poor implementations are so rare that their impact would "average" out is, at best, a contention that should be subject to empirical test. More realistically, relying on such an assumption would likely produce very inaccurate and misleading results, for the errors that result from ignoring fidelity are systematically biased. As detailed in Appendix D, including studies with poor fidelity within summary analyses produces higher ratings of poor programs and lower ratings of good programs, thus providing very misleading conclusions for consumers.

Finally, two other studies that were rejected for an alleged confound were those by Gunn and associates (2002, 2005). These works report the results of a study that included both a behavioral intervention and Reading Mastery. While it is understandable that a naïve reviewer might quickly reject these articles, a careful reading shows that the authors directly address the issue of any possible confound and, based on both their results and other works in the field, note that this was unlikely. Again, these articles, plus a third by the authors that was not included in the WWC listing, find results strongly in favor of Reading Mastery.[15]

Summary

To summarize, examination of the What Works Clearinghouse's review of Reading Mastery indicates a number of aspects that should lead consumers to question the accuracy of the report. Part of the problem involves the ways in which studies were selected for inclusion and exclusion from review. An arbitrary time frame resulted in a large number of studies being excluded. In addition, however, the Clearinghouse ignored an even larger number of studies published within the chosen time frame. It appears that the WWC made no attempt to examine any of the large and easily available meta-analyses and literature reviews and use the studies listed in those documents. In addition, the accuracy of the WWC report should be questioned because of the quality of the reviews. The reviews appear to be both inaccurate and misleading.

Inaccuracies in Other WWC Reports

As noted above, the letter of June 25, 2008 to Mathematica was written before the Reading Mastery review was posted. One issue discussed in the letter is of such concern that it deserves additional discussion in this comment: the misrepresentation of results for at least one program that was given high ratings.

The WWC has given high ratings to the Reading Recovery program, a short-term tutoring intervention. The WWC website concludes that "Reading Recovery® was found to have positive

[15] Of the remaining three peer reviewed studies examined by WWC, two were excluded from review because they did not include a comparison group, but relied on comparison of pretest and posttest scores for a single group of subjects (Humphries et al 2005 and Marchand-Martella et al 2006). A third (Kamps and Greenwood 2005) was excluded because it "does not examine the effectiveness of an intervention." Examination of this study indicates that, in fact, just the opposite is true – the study reports the results of the first few months of a randomized trial of interventions comparing several more highly structured curricula, including Reading Mastery, with less structured approaches. Although descriptive statistics indicate that the students exposed to Reading Mastery have had substantially higher rates of progress than students in the traditional settings, inferential statistics that examine differences between the settings were not included. Thus, while definitive judgments could not be made regarding the statistical significance of the results, to indicate that this study does not involve an intervention is, at best, misleading.

effects on students' alphabetics skills and general reading achievement outcomes and potentially positive effects on comprehension and fluency" (http://ies.ed.gov/ncee/wwc/reports/beginning_reading/reading_recovery/). A careful reading of the articles cited by the website, however, indicated that, in at least two cases, the WWC review did not accurately reflect the content of the research studies.

For instance, one article (Baenen, et al 1997) was cited as showing positive impacts of RR, yet even a cursory reading of the article indicates that the authors found remarkably little success, especially in the long-term, for students in the program. The article concludes (p. 176) that although about one-half of the students in the tutoring program had successfully reached first grade reading levels at the end of the year, success rates declined in subsequent years of implementation. In addition, by third grade, there was no difference in achievement scores, needs for retention, special education, or Chapter 1 assignments of students who had participated in Reading Recovery and other students. The authors concluded that Reading Recovery was very expensive to implement in relation to the benefits that it provided.

The seemingly erroneous conclusion of the WWC was brought to the attention of Mathematica in the June 25 letter. The response (Dynarski 2008, p. 4) acknowledged the results outlined above, but defended the WWC's conclusions by stating that their reports "prioritized one-year results" and that the findings regarding the results in later grades were included in a technical appendix. Yet, as teachers, parents and students would attest, how well one reads at the end of third grade is much more crucial in determining eventual success than how well one reads at the end of first grade. Learning to read is not a one-year process, but is a multi-year endeavor. If gains in first grade do not persist through third grade, the first grade gains have very little worth. Ignoring the very important conclusions of Baenen and associates would seem to do a disservice to parents and teachers.

An even more disturbing example involves an article by Iversen and Tunmer (1993), which was also cited by the WWC as supporting the conclusion that RR is effective. The major purpose of this study was to compare the standard RR program to a "modified" program that included explicit instruction in phonological skills. The major variable of interest to Iversen and Tunmer was how long children took to reach a level of competency where they could discontinue special tutoring, the major goal of a tutoring program such as RR. Students in both the unmodified Reading Recovery program and the modified program (including instruction in phonologically based elements) eventually caught up with the other children, but the students in the modified program were able to discontinue tutoring much earlier. The standard Reading Recovery program was found to be 37 percent less efficient than the modified program. In addition students in the modified program continued to have higher levels of achievement and higher rates of learning at the end of the school year. The authors provide an extensive discussion and additional analyses that demonstrate the fallacy involved in Reading Recovery about the ways in which word recognition skills develop. They clearly conclude that Reading Recovery is not an efficient method for teaching children to read and that phonological training is superior.

The WWC chose to ignore any results regarding this comparison group, which received phonological training and had superior achievement, "because it was a modified version of the

standard program." The September response to our query regarding their judgment stated, "the WWC examined the results most relevant to the question of whether Reading Recovery improves reading proficiency compared to a reasonable counterfactual." That counterfactual was, apparently, having no tutoring at all, not tutoring with a more effective program. The statement goes on to note that the results with the other comparison groups were mentioned in an appendix, implying that such information could be available for those who were interested.

Even though the WWC did acknowledge the negative results regarding Reading Recovery in a technical Appendix, the overall conclusion given on its website is certainly misleading and does not reflect the conclusions of the articles. The chance of a parent or school official accessing a technical appendix to find information that contradicts that given in the major pages of the web site is extremely remote. The actions of the WWC regarding the material on Reading Recovery not only resulted in misrepresenting the effectiveness of Reading Recovery but also that of phonologically based programs and provides further evidence that its conclusions should be seriously questioned.

Recommendations

Based on a careful review of the evidence, we suggest that consumers approach the ratings of the WWC with extreme skepticism. With respect to Reading Mastery the list of sources that were reviewed was extremely selective and tapped only a very small proportion of the extant literature. Perhaps even more disturbing, the reviews of the works that were selected were often inaccurate and misleading. We see no reason to accept their conclusion of "no evidence" as valid.

From our correspondence with the WWC there also appear to be a number of very serious policy and procedural issues that raise questions regarding the accuracy of any ratings that they have developed. As detailed above, their policies in a number of areas appear to differ substantially from traditional scientific practices. These include such areas as decisions about how to select studies to examine, classifying teacher training procedures as a confounding element, relying on replications to "average out" problems with fidelity of implementations, and focusing on one-year studies rather than multi-year results, even when such multi-year results are available and more indicative of student success.

Part of the reason that the WWC has reached erroneous conclusions may reflect the way in which its review process departs from well established traditions of scientific research. In our June, 2008, letter to Mathematica, we made several suggestions for changes to their procedures. These suggestions in many cases overlapped with those made by other critics of WWC (see Slavin 2008 and McArthur 2008). For instance, we suggested that using standard quality control mechanisms, such as two reviewers for every article with a third if results differ and a peer review process before publication of ratings, could go far to help ensure more accurate results. We suggested that the cumulative nature of the scientific enterprise should be acknowledged, and the WWC's results should be compared with the already well developed body of meta-analyses and literature reviews. When WWC's results differ from these reviews, they should investigate why and adjust the results as needed. In addition, metrics that are commonly used within the social sciences, such as measures of effect size would greatly enhance faith in the WWC's conclusions. Issues of both internal and external validity in decisions regarding acceptable

designs should be weighed. Much more accurate information could be provided if the WWC understood the wide variety of research approaches that can and must be used within real-life educational settings. Instead of simply rating studies' quality by the nature of the research design, elements related to sample size, statistical significance, substantive significance, and length and fidelity of intervention should be included, with global ratings reflecting the preponderance of evidence regarding interventions from all available data. Finally, it is crucial that the WWC ensure that reviewers are knowledgeable in the substantive areas that they are reviewing as well as in the methodological details.

Unfortunately, the concerns expressed in the correspondence with Mathematica and the WWC, as well as the concerns of earlier critics Slavin (2008) and McArthur (2008) appear to have been unheeded and, in fact, rejected. As stated in that letter, the recommendations were made "in the spirit of the intent behind the What Works Clearinghouse – a desire to provide schools and families with the most accurate information." The letter also noted that "the problems with the material posted on the website appear to be much more widespread than can be handled in a piece-meal fashion. To allow inaccurate material to remain does a disservice to schools, families and students."

The communications with Mathematica appeared before the review of Reading Mastery was posted. Unfortunately, the quality of that review has made the earlier judgments of problems with the WWC even more apparent and worthy of grave concern. If Mathematica refuses to remove the faulty information from the WWC website, consumers should be extremely wary of the ratings posted there and consult the standard social science literature for more accurate information.

References

Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.

Adams, Marilyn J. 1990. Beginning to read: Thinking and learning about print. Cambridge, Mass.: The MIT Press.

American Federation of Teachers. 1998. Building on the best, learning from what works: Seven promising reading and language arts programs. Washington, D.C.: AFT.

Anderson, R. C., E. H. Hiebert, J. A. Scott, and I. A. G. Wilkinson. 1985. Becoming a nation of readers: The report of the Commission on Reading. Washington, D.C.: National Institute of Education.

Baenen, Nancy, Alissa Bernholc, Chuck Dulaney, and Karen Banks. 1997. Reading Recovery: Long-Term Progress After Three Cohorts. Journal of Education for Students Placed at Risk 2: 161-181.

Baker, Scott K. Edward J. Kameenui, Deborah C. Simmons, and S. A. Stahl. 1994. Beginning reading: Educational tools for diverse learners. School Psychology Review 23: 372-391.

Beck, I.L. and E. S. McCaslin. 1978. An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. LRDC Report No. 1978/6 Pittsburgh.

Bond, G. and R. Dykstra. 1967. The cooperative research program in first-grade reading instruction. Reading Research Quarterly 2: 5-142.

Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.

Buechler, Mark. 2002. Catalog of School Reform Models: Program Report. Portland, Oregon: Northwest Educational Research Lab.

Chall, J.S. 1967. Learning to read: The great debate. New York: McGraw Hill.

Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.

Dynarski, Mark. 2008. "Responses to Concerns in 6/26/08 Letter from Jean Stockard," sent on September 8, 2008, from Mathematica, included in Appendix A of this report.

Foorman, B. R. 1995. Research on "the great debate": Code-oriented versus whole language approaches to reading instruction. School Psychology Review 24: 376-392.

Fukkink, R.G. and K. deGlopper. 1998. Effects of instruction in deriving word meaning from context: A meta-analysis. Review of Educational Research 68: 450-469.

Grossen, B. 1997. A synthesis of research on reading from the National Institute of Child Health and Human Development. Eugene, Oregon: University of Oregon.

Herman, R., D. Aladjam, P. McMahon, E. Masem, I. Mulligan, O. Smith, A. O'Malley, S. Quinones, A. Reeve, and D. Woodruff. 1999. *An educator's guide to schoolwide reform*. Washington, D.C.: American Institutes for Research.

Institute of Child Health and Human Development (ICHHD). 2000. Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). Washington, D. C.: U. S. Government Printing Office.

Iversen, Sandra and William E. Tunmer. 1993. Phonological processing skills and the reading recovery program. Journal of Educational Psychology 85: 112-126.

Juel, C. and C. Minden-Cupp. 2000. Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly 35: 458-492.

Kennedy, M.M. 1978. Findings from the Follow Through Planned Variation Study. Washington, D.C.: U.S. Office of Education.

McArthur, Genevieve. 2008. Does what works clearinghouse work? A brief review of FastForWord®. Australasian Journal of Special Education 32: 101-107.

National Institute of Child Health and Human Development (NICHHD). 1996. Thirty years of NICHD research: What we now know about how children learn to read. Effective School Practices 15: 33-46.

National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, D.C.: National Institute of Child Health and Human Development.

Pflaum, S., H. J. Walberg, M. L. Karigianes, and S.P. Rasher. 1980. Reading instruction: A quantitative analysis. Educational Researcher 9: 12-18.

Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.

Slavin, Robert E. 2008. Perspectives on evidence-based research in education – What works? Issues in synthesizing educational program evaluations. Educational Researcher 37: 5-14.

Smith, S., D. Simmons, M. Gleason, E. Kameenui, S. Baker, M. Sprick, B. Gunn, and C. Thomas. 2001. An analysis of phonological awareness instruction in four kindergarten basal reading programs. Reading and Writing Quarterly 17: 25-50.

Snider, V.E. 1990. Direct instruction reading with average first graders. Reading Improvement 27: 143-148.

Snow, C. E., M. S. Burns, and P. Griffin (eds.). 1998. Preventing reading difficulties in young children. Washington, D. C.: National Academy Press.

Stanovich, K.E. 1994. Romance and reality. The Reading Teacher 47: 280-291.

Appendix A
Communications with the What Works Clearinghouse and Mathematica

1) e-mail –December 2, 2007, from J. S. to WWC
2) Reply from WWC, February 8, 2008
3) JS to WWC, March 16, 2008
4) WWC to JS, April 9, 2008
5) JS to Mathematica, June 25, 2008 (letter sent by U.S. post)
6) Follow-up letter JS to Mathematica, July 31, 2008 (sent fed-ex with receipt notice)
7) Reply from PD to JS, August 6, 2008
8) JS reply to PD, August 7, 2008
9) A. S. to JS, September 6, 2008
10) Reply from Mathematica, September 8, 2008

1) e-mail –December 2, 2007, from J. S. to WWC

Sent: Sunday, December 02, 2007 12:30 PM

To: Mark Dynarski

Subject: Questions on What Works Clearing House Procedures

Dear Dr. Dynarski:

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction (NIFDI). One of my first tasks was to read as much of the background research in the area as I could. Along the way I ran into the "What Works Clearinghouse." In my review I developed several questions that I am hoping you can answer.

First, I noticed that the WWC has listed relatively few of the studies that deal with Direct Instruction. While a few were rejected because they didn't meet some specific standards of design, it appears that many others were rejected because they were published before 1985. I have not been able to find a scientific justification for selecting this date. No studies were cited that suggested children's learning styles had altered after that time or that schools' organizational structures or teachers' actions had changed. In fact, the only research article that I found to justify the methods was actually published in 1974 – 10 years prior to the beginning of your targeted dates. This article dealt with phonological awareness, and had nothing to do with systematic literature reviews and, of course, included nothing to suggest that processes related to phonological awareness have changed over time.

In all the other areas with which I am familiar we use the full range of research articles and, unless there are clear reasons established through the research literature, do not set an arbitrary date for inclusion or exclusion of results. The choice of this date seems to have particularly affected the DI literature because a great deal of the pioneering work was published before the cut-off point. Because I'm new to this particular work, I wondered if you could provide me with any explanation based in the research literature for your choice of this cut-off year. I did find a suggestion that, "if sufficient time and resources remain," studies published before 1985 might be reviewed. Do you know if this step will be taken? From my reading of the literature I believe that your limit on dates of studies may have seriously limited the selection of effective resources that educators will find on your website; and I know that this is a primary aim of your project.

Second, I noticed that a number of studies were rejected because they included children outside the K-3 grade range. I found this somewhat unusual, for the children that are most in need of effective instruction are those who have not learned to read by older ages. Because the effectiveness of DI has been so well demonstrated with regular students at the K-3 grade range, the more recent focus on other ages seems only greater evidence of its worth. This is especially

so in the case of special education students and those with behavioral and other issues. Thus, I couldn't understand the rationale of omitting all the studies that targeted these most needy students. Any explanation you might have of this decision would also be greatly appreciated.

Third, and perhaps most important, as I read studies more closely I found that the Clearinghouse's interpretation of some of the studies didn't seem to accurately reflect the actual content of the reports. I know that these reviews are often done by staff members or graduate students who might not be familiar with all of the intricacies of social research. Thus, I assume that we can send corrections and they will be added.

I look forward to hearing from you, either through e-mail (jeans@uoregon.edu) or by phone (541-346-5005). I'm sure that we share the ultimate goal of helping all students – no matter what their background – develop academic skills and appreciate your appreciate your help.

Sincerely,
Jean Stockard, Professor Emerita

2) Reply from WWC February 8, 2008

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated December 2, 2007 below.

The first question you raised concerned the cut-off year for studies that are reviewed by the WWC. To maintain a focus on current research, our reviews examine research conducted in the past 20 years. This cut-off date of 20 years is the standard for every topic reviewed by the WWC. The review for Early Childhood Education (ECE), under which Direct Instruction falls, marked its cut-off date at 1985 (the review began in 2005).

The ECE protocol states that "if sufficient time and resources remain after we have completed our review of research on interventions implemented post 1985, the ECE team will consider reviewing older research on curricula that are still in widespread use." This determination has not been made to date.

In terms of the age/grade range for the review of Direct Instruction research, the ECE team determined the parameters for the review. For further details on inclusion criteria please see the ECE protocol at http://ies.ed.gov/ncee/wwc/reports/early_ed/index.asp

In response to your third question about the qualifications of WWC topic review team members, each team is comprised of several trained professionals, each playing an integral part in the review process. The Principal Investigator (PI) for each topic is a well-known expert in his/her field and is responsible for leadership in conceptualizing the specific topic area, identifying and addressing issues during the review, and developing and reviewing topic and intervention reports developed for the topic. Leadership includes overseeing the quality in the production of the reports and making decisions, based on methodological and substantive expertise, that are not otherwise covered in the WWC protocols and procedures for report production.

The Project Coordinator (PC) is an established education researcher with relevant methodological and substantive expertise. The coordinator oversees the work of the WWC Review Team; manages that specific review; reviews research ratings; and writes and revises the work plan, protocol, and draft and final reports in collaboration with the PI.

Individual reviewers, who prepare initial summaries of studies for the WWC, are professional researchers with experience in research design and methodology. These reviewers undergo a rigorous training and certification process before conducting WWC reviews. .For more information about the staff that make up the WWC team, visit http://ies.ed.gov/ncee/wwc/overview/index.asp.

The WWC recently created a Quality Review Team to respond to concerns raised by study authors, curriculum developers or other relevant parties about WWC reviews published on our

website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. If a quality review concludes that the original review was flawed, a revision will be published. These quality reviews are one of tools used to ensure that the standards established by the Institute of Educational Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.

If you have concerns about a published WWC review that you think warrant a quality review, please send those concerns to our Help Desk at info@whatworks.ed.gov. Please identify the study in question, the specific issue(s) that you think were handled incorrectly, and where relevant, explain what you think is the correct interpretation of the study.

I hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

3) March 16, 2008 – JS to WWC

From: Jean Stockard [mailto:jeans@uoregon.edu]
Sent: Sunday, March 16, 2008 11:59 AM
To: WhatWorks
Subject: re: Questions on What Works Clearing House Procedures

Thank you for your reply of February 8 to my query of December 2. I am pleased to hear that you have established a Quality Review Team. Since my initial query I have continued to examine the reviews and have found quite a few that appear very problematic. I, and my colleagues, will be submitting our concerns to you within the coming weeks.

Unfortunately, I'm afraid that your reply failed to address two of my concerns. Your responses simply restated WWC's policies regarding the cut-off year for studies and the age-grade parameters without providing the requested justification. Again, I would be very interested in knowing the scientific and pedagogical justifications for not incorporating the entire corpus of work in your reviews and for limiting your reviews of studies to only those that included children in grades K-3.

My earlier, rather lengthy, explanation of my concerns is included in this e-mail. If you wish me to clarify any part of my query, please let me know. Thank you for your attention to this matter.

Sincerely,

Jean Stockard, Ph.D.
Emerita Professor

4) WWC to JS, April 9, 2008

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated March 16, 2008 below.

In response to your first concern, the cut-off date of 20 years is a parameter set to ensure the research reviewed is most-relevant to classrooms as they operate today, and to ensure that the scope of studies to be reviewed is manageable. That said, for every topic reviewed by the WWC, the Principal Investigator (PI) is given the flexibility to include studies published earlier, if they think the expansion is important for the review. In this case, the decision was made to keep the 20 year parameter in place.

In terms of the age-grade parameters established, this is determined by the PI and depends on the topic area and the studies under review. In this case, because the focus of the topic is on early childhood education, the age-grade parameter was set to K-3. Other topic areas have also focused on a subset of age-grade ranges. As the WWC expands, we anticipate expanding these topics to examine outcomes for other age-grade ranges.

The WWC solicited nominations from many sources for the topic areas and prioritized the topic areas based on the following criteria:

- potential to improve important student outcomes;
- applicability to a broad range of students or to particularly important subpopulations;
- policy relevance and perceived demand within the education community; and
- likely availability of scientific studies.

More information about this process and the specific topic areas can be found at http://ies.ed.gov/ncee/wwc/reports/.

Three Direct Instruction interventions were reviewed in two topic areas; the spread across topics was due to the samples included in the studies. Direct Instruction falls under the Early Childhood Education (ECE) area since it includes studies with preschool and kindergarten children where the majority (60% or more) of children in the sample are in preschool. This criterion for inclusion can be found at http://ies.ed.gov/ncee/wwc/pdf/ECE_protocol.pdf. One Direct Instruction intervention is also included in the English Language Learners topic area since it was used to supplement reading instruction for Spanish speaking students in grades K-3.

We hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

5) JS to Mathematica, June 25, 2008 (letter sent by U.S. Post)

June 25, 2008

Anita A. Summers, Chairperson and
Paul T. Decker, President and CEO
Mathematica, Policy Research Inc.
P.O. Box 2393
Princeton, NJ 08543-2393

Dear Drs. Summers and Decker:

I write to express my deep concerns with the material posted on the What Works Clearinghouse, which was established "to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education." I understand that Mathematica has now assumed responsibility for the Clearinghouse. I also know that Mathematica has a well established reputation for high quality research. I write to you in the hope that this correspondence can help you correct the errors in the current reports and establish future review procedures that are in accordance with standard scientific practices. The goals of the What Works Clearinghouse are too important to allow a flawed review process to continue or faulty reviews to be posted. None of us wants parents or schools to be given faulty information about the most effective curriculum for their students.

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction. In beginning to familiarize myself with the area I, of course, reviewed major analyses of Direct Instruction. As you no doubt know, in recent years, meta-analyses have become the most commonly accepted method for examining large bodies of research literature. The most recent meta-analysis of the achievement effects of comprehensive school reform models was conducted by Geoffrey Borman and associates (2002, Review of Educational Research). Their examination found that Direct Instruction was the most researched of any of the models and had the most studies conducted by third parties (those not affiliated with the developers). As they put it, "the research base for Direct Instruction is very extensive and of very good quality" (p. 187). Among the many different models used in their analysis, Direct Instruction had the strongest evidence of effects on student achievement.

In my reviews I found that this conclusion simply echoes that found in earlier meta-analyses and literature reviews. The overwhelming conclusion of the education community for many years, based on solid experimental evidence, is that Direct Instruction is one of, if not the, most effective curriculum currently available for teaching reading and mathematics.

In my examination of the literature I also reviewed the reports on Direct Instruction that had been prepared by the What Works Clearinghouse and was shocked and dismayed to find that the WWC conclusions were in marked contrast to those of the extant literature. My concerns with what I found prompted me to write to the Clearinghouse on two previous occasions (December 2, 2007 and March 16, 2008). The replies I received did not clearly address my

concerns. Even worse, my continuing examination of the work of the Clearinghouse has raised even more serious concerns about the quality of the reviews.

In recent months others have expressed similar concerns regarding the Clearinghouse's conclusions and procedures. Two examples of these critiques are Robert Slavin's article in the January/February, 2008 issue of Educational Researcher and Genevieve McArthur's article in the April, 2008 issue of the Australasian Journal of Special Education. While Slavin and McArthur focused their concerns on reviews related to Success for All (Slavin) and FastForWord® (McArthur), many of the concerns they raise parallel the issues I discovered in my review of the judgments of articles regarding Direct Instruction.

Below I outline in greater detail the extent of my concerns. I first discuss what I see as severe limitations of the process used to decide what studies to include in the review process and then give examples of very severe errors in the reviews themselves. I end with a few suggestions regarding what needs to be done to provide the most accurate information to the education community and to parents.

Problems with Exclusion/Inclusion Decisions

Some of the most disturbing elements of the WWC process involve the ways in which studies are selected for reviews. Some of these decisions arbitrarily limit the range of studies included and thus the information available to schools and parents. For instance, as I noted in my earlier communications, the decision to reject all studies for inclusion that were published before 1985 appears to have no scientific justification. The reply to my April 9 query claimed that the "cut-off date of 20 years is a parameter set to ensure the research reviewed is most relevant to classrooms as they operate today." I have not been able to find any studies cited on the website, or elsewhere, that document that children's learning styles, classrooms, teachers' actions, or schools' organizational structures have drastically changed over that time period. Most high quality studies conducted before 1985 are still relevant today. As I stated in earlier correspondence, in all the other areas with which I am familiar we use the full range of research articles and do not set an arbitrary date for inclusion or exclusion of results.

Similarly, some areas of the review have a very restrictive grade range for studies that will be included. For instance, for reviews of work on beginning reading, a number of studies, of both Direct Instruction and other approaches as well, were rejected because they included children outside the K-3 range, such as grades 3-5 or grades K-4. It is hard to understand why such studies would be summarily rejected.

The combination of the decision to limit reviews to work published after 1985 and to a narrow band of grades directly affected the numbers of studies of Direct Instruction that were reviewed. Even worse, this decision ignored the cumulative nature of science and research inquiry within education. Many studies of Direct Instruction with only K-3 students were completed before 1985. Given the consistently positive results from that work, the research focus then moved to populations of students with more extensive needs, such as older children who had not yet learned to read. Yet these studies were omitted because they were not focused on the general population. The real losers from these decisions are, of course, children and their

families who have been denied accurate and complete information, collected over a long span of time and from many different populations, about the most effective programs.

Some of the decisions regarding exclusion or inclusion of curricula and studies seem to have occurred with no stated justification. For instance, Direct Instruction's Reading Mastery program has long been acclaimed as a highly effective reading program, with strong research supporting its development and documenting its success. Yet, the WWC website states that they chose not to include Reading Mastery in their review of Beginning Reading studies (apparently not realizing that the RITE study noted below involved Reading Mastery). No explanation is given as to why this curriculum, which directly addresses beginning reading and which has a very large research base, was ignored. Again, the real losers with this decision are the students.

Finally, the range of work considered has been limited by downgrading findings from studies that do not incorporate strict random assignment. While we all know that random assignment is the "gold standard" for experimental work, the WWC's over-reliance on this criterion ignores the realities of how school organizations work. With this strict attention to random assignment, other aspects of research designs that are an even greater threat to internal validity can be ignored. The most important of these is no doubt ensuring the fidelity of treatment implementation, making sure that a program is implemented as the developers designed it. The complete rationale behind this concern is too lengthy to include in this letter. I worry that this criterion has involved a sterile and unthinking application of methodological rules that may work for growing corn in a field or worms in a lab. Yet they may be inappropriate, at best, and potentially harmful, at worst, for students in real-life schools and real-life neighborhoods where fidelity of treatment implementation is much more tenuous.

While the concerns noted above have resulted in fewer studies being included, another procedural decision of WWC has resulted in the inclusion of many studies that should probably not have been considered. Unlike most scholarly writing, the WWC material that I have reviewed has surprisingly little relationship to the extant published literature. As I have explored the studies that were reviewed for the website I found that many of them were very difficult to find and, even more shocking, did not appear in peer-reviewed journals, the standard mark of academic respectability. A surprisingly large number of the reports that were considered to have met the evidence standards, either fully or with reservation, were unpublished manuscripts. In some cases, I had to write to the original developers several times before receiving the manuscripts cited in the reviews. In contrast, numerous Direct Instruction articles that were rejected for review have been published in peer-reviewed journals and are available to the general public.

Finally, and perhaps most important, as I have read the studies and their evaluations more closely I have found that *the Clearinghouse's interpretations do not accurately reflect the actual* content of the reports. Below I give some examples of the problems that I have found to date. A very large proportion of the reviews I've read have had little correspondence to the articles themselves.

Quality Problems with the Reviews

One of the programs that received a high rating from WWC in Beginning Reading is Reading Recovery®. Most of the other programs reviewed by WWC, such as Direct Instruction, are designed to be used with entire classrooms of students as part of the regular instructional program. In contrast, Reading Recovery (RR) is a "pull-out" program, one that is applied only when students are having trouble in their regular classrooms. The inclusion of the tutorial based Reading Recovery in the Beginning Reading Category is, in fact, rather surprising, given the WWC's stated objective of reporting on "comprehensive … programs," those that enhance "whole school literacy," and "basals/textbooks intended for whole-school/whole-classroom use" (http://ies.ed.gov/ncee/wwc/PDF/BR_protocol.pdf, p. 5).

The WWC website concludes that "Reading Recovery® was found to have positive effects on students' alphabetics skills and general reading achievement outcomes and potentially positive effects on comprehension and fluency" (http://ies.ed.gov/ncee/wwc/reports/beginning_reading/reading_recovery/). A careful reading of the articles cited by the website, however, indicates that the reviewers apparently did not understand basic elements of the program or the reported research.

For instance, the article by Baenen and associates (1997) is one of the articles cited as supporting WWC's positive conclusions. In this study first graders were randomly assigned to participate in the pull-out tutoring program or to remain in their regular classroom. Note that the design does not compare Reading Recovery to an alternative curriculum, which would be a rigorous test of the treatment, but instead to a no-treatment situation. Even with this built-in advantage, the authors found remarkably little success, especially in the long-term, for students in the Reading Recovery program. The article concludes (p. 176) that although about one-half of the students in the tutoring program had successfully reached first grade reading levels at the end of the year, success rates declined in subsequent years of implementation. In addition, by third grade, there was no difference in achievement scores of students who had participated in Reading Recovery and other students. Two years after the treatment there were no differences in needs for retention, special education or Chapter 1 assignments. The authors concluded that Reading Recovery was very expensive to implement in relation to the benefits that it provided.

An even more disturbing example involves the article by Iversen and Tunmer (1993), which was also cited by the WWC as supporting the conclusion that RR is effective. The major purpose of this study was to compare the standard RR program to a "modified" program that included explicit instruction in phonological skills. The WWC chose to ignore any results regarding this comparison group "because it was a modified version of the standard program." Unfortunately, this decision resulted in completely ignoring the major conclusions of the article and thus misrepresenting the effectiveness of Reading Recovery and phonologically based programs as well.

The major variable of interest to Iversen and Tunmer was how long children took to reach a level of competency where they could discontinue special tutoring, the major goal of a tutoring program such as RR. Students in both the unmodified Reading Recovery program and the modified program (with instruction in phonologically based elements added) eventually

caught up with the other children. But the most important finding was that the students in the modified program were able to discontinue the tutoring much earlier. The standard Reading Recovery program was found to be 37 percent less efficient than the modified program. In addition students in the modified program continued to have higher levels of achievement and higher rates of learning at the end of the school year. The authors provide an extensive discussion and additional analysis that demonstrate the fallacy involved in Reading Recovery about the ways in which word recognition skills develop. They clearly conclude that Reading Recovery is not an efficient method for teaching children to read and that phonological training is superior.

Similar, very serious, problems occurred with the review of studies of Direct Instruction, but the mistakes with these articles led the reviewers to dismiss the studies out of hand. One of the most egregious cases involves the study by Carlson and Freeman (2002). The stated reason for dismissing this study was that "there was a confound, with the Direct Instruction intervention being modified or combined with other interventions." In fact, there was only one intervention, and it appears that the reviewer simply did not understand the nature of the intervention well enough to accurately read the article. As with virtually all other studies of Direct Instruction, the analysis, which used sophisticated multilevel techniques, had results strongly in favor of the DI curriculum.

It is possible that the reviewers responded to this section of the description of the intervention, which was termed "RITE" and used the Reading Mastery program of Direct Instruction: "In addition to the teaching of skills directly related to the RM curricula, the RITE program also strives to provide teachers with strong classroom management techniques" (p. 143). Strong classroom management is part and parcel of the Direct Instruction approach. Part of the reason that DI is so successful is that it provides not just well designed curricular materials but well developed, research-based guidance on how the curriculum should be administered. These guidelines are well documented in the various guides to the programs and should have been part of the knowledge base of a competent reviewer. The classroom management was not a confounding element of the intervention, but was an integral part of the curriculum.

In other cases, articles regarding a Direct Instruction curriculum were rejected for methodological issues, while those with a different curriculum, but the same methodological approach, were accepted. For instance, Waldron-Soler, et al (2002) found that students instructed in the DI program, Language for Learning, had significantly higher achievement than those in programs without explicit language instruction, using analysis of covariance to control for pretest scores. The study was rejected for consideration because "the intervention and comparison groups cannot be considered equivalent at baseline, even with the use of covariates in the analysis." However, the published article includes means and standard deviations for both the experimental and control groups. I calculated simple t-tests and found that in fact there were no significant differences between the groups at pre-test. Such a test was apparently not done by the reviewers, despite their stated conclusion.

To help understand why the Waldron-Soler article was dismissed, I examined studies of Success for All that WWC deemed had met the minimal requirements for inclusion. (Success for All is a reading program that has been found to be effective almost as often as Direct Instruction

and was modeled on many elements of DI.) Three of these papers, none of which had been published in a peer reviewed journal, also used analysis of covariance to adjust for pretest scores (Dianda and Flaherty 1995; Ross, Alberg and McNelis, 1997; and Ross, McNelis, Lewis and Loomis, 1998). Why this adjustment method would pass muster with the unpublished articles regarding Success for All, but be dismissed with the published article regarding Direct Instruction is very unclear and disturbing.

Similarly, two articles by Tobin (2003, 2004) on the Horizons DI reading program were rejected because the study groups were supposedly "incomparable." The studies, which were again published in peer reviewed journals, used a technique of matching students in the experimental and treatment groups on pretest measures. Tobin (2003) reported that there were no statistically significant pretest differences between the groups, and, as usual, the results supported the superiority of the DI program. Again I looked at the accepted studies of Success for All to try to understand why Tobin's articles had been rejected. I found that one of the accepted SFA studies (Smith, Ross, Faulks, et al, 1993) also individually matched students on pretest scores. Unfortunately, this unpublished paper only reports the results in graphical form and it was impossible to actually examine the means and standard deviations to test for any differences. Again, I am baffled as to why the unpublished Smith et al piece, which omitted crucial methodological details, would be accepted while the published Tobin pieces, with similar methodology and including the statistical details, were rejected.

Recent innovations to the WWC website have only compounded these errors. The "create your own summary" feature (http://ies.ed.gov/ncee/wwc/Reports/Topic.aspx?tid=01) invites viewers to create a summary of the reports by simply clicking a button. The resulting page lists interventions in the order of the "improvement index" score, implying that those at the top of the list are the most effective while those at the bottom of the list are least effective.

One of the programs often listed near the top of the Beginning Reading lists is Reading Recovery; but, as explained above, the reviews of the program were fatally flawed and inaccurate. Another program rated near the top is Early Intervention in Reading (EIR). Like Reading Recovery, EIR involves special tutoring for low performers, yet, as with the reviews of Reading Recovery, the judgment of EIR's effectiveness is questionable. The conclusion regarding EIR is based on one unpublished study, involving 59 students. Half of these students received tutoring in addition to their classroom instruction and their reading achievement was compared to those who received no additional systematic instruction. While the children who received tutoring did better than those who did not, there is no way to tell if EIR is superior to another type of intervention or if it is simply the additional practice time or attention that made the difference.

In short, given the flawed process of selecting studies for review and the extremely poor quality of the review process, any "effectiveness ratings" are undoubtedly inaccurate and misleading. To prominently display these ratings and encourage their use does a great disservice to schools and parents throughout the country and will eventually, I believe, harm Mathematica's reputation of scientific integrity.

What Should Be Done

I bring these very serious concerns to your attention in the hope that you will correct the mistakes that have been made and establish procedures that will prevent them from occurring in the future. The articles by Slavin and McArthur provide a number of very good suggestions. For instance, Slavin provides very cogent suggestions regarding the development of program ratings that more accurately reflect the strength of the evidence, altering the decision process regarding acceptable study designs, and giving greater weight to larger studies and those of longer duration. McArthur suggests that reviews be limited to studies that have passed peer review, that measures be adjusted to more accurately reflect the concepts that are supposedly being rated, that the method of rating programs more accurately reflect the results and, most important, that reviewers be familiar with the subject area and that all summaries be reviewed by independent experts in the area before publication on the website.

I fully support these suggestions and applaud their common notion that the WWC should use standard methods of scientific procedure and review and that those who do the reviews should be familiar with the areas under study. Acknowledging that there is overlap with the suggestions of Slavin and McArthur, I will list my own recommendations:

1) Make the criteria used in searching for studies explicit and replicable.

2) Concentrate on peer reviewed studies first.

3) Ensure that all studies accepted are available in full text for review by others.

4) Use all the available evidence regarding interventions, with no arbitrary cut-off dates or age or grade ranges that unfairly exclude a body of evidence.

5) Ensure that reviewers are knowledgeable in the substantive areas that they are reviewing as well as in the methodological details.

6) Develop standard quality control mechanisms, such as two reviewers for every article with a third if results differ and a peer review process before publication of ratings.

7) Develop procedures to ensure comparability from one area of the review to another.

8) Compare results that are obtained with the already well developed body of meta-analyses and literature reviews. When WWC's results differ from these reviews, investigate why and adjust results as needed.

9) Use measures of effect size that are more commonly used within the social sciences and comparable to those used in other review studies. Include elements related to sample size, statistical significance, substantive significance, and length of intervention.

10) Weigh issues of both internal and external validity in decisions regarding acceptable designs and recognize the wide variety of approaches that can and must be used within real-life educational settings.

11) Organize topics in ways that reflect how schools actually operate, such as core reading programs, supplemental programs, etc.

While all of these changes are vital to ensuring that future reviews are accurate, they do not deal with the faulty reviews that are already posted on the website. Given the very serious concerns that Slavin, McArthur, and I (and perhaps others of whom I'm not aware) have raised regarding the content of the WWC website, I believe that the reviews and ratings that are now posted must be removed until they can be reviewed for accuracy with well designed quality control measures. The problems with the material posted on the website appear to be much more widespread than can be handled in a piece-meal fashion. To allow inaccurate material to remain does a disservice to schools, families and students. I know that the vast majority of the work

posted on the website was not done under Mathematica's guidance. Continuing to post such faulty material can only, unfairly, sully Mathematica's reputation.

Thank you for attending to my concerns. Please know that they are presented in the spirit of the intent behind the What Works Clearinghouse – a desire to provide schools and families with the most accurate information. If you wish to discuss any of my concerns further, please do not hesitate to contact me. I look forward to hearing from you.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
jstockard@nifdi.org
Toll free phone # 877-485-1973

6) Follow-up letter, JS to Mathematica, July 31, 2008, sent with return receipt via Fed-ex.

July 31, 2008

Anita A. Summers, Chairperson and
Paul T. Decker, President and CEO
Mathematica, Policy Research Inc.
P.O. Box 2393
Princeton, NJ 08543-2393

Dear Drs. Summers and Decker:

It has now been more than a month since I wrote you regarding my concerns with the material posted on the What Works Clearinghouse website. However, I have received no response or even an acknowledgement of my letter. In case you did not receive the original communication, I have included a copy in this package. I have also sent this letter via Federal Express so that I may confirm that it has been received.

As you know, I believe that the analyses of curricular material posted on the WWC website are fatally flawed. They provide very erroneous information to educators and parents. Because Mathematica only recently assumed responsibility for the WWC and because I know that Mathematica has had, in the past, a reputation for high quality work, I felt it important to let you know of these problems privately so that you could address them.

The lack of response to my earlier letter suggests, however, that my concerns, as well as those of others, have not been given serious credence. Given the serious nature of the issues involved, unless I hear from you within the next 2 weeks, I will have to pursue alternative means of dealing with this matter. I am available to speak via phone (toll-free at 877-485-1973) or to communicate via e-mail (jstockard@nifdi.org) if you prefer those means of communication. I hope to hear from you soon.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
jstockard@nifdi.org
Toll free phone # 877-485-1973

7) Decker to JS, August 6, 2008

Response to Your Letter

Dear Dr. Stockard:

Thank you for prompting me regarding a reponse to your concerns about the What Works Clearinghouse (WWC). I received your earlier letter and reviewed it carefully. As you know, your original letter contained a substantial amount of material as well as 11 specific recommendations to change the criteria and processes supported by the Clearinghouse. I therefore have asked for input from both our team working on the project as well as our client for this work, the Institute of Education Sciences (IES) at the U.S. Department of Education, to develop an appropriate response to your letter.

As you probably know already, Mathematica is administering the WWC under contract to IES, and our work is bound by the statement of work contained in the contract. Furthermore, we work closely with IES to determine how best to fulfill the statement of work as the details of the work unfold. Hence, any potential changes that Mathematica could make in response to your recommendations, including changes in the WWC web site, may be limited by the contract or by the guidance of our client.

Thank you for your attention to the What Works Clearinghouse initiative. I appreciate your thoughtful review, and I look forward to sending you a lengthier response.

Paul Decker

_____

Paul Decker
President and Chief Executive Officer
Mathematica, Inc.
Box 2393
Princeton, NJ 08543
609-275-2290

8) JS to Decker, August 7, 2008

Dear Dr. Decker:

Thank you for your response to my letters. I look forward to receiving your lengthier reply. Might you have any idea of the time frame for this response?

Again, thank you for attending to my concerns.

Sincerely,

Jean Stockard

9) A.S. to J.S., September 6, 2008

Ms. Stockard: Thank you for copying me on your Aug. 29 letter to Paul Decker. I am
following up on it with him, and am assured that he is addressing the issues meticulously. I will continue to connect with him on this matter.

Sincerely,

Anita A. Summers

10) Mathematica reply, September 8, 2008

Note that this has two part: The first is a cover letter from Paul Decker the President and CEO of Mathematica. The second is a more detailed response to the June 25 letter, apparently prepared by Mark Dynarski. This is referred to in the text as Dynarski, 2008.

MATHEMATICA
Policy Research, Inc.

Paul T. Decker
President and Chief Executive Officer

P.O. Box 2393
Princeton, NJ 08543-2393
Telephone (609) 799-3535
Fax (609) 799-0005
www.mathematica-mpr.com

September 8, 2008

Dr. Jean Stockard, Ph.D.
Director of Research
National Institute of Direct Instruction
P.O. Box 11248
Eugene, OR 97440

Dear Dr. Stockard:

Thank you for your June 25 letter concerning the What Works Clearinghouse (WWC). Mathematica Policy Research wants to ensure the quality and accuracy of all information contained in the WWC, and it works with study authors, curriculum developers, and consumers of the WWC to correct inaccurate information. We therefore appreciate your interest in and feedback on the WWC.

The WWC was designed to establish and apply a set of standards to identify rigorous research related to the impacts of education interventions. Studies that meet these standards are considered to have designs with causal validity, which gives practitioners increased confidence that the results of the study actually reflect the true impact of the intervention being examined. These standards were developed by leading experts in education research methodology, and they are applied through a systematic review process that includes repeated checks for quality and accuracy. They represent research criteria that we believe to be vital to the purpose of the WWC.

I have closely reviewed the concerns in your letter. I have also asked for input from Mark Dynarski, the director of the WWC effort, and representatives from the Institute for Education Sciences at the U.S. Department of Education. Based on my review and the input I received, I am convinced that the standards being used by the WWC adequately reflect the original mission of the WWC, as outlined above, and that those standards are being applied appropriately and consistently. In the attached document, Mark Dynarski provides more detailed responses to the issues you raise concerning the WWC's review of Direct Instruction.

Your letter makes several suggestions for the WWC moving forward. Some suggestions are counter to the systematic review process that is the foundation of the WWC. For example, as explained in the attachment, the criteria for date and age ranges being reviewed are established with specific rationales, and the systematic review process gives equal weight to peer-reviewed and non-

An Affirmative Action/Equal Opportunity Employer

peer-reviewed sources. As we expand the WWC's activities into new topic areas, we will be mindful of the suggestions you raised in your letter.

I hope this information clarifies the WWC mission and the way in which the WWC operations support that mission.

Sincerely,

(b)(6)

Paul T. Decker

RESPONSES TO CONCERNS IN 6/25/08 LETTER FROM JEAN STOCKARD

Contrasting conclusions between WWC and extant literature

The letter states that WWC conclusions differ from extant literature including meta-analyses and literature reviews. It is important to note that the WWC is designed to produce a systematic review of literature. A sound definition of a "systematic review" is in the recent publication *Knowing What Works in Health* Care by the Institute of Medicine (IOM):[1] *A systematic review is a scientific investigation that focuses on a specific question and uses explicit, preplanned scientific methods to identify, select, assess, and summarize similar but separate studies.* (pg. 82)

Consistent with the recommendations in the IOM report, the WWC applies evidence-based methodological standards consistently to each study it reviews. These standards, which are available on the WWC website, were developed by leading education research methodologists.

The systematic review conducted by the WWC goes beyond the procedures performed in meta-analyses. Again using the IOM definition, a meta-analysis "quantitatively combines the results of similar studies in an attempt to allow inference from the sample of studies included to the population of interest" (pg. 82). However, as Robert Slavin has noted, meta-analyses rarely describe even one study in any detail.[2] The WWC uses meta-analytic techniques to summarize the results of studies that meet WWC standards. However, the important distinction is that the WWC uses a rigorous set of standards, applied consistently, to determine *which* studies are included in the meta-analytic computations. This ensures that WWC summary measures are based only on studies with causal validity.

It is not surprising that WWC findings may differ from those of other analyses. Some meta-analyses may not have such rigorous standards for including studies; others may have standards that differ from those used by the WWC. Because of these differences, the WWC cannot judge the results of its systematic reviews by how they compare to other analyses. Rather, it works to ensure that its standards appropriately identify studies with strong causal validity and applies those standards consistently to each study reviewed.

---

[1] Institute of Medicine (IOM). *Knowing What Works in Health Care: A Roadmap for the Nation.* Washington, DC: The National Academies Press. 2008.

[2] Slavin, R.E. (1995). Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, Vol.48, No.1 pp.9-18.

Inclusion and exclusion procedures

Limiting studies to 1985 or later

The WWC's default time period for reviews is a study publication date of 1985 or later. This timeframe, which was established in 2005, is used for two reasons. First, by limiting reviews to research to this time period, WWC reviews reflect reasonably current research. In particular, the time period range ensures that effect sizes and improvement indices are based on a counterfactual condition that reflects classrooms as they operate within a recent time period. Second, the timeframe ensures that the research reviewed is examining versions of interventions that are most likely to be available to practitioners today.

WWC principal investigators have the option to expand the period for which studies can be reviewed, if they believe that important research will be excluded. The principal investigators for the Beginning Reading area chose to maintain the default period in large part to maintain currency with the classroom context for beginning readers. For instance, the fact that preschool enrollment has increased,[1] combined with the fact that more preschool and kindergarten programs run full-day,[2] means that students in the early grades may be better prepared to receive reading instruction today than students 25 years ago. Moreover, it is possible that any changes in reading readiness over this period have not been evenly distributed, since differences in reading ability by socioeconomic status and race are apparent at the kindergarten level.[3] Other contextual factors have changed over the past 20 years, including advances in teacher training, increases in home literacy activities, and changes in the content of and variety of curricula used in classrooms.

Any of these changes could have implications for the effectiveness of an intervention. If school readiness has increased, than an intervention that was effective 25 years ago may not be effective in more recent years. If teachers are receiving stronger training and using newer curricula, the counterfactual condition against which interventions were measured 25 years ago have changed, and possibly with it the magnitude of its effects. The Beginning Reading principal investigators judged that they had an inadequate basis for

---

[1] The proportion of 3 and 4 year olds enrolled in school increased from 37 percent in 1980 to 56 percent in 2006. See Snyder, T.D., Dillow, S.A., and Hoffman, C.M. *Digest of Education Statistics 2007*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2008.

[2] The proportion of pre-kindergartners and kindergartners attending full-day pre-kindergartners or kindergarten increased from 32 percent in 1980 to 59 percent in 2006. See Snyder, Thomas D. *Mini-Digest of Education Statistics, 2007*. Washington, DC: National Center for Education Statistics, Institute of Educational Sciences, U.S. Department of Education, 2008.

[3] For example, the Early Childhood Longitudinal Study of Kindergarteners (ECLS-K) found gaps in the reading knowledge and skills of kindergarteners by race: black and Hispanic children scored just under half of a standard deviation below whites on a test of reading knowledge and skills. Analysis of the ECLS-K and other surveys (i.e., Children of the National Longitudinal Study of Youth and the Infant Health and Development Program) show that socioeconomic status accounts for about half of the standard deviation of racial differences in reading test scores. See Duncan, G. and K. Magnuson. "Can Family Socioeconomic Resources Account for Racial and Ethnic Test Score Gaps?" *Future of Children*, Vol. 15, No. 1, Spring 2005.

assuming that effects of interventions measured more than 20 years ago would be experienced if schools adopted those interventions today.

Restrictive grade range for Beginning Reading studies

The letter raised concerns that the WWC's exclusion of research conducted on children outside of the kindergarten to 3rd grade age range was too restrictive. Because the reviews are focused on assessing interventions for beginning reading, the principal investigators, in coordination with the Institute of Education Sciences (IES), concluded that the review should focus on intervention effects for children in kindergarten through third grade (roughly ages 5 to 8). Some studies examine effects of interventions for students within the specified grade range and also students in higher grades. These studies are included in the review when the WWC is able to isolate the effects for the students who fall within the Beginning Reading grade range. The studies are excluded from the review if the results cannot be disaggregated to isolate effects for relevant grade range.

The grade range criterion is important for the integrity of the review process. The WWC-computed improvement indices and effect sizes are intended to reflect the effect of the intervention on the population in question. Including students above the topic area age range could lead to misstatements about intervention effects on children within the grade range.

It should be noted that the WWC attempts to determine a study's effects for the relevant grade range. When authors present findings aggregated across a broader age range and indicate that findings were analyzed for the relevant age range, it is standard WWC procedure to contact authors to request findings disaggregated for the grade range. Sometimes design limitations or other factors preclude authors from providing disaggregated results. In such cases, the WWC excludes the study.

Fidelity of treatment implementation

The letter notes that the WWC review process may downplay implementation fidelity. Definitions of implementation fidelity vary and many studies include little information to gauge fidelity, especially information about whether an intervention has been implemented within normal operating regimes of districts, schools, and teachers, not under specialized laboratory conditions. Moreover, there is no standard metric with which to rate and assess fidelity across studies that assures comparability.

The WWC's approach emphasizes the importance of replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies. Intervention reports include an "extent of evidence" classification that allows practitioners to place more weight if they choose on interventions for which the extent of evidence is large, meaning the results are drawn from multiple studies and a large number of classrooms and students.

Concerns about interventions and studies reviewed by the WWC

Reading Recovery

The letter expressed concern that Reading Recovery is an intervention outside of the Beginning Reading protocol. The Beginning Reading protocol states that interventions that target specific populations (for example, readers below grade level, and at-risk students) are eligible for the review. Reading Recovery is a short-term tutoring intervention program intended to serve the lowest achieving first-grade students (i.e., those in the bottom 20 percent). As such, it falls within the Beginning Reading protocol.

The letter expressed concerns that the reviews of studies of Reading Recovery mischaracterized the findings from those studies. For both the Baenen et al. (1997) and Iversen and Tunmer (1993) studies, the results presented by the WWC review represent the findings when the WWC standards and procedures are applied to these studies.

With respect to the Baenen et al. study, it is important to note that the beginning reading protocol prioritized one-year results. In effect, the Beginning Reading review is only intended to examine whether beginning reading interventions have an effect within one year. This one-year period is applied consistently to each study reviewed to ensure the results can be compared across studies and interventions. The Baenen et al. study's two- and three-year general reading achievement measures are not ignored, however; they are presented in Appendix A4.4 of the Reading Recovery Technical Appendices (http://ies.ed.gov/ncee/wwc/pdf/techappendix01_209.pdf).

With respect to the Iversen and Tunmer study, and consistent with the protocol, the WWC examined the results most relevant to the question of whether Reading Recovery improves reading proficiency compared to a reasonable counterfactual. That the study examined other comparisons is not ignored however. Appendices A4.1, A4.2 and A4.3 present results from other comparison groups. As with any study it reviews, the WWC does not base the findings of its review on the conclusions drawn by the authors.

We disagree that studies that compare an intervention to a no-treatment condition (as was done in Baenen et al.) provides a "built-in advantage," as the letter suggests. Many practitioners are interested in knowing whether an intervention is effective relative to customary classroom practices. The WWC reviews studies comparing treatments to no-treatment as well as studies comparing one treatment to another. In each case, the counterfactual is clearly documented in the review.

Exclusion of Reading Mastery Program

The letter expresses concern that the Reading Mastery Program is excluded from review by the WWC. The WWC has reviewed studies of Reading Mastery. A report summarizing the results of those reviews was published on the WWC website on August 12, 2008.

Inclusion of unpublished manuscripts for review

The letter suggests that by reviewing unpublished manuscripts, the WWC reviews studies of lower quality. Reviewing only studies that have passed a peer reviewing process could neglect unpublished information that may contain important and different findings. Mark Lipsey and David Wilson caution against this type of publication bias in *Practical Meta-Analysis* (2001). The damage such publication bias can cause in a systematic review is substantial. In *Knowing What Works in Healthcare*, the IOM notes that publication bias is well-established and that systematic reviews need to take steps to counter it, because otherwise "harmful interventions may appear to be worthwhile and beneficial interventions may appear to be useless" (page 97). For this reason WWC procedures are explicitly designed to review evidence from sources other than those published in journals. However, regardless of publication status, studies must meet the same evidence standards.

The letter asserts that the WWC should make available to the public all the research it reviewed. The more important issue is that the WWC provides an explicit citation to the study and thereby enables readers to obtain studies they wish to review. Unpublished information obtained from authors in response to queries that arise in reviews also are made available to requestors, and authors sign a form indicating that the information they provide to the WWC is available to the public.

Exclusion of Carlson and Francis (2002) study

The letter expressed concern over the WWC review's conclusion that there was a confound with the Direct Instruction intervention in Carlson and Francis (2002).[1] The WWC standard regarding intervention confounds was established to ensure that the results in a WWC review reflect what educators can expect if they implement the intervention being reviewed. A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program.

Exclusion of Waldron-Soler et al. (2002) study

As described in the protocols for the WWC Early Childhood Education and Beginning Reading reviews, to establish baseline equivalence, treatment and comparison groups must differ on the pretest measures by less than half a standard deviation, or the differences must be insignificant in an adequately powered statistical test. The Waldron-Soler (2002) study reported pretest differences exceeding half a standard deviation on at least two measures.

---

[1] The letter references Carlson and Freeman (2002). We have no record of a Carlson and Freeman (2002) study, and have assumed the letter refers to Carlson, C.D., & Francis, D.J. (2002). "Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE)." *Journal of Direct Instruction*, 3(1), 29-50.

The same standards were applied to the three Success for All studies you cite (Dianda and Flaherty, 1995; Ross, Alberg and McNelis, 1997; and Ross, McNelis, Lewis and Loomis, 1998). Unlike the Waldron-Soler study, pretest differences between the treatment and comparison groups were less than one half of a standard deviation.

Exclusion of Tobin (2003, 2004) studies

The Tobin (2003) study indicated that treatment and comparison groups were substantively different on pretest measures, with the differences exceeding half a standard deviation. Because the Tobin (2004) study follows the same students for additional time, it did not meet standards for the same reason.

The same standards were applied to the Success for All study cited in the letter (Smith, Ross, Faulks et al., 1993). Unlike the Tobin studies, pretest differences between the treatment and comparison groups were not statistically significantly different, and they were less than one half of a standard deviation. In keeping with WWC procedures, the WWC obtained information on pretest differences through communication with the study authors.

The "Create your own summary" feature

The letter expresses concern about the "create your own summary" feature of the WWC website because the results sort interventions by the magnitude of the improvement index. Because the improvement index calculations accurately reflect the application of the WWC evidence standards and effect size computations to the studies of those interventions, the WWC believes it is informative and useful to practitioners to sort results by improvement index.

The improvement index is one summary measure of the effect of beginning reading interventions. Users can also sort interventions alphabetically (by intervention name), by evidence rating, and by extent of evidence indices.

Appendix B
A Partial List of Studies of Reading Mastery and its Precursors Completed before 1985

Note: The letters in parentheses at the end of a citation refer to the source from which it was obtained. These sources are listed at the end of this appendix.

Apffel, J.A., J. Kelleher, M.S. Lilly, and R. Richardson. 1980. Developmental reading for moderately retarded children. Education and Training of the Mentally Retarded 10: 229-235. (b) (g)

Beck, I. L. and E. S. McCaslin. 1978. An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. LRDC Report No. 1978/6. Pittsburgh: University of Pittsburgh Learning Research and Development Center. (c)

Becker, W. C. and D. W. Carnine. 1980. Direct instruction: An effective approach to educational intervention with the disadvantaged and low performers. Pp. 429-473 in B.B. Lahey and A. E. Kazdin (eds.), Advances in Clinical Child Psychology, Volume 3, New York: Plenum. (a)

Becker, W. C. and S. Engelmann. 1976. Analysis of achievement data on six cohorts of low-income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model. Eric Document Reproduction Service No. ED 145922. (a)

Bock, G. and L. B. Stebbins. 1977. Education as experimentation: A planned variation model, Volume IV-B effects of Follow Through models. Cambridge, MA.: Abt Associates (ERIC Document Reproduction Service No. ED 148491). (a)

Booth, A., D. Hewitt, W. Jenkins, and A. Maggs. 1979. Making retarded children literate: A five-year study. The Australian Journal of Mental Retardation 5: 257-260. (d) (i)

Bowers, W. M. l972. An evaluation of a pilot program in reading for culturally disadvantaged first grade students. Doctoral dissertation, University of Tulsa, 1972. (Eric Document Reproduction Service No. ED 073439). (a) (b)

Bracey, S., A. Maggs, and P. Morath. 1975. Effects of a direct phonic apporach in teaching reading with six moderately retarded children: Acquisition and mastery learning stages. Slow Learning Child 22: 83-90. (b) (d) (i)

Branwhite, A. B. 1983. Boosting reading skills by Direct Instruction. British Journal of Educational Psychology 53: 291-298. (b) (c) (d) (g) (i)

Carnine, D. 1977. Phonics versus look-say: Transfer to new words. The Reading Teacher 30: 636-640. (h)

Carnine, D. 1980. Phonic versus whole-word correction procedures following phonic instruction. Education and Treatment of Children 3: 323-329. (h)

Carnine, D. and R. Gersten. 1983. Effectiveness of Direct Instruction in teaching selected reading comprehension skills: Preliminary Draft. Paper presented at American Educational Research Association (April 11-15, 1983). (b)

DuPree, T. J. 1976. Brief history of Cherokee schools, 1804-1976. BIA Education Research Bulletin 4: 3-11. (ERIC Document Retrieval Service No. ED 127051). (a)

Gersten, R., W.C. Becker, T. J. Heiry, and W. A. T. White. 1984. Entry IQ and yearly academic growth of children in Direct Instruction programs: A longitudinal study of low SES children. Educational Evaluation and Policy Analysis 6: 109-121. (e)

Gersten, R. and A. Maggs. 1982. Teaching the general case to moderately retarded children: Evaluation of a five-year project. Analysis and Intervention in Developmental Disabilities 2: 329-343. (i)

Gersten, R., A. Brockway, and N. Henares. 1983. The Monterey DI program for students with limited English (ESL). Direct Instruction News, 2(4): 8-9. (a)

Gordan, M.B. (ed.) 1971. DISTAR instructional system: Summaries of case studies on the effectiveness of the DISTAR instructional system. Chicago: Science Research Associates. (a)

Gregory, R. P. and B. G. Warburton (1983). DISTAR Reading and remedial children in an infant school. School Psychology International 4: 169-172. (i)

Haring, N.G. and D. A. Krug. 1975. Evaluation of a program of systematic instructional procedures for extremely poor retarded children. American Journal of Mental Deficiency 79: 627-631. (b) (i)

Kastner, S. and M. Hollingshead. 1973. An evaluation of ESEA Title I Programs, Community School District 15. New York: New York University, Center for Educational Research and Field Services, School of Education. (ERIC Document Reproduction Service No. ED 087842). (a)

Kaufman, M. 1973. The effect of the DISTAR Instructional System: An evaluation of the 1972-1973 Title I program of Winthrop, Massachusetts. (ERIC Document Reproduction Service No. ED 110171). (a)

Kaufman, M. 1974. The effect of the DISTAR Instructional System: An evaluation of the 1973-1974 Title I program of Winthrop, Massachusetts. (ERIC Document Reproduction Service No. ED 110170). (a)

Lewis, A. 1982. An experimental evaluation of direct instruction programme with remedial readers in a comprehensive school. Educational Psychology 2: 121-135. (g)

Lloyd, J., D. Cullinan, E.D. Heins, and M.H. Epstein. 1980. Direct instruction: Effects on oral and written language comprehension. Learning Disabilities Quarterly 3: 70-76. (g)

McCabe, T.A. 1974. The DISTAR Reading and Language Program: Study of its effectiveness as a method for the initial teaching of reading. Doctoral dissertation, University of Massachusetts, 1974. (ERIC Document Retrieval Service No. ED 102498).

Meyer, L. A. 1983. Long-term academic effects of Direct Instruction Follow Through. Technical report No. 299. Champaign: University of Illinois at Urbana-Champaign, Center for the Study of Reading. (ERIC Document Retrieval Service No. 237932). (a)

Newark (NJ) Board of Education. 1974. Program to improve the informational processoing of children with reading and learning problems. ERIC 106 826. (b)

Ogletree, E. J. 1976. A comparative study of the effectiveness of DISTAR and eclectic reading methods for inner-city children. Chicago: Chicago State University. (ERIC Document Reproduction Service No. ED 146544). (a) (b)

Ogletree, E.J. 1977. Does DISTAR meet the reading needs of inner-city kindergarten pupils? ERIC 146 303. (b)

Rawl, R. K. and F. S. O'Tuel. 1982. A comparison of three prereading approaches for kindergarten students. Reading Improvement 19: 205-211. (a) (b)

Richardson, E., B. Dibenedetto, A. Christ, M. Press, and B. Winsbert. 1978. An assessment of two methods for remediating reading deficiencies. Reading Improvement 15: 82-95. (a) (b) (d) (g) (i)

Sassenrath, J. M. and R. E. Maddux. 1974. Language instruction, background, and development of disadvantaged kindergarten children. California Journal of Educational Research 25: 61-68. (a)

Serwer, B. L., B. J. Shapior, and P. P. Shapiro. 1973. Comparative effectiveness of four methods of instruction on the achievement of children with specific learning disabilities. Journal of Special Education 7: 241-249. (b) (d)

Singer, B. 1973. The effects of structured instruction on kindergarten pupils: Final report. Washington, D.C.: U.S. Office of Education. (ERIC Document Reproduction Service No. ED 087564). (a)

Stebbins, L. B., R.G. Pierre, E. C. Proper, R. B. Anderson, and T. R. Cerva. 1976. Education as experimentation: A planned variation model (Vols. 3A-3B). Cambridge, MA: Abt Associates. (ERIC Document Reproduction Service No. ED 148489). (a)

Stein, C. L. and J. Goldman. 1980. Beginning reading instruction for children with minimal brain dysfunction. Journal of Learning Disabilities 13: 52-55. (b) (g) (i)

Summerell, S. and G. G. Brannigan. 1977. Comparison of reading programs for children with low levels of reading readiness. Perceptual and Motor Skills 44: 743-746. (a) (b) (d) (g)

Williamson, F. 1970. DISTAR reading: Research and experiment. (ERIC Document Reproduction Service No. ED 045318). (a) (b)

Note: Letters in parentheses refer to the source(s) providing the citation.

a. Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.
b. Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.
c. McGraw-Hill, 2002. "Appendix: Research supporting instruction in Reading Mastery: A Selected Annotated Bibliography." Results with Reading Mastery. ERIC ED 469 523.
d. Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.
e. Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.
f. NIFDI research staff found this citation independently.
g. Grossen, Bonnie. n.d. "References for Reading Mastery Reading Review," The Research Base for Reading Mastery, SRA, Eugene, Oregon: University of Oregon. (http://darkwing.uoregon.edu/~adiep/rmref.htm).
h. National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washing ton, D.C.: National Institute of Child Health and Human Development.
i. Kinder, D., R. Kubina, N.E. Marchand-Martella. 2005. Special education and Direct Instruction: An effective combination. Journal of Direct Instruction 5: 1-36.

Appendix C
Studies of Reading Mastery Published 1985 or Later But Not Included in the WWC Review

Note: The letters in parentheses at the end of a citation refer to the source from which it was obtained. These sources are listed at the end of this appendix.

Ball, E. and B. Blachman. 1991. Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? Reading Research Quarterly 26: 49-66. (d)

Bateman, B. 1991. Teaching word recognition to slow-learning children. Reading, Writing, and Learning Disabilities 7: 1-16. (b)

Brent, G. and N. DiObilda. 1993. Effects of curriculum alignment versus direct instruction on urban children. Journal of Educational Research 86: 333-338. (a)

Brett, A., L. Rothlein, and M. Hurley. 1996. Vocabulary acquisition from listening to stories and explanation of target words. Elementary School Journal 96: 415-422. (d)

Bryne, B. and R. Fielding-Barnsley. 1991. Evaluation of a program to teach phonemic awareness to young children. Journal of Educational Psychology 83: 451-455. (d)

Butler, P.A. 2003. Achievement outcomes in Baltimore City Schools. Journal of Education for Students Placed at Risk 8: 33-60. (e)

Chamberlain, L.A. 1987. Using DI in a Victoria, B.C. resource room. ADI News 7 (1): 7-8. (i)

Cole, K.N. Dale, P.S. and P. E. Mills. 1991. Individual differences in language delayed children's responses to direct and interactive preschool instruction. Topics in Early Childhood Special Education 11: 99-124. (i)

Cole, K. N., P.S. Dale, P.E. Mills, and J. R. Jenkins. 1993. Interaction between early intervention curricula and student characteristics. Exceptional Children 60: 17-28. (i)

Cooke, N.L., S.L. Gibbs, M. L. Campbell, and S. L. Shalvis. 2004. A comparison of Reading Mastery Fast Cycle and Horizons Fast Track A-B on the reading achievement of students with mild disabilities. Journal of Direct Instruction 4: 139-151. (i)

Cunningham, A. 1990. Explicit versus implicit instruction in phonemic awareness. Journal of Experimental Child Psychology 50 :249-444. (d)

Dale, P.S. and K.N. Cole. 1988. Comparison of academic and cognitive programs for young handicapped children. Exceptional Children 54: 439-447. (i)

Darch, C. and E. Kameenui. 1987. Teaching critical reading skills to learning disabled children. Learning Disability Quarterly 10: 82-92. (g)

Diobilda, N. and G. Brent. 1986. Direct Instruction in an urban school system. Reading Instruction Journal 29: 2-5. (e)

Dowdell, T. 1996. The effectiveness of Direct Instruction on the reading achievement of sixth graders. (ERIC Document Retrieval Service No. ED 396268). (a) (d)

Fredrick, L.D., M.C. Keel, and J. H. Neel. 2002. Making the most of instructional time: Teaching reading at an accelerated rate to students at risk. Journal of Direct Instruction 2: 57-63. (f)

Gersten, R. and D. Carnine. 1986. Direct Instruction in reading comprehension. Educational Leadership 44: 68-78. (b)

Gersten, R., C. Darch, and M. Gleason. 1988. Effectiveness of a Direct Instruction academic kindergarten for low-income students. The Elementary School Journal 89: 227-240. (e) (h)

Gersten, R., T. Keating, and W. Becker. 1988. The continued impact of the direct instruction model: Longitudinal studies of follow through students. Education and Treatment of Children 11: 318-327. (h)

Gunn, B., A. Biglan, K. Smolkowski, and D. Ary. 2000. The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. Journal of Special Education 2: 90-103. (c) (d)

Haskell, D. W., B. R. Foorman, and P. R. Swank. 1992. Effects of three orthographic/phonological units on first-grade reading. Remedial and Special Education 13: 40-49. (c) (d)

Herrera, J. A., C. H. Logan, P. G. Cooker, D. P. Morris, and D. E. Lyman. 1997. Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children. Learning disability or learning difference? Reading Improvement 34: 71-89. (d)

Jenkins, J. R., B. Matlock, and T. A. Slocum. 1989. Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. Reading Research Quarterly 24: 215-235. (d)

Juel, C. 1988. Learning to read and write: A longitudinal study of 54 children from first through fourth grade. Journal of Educational Psychology 80: 437-447. (d)

Juel, C. and C. Minden-Cupp. 2000. Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly 35: 458-492. (d)

Kamps, D.M., H.P. Wills, C.R. Greenwood, S. Thorne, J.F. Lazo, J.L. Crockett, J.M. Akers, and B. OL. Swaggart. 2003. Curriculum influences on growth in early reading fluency for students with academic and behavioral risk. Journal of Emotional and Behavioral Disorders 11: 211-224. (f)

Kaiser, S. K. Palumbo, R.C. Bialozor, and T. F. McLaughlin. 1989. Effects of Direct Instruction with rural remedial students: A brief report. Reading Improvement 26: 88-93. (b) (c) (g)

Keel, M.C., L. D. Frederick, T. A. Hughes, and S.H. Owens. 1999. Using paraprofessionals to deliver Direct Instruction reading programs. Effective School Practices 18: 16-22. (e)

Kuder, S.J. 1990. Effectiveness of the DISTAR reading program for children with learning disabilities. Journal of Learning Disabilities 23: 69-71. (b) (d) (i)

Kuder, S. J. 1991. Language abilities and progress in a Direct Instruction reading program for students with learning disabilities. Journal of Learning Disabilities 24: 124-127. (b) (d)

Leach, D. and S. Siddall. 1990. Parental involvement in the teaching of reading: A comparison of hearing reading, paired reading, pause, prompt, praise, and direct instruction methods. British Journal of Educational Psychology 60: 349-355. (h)

Ligas, M. R. and D. W. Vaughan. 1999. Alliance of quality schools: 1998-99 evaluation report. Broward , Florida: Broward County Schools. (a)

MacIver, M.A., E. Kedmper, and S. Stringfield. 2000. The Baltimore Curriculum Project: Fourth year report. Report for the Abell Foundation. Baltimore: Johns Hopkins University, Center for Social Organization of Schools. (a)

Marston, D., S. Deno, D. Kim, K. Diment, and D. Rogers. 1995. Comparison of reading intervention approaches for students with mild disabilities. Exceptional Children 62: 20-37. (h) (i)

Mathes, P.G. and T.J. Proctor. 1988. Direct Instruction for teaching “hard to teach” students. Reading Improvement 25: 92-97. (b)

Neely, M. 1995. The multiple effects of whole language, precision teaching, and Direct Instruction on first-grade story-reading. Effective School Practices 14: 33-42. (d)

O’Connor, R. E. and J. R. Jenkins. 1995. Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. The Journal of Special Education 29: 255-275. (i)

O’Connor, R., J. Jenkins, K. N. Cole, and P.E. Mills. 1993. Two approaches to reading instruction with children with disabilities: Does program design make a difference? Exceptional Children 59: 312-323. (b) (d) (i)

Rawl, R. K. and F. O'Tuel. 1982. A comparison of three prereading approaches for kindergarten students. Reading Improvement 19: 205-211. (d)

Reutzel, D. R. and P. M. Hollingsworthy. 1993. Effects of fluency training on second graders' reading comprehension. Journal of Educational Research 86: 325-331. (d)

Ross, S., J. Nunnery, E. Goldfeder, A. J. McDonald, R. Rachor, M. Hornbeck, et al. 2004. Using school reform models to improve reading achievement: A longitudinal study of Direct Instruction and Success for All in an urban district. Journal of Education for Students Placed at Risk 9: 357-389. (e)

Salerno, C. 1992. A comparison of classrooms using a meaning-centered approach and a code-centered approach. ADI News 11 (2): 26-29. (d)

Sexton, C. W. 1989. Effectiveness of the DISTAR Reading I Program in developing first graders' language skills. Journal of Educational Research 82: 289-293. (a) (b) (d) (g)

Smith, S., D. Simmons, M. Gleason, E. Kameenui, S. Baker, M. Sprick, et al. 2001. An analysis of phonological awareness instruction in four kindergarten basal reading programs. Reading and Writing Quarterly 17: 25-50. (d)

Snider, V. E. 1990. Direct Instruction reading with average first-graders. Reading Improvement 27: 143-148. (b) (g) (h)

Stein, M. 1990. *Reading research ... and Reading Mastery*. Chicago: SRA. (b)

Tobin, K. 2000. A comparison between Horizons Fast Track A-B and Silver, Burdett, and Ginn reading curricula in first grade: June 2000 final report. Oregon: National Institute for Direct Instruction. (a) (e)

Torgesen, J., S. Morgan, and C. Davis. 1992. Effects of two types of phonological awareness training on word learning in kindergarten children. Journal of Educational Psychology 84: 364-370. (d)

Traweek, D. and Berninger, V. 1997. Comparisons of beginning literacy programs: Alternative paths to the same learning outcome. Learning Disability Quarterly 20: 160-168. (a) (d) (h)

Umbach, B. T., C. Darch, and G. Halpin. 1987. Teaching reading to low performing first graders: A comparison of two instructional approaches. ERIC 290 130. (b)

Umbach, B., C. Darch, and G. Halpin. 1989. Teaching reading to low performing first graders in rural schools: A comparison of two instructional approaches. Journal of Instructional Psychology 16: 23-30. (h)

Urdegar, S. M. 1998. Evaluation of the Success for All Program, 1997-98. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research. (a)

Urdegar,k S. M. 2000. Evaluation of the Success for All Program, 1998-99. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research. (a)

Weinstein, G. and N. L. Cooke. 1992. The effects of two repeated reading interventions on generalization of fluency. Learning Disability Quarterly 15: 21-28. (d)

Varela-Russo, C., K. A. Blasik, and M. Ligas. 1998. Alliance of quality schools evaluation report. Ft. Lauderdale, FL: School Board of Broward County. (a)

Yu, L. and R. Rachor. 2000. The two-year evaluation of the three-year Direct Instruction program, in an urban public school system. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 441831). (a) (e)

Note: Letters in parentheses refer to the source(s) providing the citation.

a. Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.
b. Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.
c. McGraw-Hill, 2002. "Appendix: Research supporting instruction in Reading Mastery: A Selected Annotated Bibliography." Results with Reading Mastery. ERIC ED 469 523.
d. Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.
e. Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.
f. NIFDI research staff found this citation independently.
g. Grossen, Bonnie. n.d. "References for Reading Mastery Reading Review," The Research Base for Reading Mastery, SRA, Eugene, Oregon: University of Oregon. (http://darkwing.uoregon.edu/~adiep/rmref.htm).
h. National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washing ton, D.C.: National Institute of Child Health and Human Development.
i. Kinder, D., R. Kubina, N.E. Marchand-Martella. 2005. Special education and Direct Instruction: An effective combination. Journal of Direct Instruction 5: 1-36.

Appendix D
The Impact on Summary Ratings of Ignoring Fidelity of Implementation

As noted in the text of this report (p. 11), the WWC has indicated that they have chosen to ignore ratings of fidelity of implementation and, instead, to rely on "replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies" (Dynarski, 2008, p. 3, in Appendix A of this report). As noted in the text, this assumption is seriously flawed. In fact, for both exemplary programs and for programs that are ineffective, poor implementation of a program would, very likely, produce results that are systematically biased. This has very serious consequences for consumers, leading to minimizing the positive impact of good programs as well as the negative impact of poor programs.

To illustrate these results the tables below examine the possible implications of inadequate fidelity of implementation for two different hypothetical curricular programs: one that is more effective than a control group and one that is less effective than the control group. We use effect sizes (the difference between the mean scores of experimental and control groups divided by the common standard deviation) as the metric of comparison. Basic to our analysis is the assumption of "regression toward the mean," the well established statistical phenomenon where those with high scores (or low scores) on a measure will tend to have scores that are closer to the mean (lower for those who are high and higher for those who are low) at later testing periods. Simply through regression toward the mean we would expect those who score lower than a mean at pretest to score higher (and closer to the mean) at posttest. For sake of illustration we will assume that the scores involved are normal curve equivalent (NCE) scores with a control group mean of 50 and a standard deviation of 21.

Consider first a program that is effective. The first column of Table D-1 gives four possible values of the experimental group that we will assume are the "true" values and the second column gives the value of Cohen's d that results when comparing these experimental means with those of the control group. It can be seen that the d values vary from .24, when the experimental mean is 55, to .95, when the experimental mean is 70. The higher values in this column are, in fact, similar to those often obtained in studies of Direct Instruction, especially when it is implemented with fidelity.

The other columns in part A of Table D-1 examine the likely result if the effective program is implemented with less than optimal fidelity, but the impact is simply the "random" effects that the WWC assumes will occur. In these calculations the means of the experimental group do not change, as implied by the WWC assumption. However, the standard deviation is assumed to change, for with lower levels of fidelity, even if the means stay constant, greater variability would be expected. This is at the basis of the WW notion of "averaging out." Three different values of the standard deviation are given, and it can be seen that in all situations the value of the effect size becomes lower than the "true" value, often substantially so. (Compare the "true" values in the second column with the other values of d going across each row.)

Panel B of Table D-1 presents results that are probably more likely if an effective program is implemented with less than adequate fidelity. In such a situation it would be logical to expect that performance would be lower – that the mean value would decline. As with Panel

A, the first two columns of Panel B give the “true” experimental mean and effect size for comparison purposes. The next two columns present the effect sizes that would occur if the average value of the experimental group declined by either 5 points (column 3) or by 10 points), but assuming no change in the standard deviation. Again, all of the effect sizes are smaller. Finally, the calculations in the last three columns of Panel B assume that lower fidelity with an effective program has two effects: the average score is lower than it would otherwise be and the standard deviation is larger, as in Panel A. Again, of course, the effect sizes are substantially lower and the declines are marked.

Table D-1: Effect Sizes of Comparisons with Effective Programs with Different Fidelity Conditions

A. Random Influences of Fidelity Problems

Mean, Experimental Group	"True" d	s.d. = 25	s.d. = 30	s.d. = 35
55	0.24	0.20	0.17	0.14
60	0.47	0.40	0.33	0.29
65	0.71	0.60	0.50	0.43
70	0.95	0.80	0.67	0.57

B. Systematic Influences of Fidelity Problems

		Systematic - Experimental mean smaller, s.d. the same		Systematic - Exp Mean smaller, s.d. larger		
Mean, Experimental Group	“True” d	Mean 5 points less	Mean 10 points less	mean 5 points less, s.d. = 25	mean 10 points less, s.d. = 25	mean down by 10 points, s.d. = 30
55	0.24	0.00	-0.24	0.00	-0.20	-0.17
60	0.47	0.24	0.00	0.20	0.00	0.00
65	0.71	0.47	0.24	0.40	0.20	0.17
70	0.95	0.71	0.47	0.60	0.40	0.33

Note: If the impact of fidelity implementation is random, it is assumed that this affects only the common standard deviation and not the mean. If the impact of fidelity implementation is systematic, both the mean and the standard deviation can be affected. Cohen's d is calculated by subtracting the mean of the experimental group from the mean of the control group and dividing by the common standard deviation. The "true" calculations assume that the standard deviation is 21, the value of the control group = 50 and experimental mean varies as shown in each of the rows of the table.

Now consider the possible implications of poor fidelity of implementation of a program that is, in reality, ineffective and, in fact, produces poorer results than the control group. These results are shown in Table D-2. As with Table D-1, the results with the “true” values are shown

in the first two columns of each panel. It can be seen that the "true" effect sizes vary from -.24 to -.95.

The results in Panel A illustrate what would happen if the fidelity problems produce only random changes: the means would stay the same, but the standard deviations would become larger. It can be seen that, in all cases, the absolute values of the effect sizes become smaller – that is, less negative.

Table D-2: Effect Sizes of Comparisons with Ineffective Programs with Different Fidelity Conditions

A. Random Influences of Fidelity Problems

Mean, Experimental Group	"True" d	s.d. = 25	s.d. = 30	s.d. = 35
45	-0.24	-0.20	-0.17	-0.14
40	-0.47	-0.40	-0.33	-0.29
35	-0.71	-0.60	-0.50	-0.43
30	-0.95	-0.80	-0.67	-0.57

B. Systematic Influences of Fidelity Problems

		Systematic - Experimental mean larger, s.d. the same		Systematic - Experimental Mean higher, s.d. smaller		
Mean, Experimental Group	"True" d	mean 5 points higher	mean 10 points higher	mean 5 points higher, s.d. = 18	mean 10 points higher, s.d. = 18	mean 10 points higher, s.d. = 15
45	-0.24	0.00	0.24	0.00	0.28	0.33
40	-0.47	-0.24	0.00	-0.28	0.00	0.00
35	-0.71	-0.47	-0.24	-0.56	-0.28	-0.33
30	-0.95	-0.71	-0.47	-0.83	-0.56	-0.67

Note: If the impact of fidelity implementation is random, it is assumed that this affects only the common standard deviation and not the mean. If the impact of fidelity implementation is systematic, both the mean and the standard deviation can be affected. Cohen's d is calculated by subtracting the mean of the experimental group from the mean of the control group and dividing by the common standard deviation. The "true" calculations assume that the standard deviation is 21, the value of the experimental mean varies as shown in each of the rows of the table.

The results in Panel B illustrate the results if the impact of poor fidelity is systematic. With programs that are ineffective, it would be expected that poor implementation would lead to higher average scores (through regression to the mean) and smaller standard deviations. The

smaller standard deviations result from having values that are "less ineffective." As can be seen in Table D-2, the result is that the effect sizes are less negative. That is, the extent to which the programs are truly ineffective, as shown by the "true" values of d, is disguised.

The results presented above seem to clearly refute the assumption that guides the WWC's approach to considering the fidelity of implementation in deliberations. Even if the results of poor fidelity of implementation were random, the efficacy of both good and poor programs would be misrepresented. If, as is more likely, the results of poor implementation are not random, the impact would be even greater. Good programs appear less effective and poor programs appear better than they actually are. The WWC's decision regarding consideration of the fidelity of implementation appears to be seriously misguided.

Subject: RE: Machinations of What Works Clearinghouse

-----Original Message-----
From: Sanchez, Susan [mailto:Susan.Sanchez@ed.gov]
Sent: Wednesday, October 08, 2008 12:43 PM
To: Mark Dynarski; Jill Constantine; Scott Cody
Subject: FW: Machinations of What Works Clearinghouse

Fyi. Will be following up with you.

-----Original Message-----
From: Cottingham, Phoebe
Sent: Wednesday, October 08, 2008 8:49 AM
To: Sanchez, Susan
Subject: FW: Machinations of What Works Clearinghouse

fyi

----- Original Message -----
From: zig@nifdi.org <zig@nifdi.org>
Sent: Tue Oct 07 20:24:53 2008
Subject: Machinations of What Works Clearinghouse

Attached are two papers I think will interest you. Both deal with the same
issue, the casuistry of What Works Clearinghouse. The background for the papers is that WWC reported that it found no studies for beginning reading
that provided evidence of effectiveness for Reading Mastery. The research director for The National Institute for Direct Instruction, Jean Stockard,
wrote a letter to WWC challenging some of the reviews of Reading Mastery studies. WWC responded on September 8, 2008 with a letter that defended every issue that Jean brought up in her earlier letter. Mathematica concurred that WWC was justified on every point made in the defense letter.

I wrote a critique of the Sept. 8 letter. It appears as attachment A. This
paper focuses on the logic of WWC's assertions and basically shows that the probability that the rejection of all Reading Mastery studies could not have happened by chance. The paper also shows the illogic and contradictions in WWC's rationales and assertions.

Jean Stockard wrote a response to WWC, which appears as attachment B. Her paper provides more empirical data, including some statistical models that
contradict WWC's assertions. (See her appendix D for instance. I think it is very clever.)

My hit is that the scientific community should rain down condemnation on WWC. Let me know what you think. Is it time for the Inspector General to do something to earn his salary and benefits?

Any help, comments, or criticisms are greatly appreciated.

Thanks.

Siegfried Engelmann

Reading Achievement in a Direct Instruction School and a "Three Tier" Curriculum School

Jean Stockard
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

November 2008

Technical Report 2008-05



© 2008 National Institute for Direct Instruction
PO Box 11248 • Eugene, OR 97440
1-877-485-1973

Table of Contents

	Page
Executive Summary	iii
List of Tables	iv
List of Figures	v
Report	1
Oral Reading Fluency	1
Other Measures	3
Summary	4
Technical Appendix	6
Methodology	6
The Sample	6
Measures	7
Analysis	8
Results	9
Oral Reading Fluency	9
Other Indicators	13

Reading Achievement in a Direct Instruction School and a "Three Tier" Curriculum School

Executive Summary

This report examines data from two schools within the same Oregon school district. One school adopted the *Reading Mastery* Direct Instruction program as the core reading curriculum for all primary children, while the other used a "three tiered" model, occasionally employing DI for students that teachers felt would benefit from the instruction.

Data were available for 2 cohorts of students who were in the schools from kindergarten through third grade. There were almost equal numbers of students from each school and there were no significant differences between the schools in the students' eligibility for free or reduced lunch, their racial-ethnic characteristics, or their special education designation. A slightly modified version of DIBELS was used to assess achievement gains over the time span of the study.

Key findings include:

- Students in the Direct Instruction School had statistically significant higher gains in Oral Reading Fluency (ORF) from first through third grade than students in the Control School. These differences were especially marked for students in special education.
- Almost all of the effect sizes comparing ORF of the two groups surpassed the usual criterion of educationally significant (.25). By the end of third grade, the effect size for ORF for the total group of students was .42, while the effect size for special education students was .73.
- The odds of a child in the DI school having levels of oral reading fluency that would indicate a possibility of academic failure were less than half that of a child in the control school.
- Similar results occurred with the measures of onset recognition fluency, phonemic segmentation fluency, and nonsense word fluency, consistently favoring the Direct Instruction students.
- There were no differences between the groups in letter naming fluency between the two groups in kindergarten. This result was expected because the DI curriculum does not emphasize learning of letter names at that level.

Results add to the literature that documents the higher levels of achievement and stronger achievement gains of students receiving Direct Instruction. They also indicate the extent to which having Direct Instruction as a core curriculum enhances achievement of all students, both those receiving special education and those only in general education programs.

List of Tables

	Page
1. Characteristics of the DI and Control School, 2001-02	6
2. DIBELS Measures used in analysis, number and year of administration	8
3. Means and standard deviations, oral reading fluency measures, by school and grade, total sample and special education students	9
4. Analysis of variance, ORF by school and special education status	11
5. Percentage of students “at risk” and “at low risk” of not meeting literacy goals by school and administration	12
6. Analysis of variance and descriptive results for letter naming fluency, onset recognition fluency, phonemic segmentation fluency, and nonsense word fluency with schools and special education status as factors	14

List of Figures

Page

1. Oral Reading Fluency, First to Third Grade, Non-Special Education Students 1

2. Oral Reading Fluency, First to Third Grade, Special Education Students 2

3. Percentage of Students at Low Risk of Failure, End of School Year, by Grade and Group 3

4. Percentage of Special Education Students at Low Risk of Failure, End of School Year, by Grade and Group 3

Reading Achievement in a Direct Instruction School
and a "Three Tier" Curriculum School

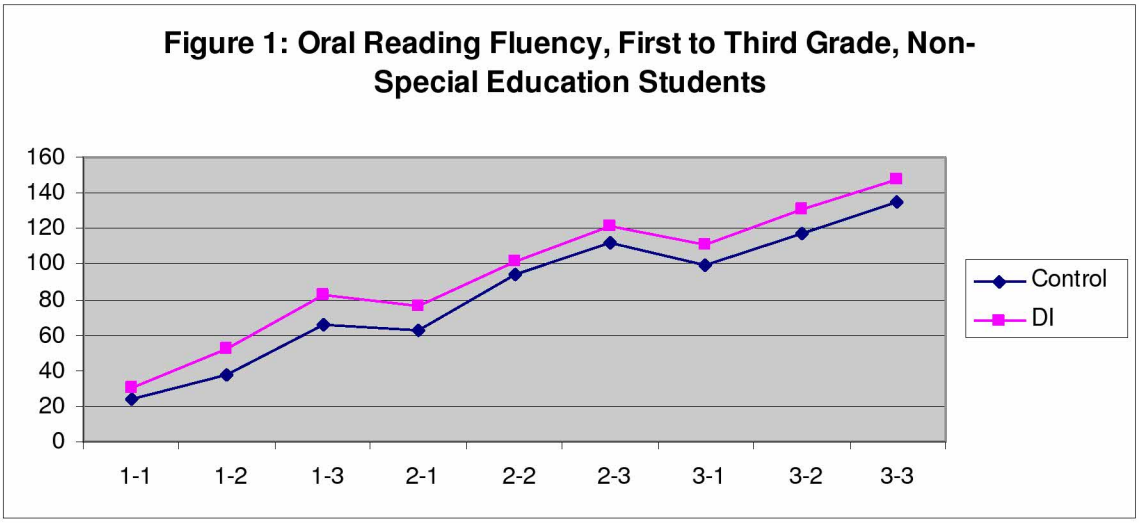
This report examines data from two elementary schools in the Bethel, Oregon, school district. One school adopted the Direct Instruction program, *Reading Mastery*, as the core reading curriculum for the primary grades. The other, termed the Control School, used a traditional curriculum, employing a "three tiered" model and occasionally using Direct Instruction for students that teachers felt would benefit from the program. This report examines the growth in reading achievement of two cohorts of students from kindergarten through third grade in each of these schools. Full details on the methodology and the statistical results are included in a technical appendix.

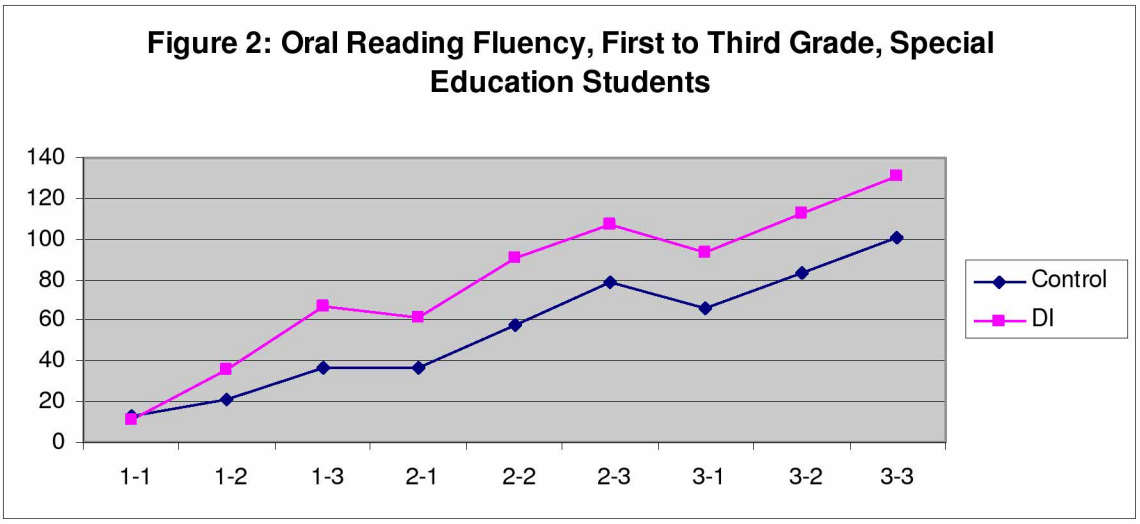
The analysis focuses on 169 students who were enrolled in their respective schools from Kindergarten through third grade. There were no differences between the students in the two schools in their race-ethnicity, receipt of free or reduced lunch, or proportion receiving special education services. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores were routinely gathered to monitor students' progress. The DIBELS measures have high statistical reliability and can be compared against established benchmarks that indicate the level at which students should achieve to reach generally accepted literacy goals.

Oral Reading Fluency

The most important DIBELS measure is Oral Reading Fluency (ORF), which taps students' ability to accurately and fluently read short passages. Figures 1 and 2 illustrate the Oral Reading Fluency scores of students in the Control and Direct Instruction schools from the beginning of first grade through the end of third grade. Figure 1 gives data for the students in general education, and Figure 2 gives data for students in Special Education. In early first grade the students in the two schools had similar average scores,

Figure 1: Oral Reading Fluency, First to Third Grade, Non-Special Education Students





but over time the students in the Direct Instruction School had significantly stronger gains, on average, than those in the control schools. The difference in these gains occurred for both the general education and special education students, but was significantly stronger among the special education students. All of the differences were both statistically significant and substantively strong.

Effect sizes are a commonly used metric to summarize differences between two groups. As explained more fully in the Appendix, effect sizes of .25 or larger have traditionally been considered "educationally significant." At the beginning of first grade the effect sizes summarizing the difference between the two groups were quite small. This would be expected given the groups' general equivalence at the start of their schooling experience. But, after that time with greater exposure to the chosen curricula, the effect sizes became quite large, ranging up to .94 for special education students. Only one of the 24 effect sizes calculated for data after the beginning of first grade failed to meet the .25 criterion of educational significance.

Another way of examining the differences between the groups is to compare students' Oral Reading Fluency scores to standard benchmarks for adequate progress. Figures 3 and 4 summarize this information for the total group of students and for special education students. The data in these figures indicate the percentage of students whose ORF scores at the end of each grade indicated that they were at "low risk" for failure in the future. In other words, the data tell the percentage of students who were "on track" for success, as compared to other students in their grade throughout the nation.

The results are clear. For both the entire group of students and for those in special education, the students in the Direct Instruction school were much more likely at each grade to be on track for future success. In the DI school over three-fourths of all the students and well over half of the special education students were meeting benchmarks at the end of each grade that would predict future success. Corresponding figures in the Control School were much lower, especially among the special education students. Calculations included in the appendix indicate that the odds that a child in a DI school

would be at risk of future academic failure were less than half, and often far less than half, that of a child in the Control school.

Figure 3: Percentage of Students at Low Risk of Failure, End of School Year, by Grade and Group

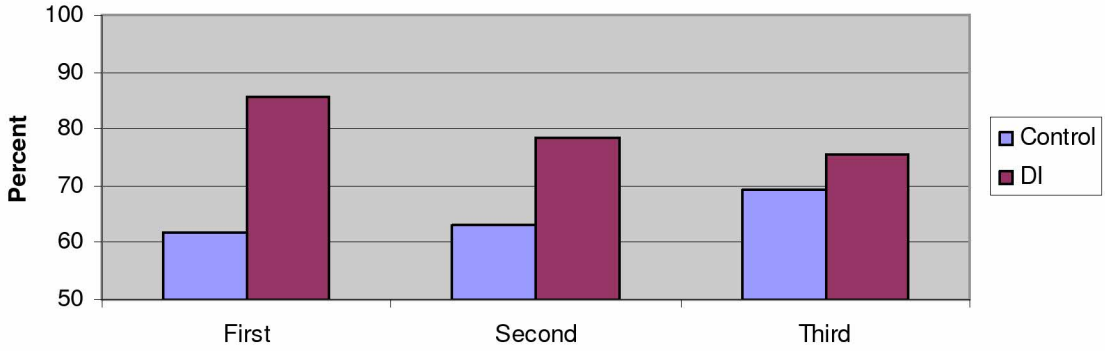
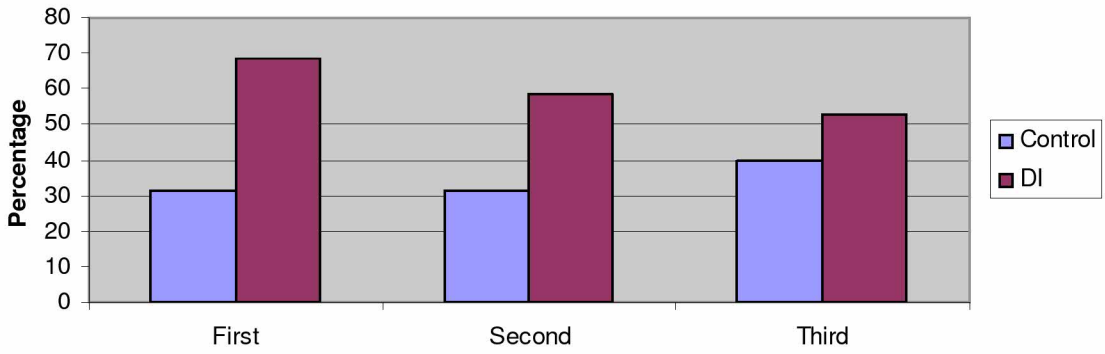


Figure 4: Percentage of Special Education Students at Low Risk of Failure, End of School Year, by Grade and Group



Other Measures

Several other DIBELS measures were administered to students in kindergarten and first grade, and the results with these measures support the findings summarized above. Two measures were administered from mid-year in kindergarten through the end of first grade: Phonemic Segmentation Fluency (PSF), which measures the ability to separate words into phonemes, and Nonsense Word Fluency (NWF), which measures the ability to read phonetic nonsense words. Students in the Direct Instruction school had significantly greater gains from kindergarten through the end of first grade than students in the Control School in their scores on both of these measures. The gains in phonemic segmentation were especially large for special education students in the DI school.

One measure was administered only at the beginning and middle of the kindergarten year: Onset Recognition Fluency (OnRF), which taps the ability to recognize the initial sound of words. Students in the DI school had slightly higher scores on this measure, but the differences were not statistically significant. Finally, a measure of Letter Naming Fluency (LNF) was administered during kindergarten and at the beginning of first grade. The Direct Instruction curriculum does not emphasize learning letter names during those years, for research indicates that instruction in letter names at an early age is largely unrelated to later reading achievement. As expected, students in the Control school had significantly higher scores on the LNF measure.

Summary

This report analyses the achievement of students enrolled from kindergarten through third grade in two schools in the same Oregon district. One of the schools instituted Direct Instruction as a core primary reading curriculum, while the other occasionally used DI as a supplementary program for students deemed in need of additional help. There were no significant differences between the students in the two schools in race ethnicity, eligibility for free or reduced lunch, or special education status.

Data from regularly administered DIBELS tests were available on all students in two cohorts from first through third grade. Statistical analyses indicated strong and statistically significant results in favor of the DI curriculum. Numerous calculations indicate that the differences were substantively strong and educationally meaningful. Students in the DI school had significantly greater gains in oral reading fluency, nonsense word fluency, and phonemic segmentation fluency, the skills most closely related to continued success in reading. The changes in onset recognition fluency also favored DI, but were not statistically significant. Changes in letter naming fluency, a skill not shown by the literature to be related to later success in reading and not included in the Direct Instruction curriculum in the first months of instruction, were greater in the control school. This result was expected. Changes in favor of direct instruction were stronger among special education students for both oral reading fluency and phonemic segmentation fluency.

The presence of numerous data points provides additional strength for these conclusions. The superior achievement of students in the Direct Instruction program was found with different measures and at virtually all testing periods. The effect sizes were consistently strong and educationally important, resulting in students in the DI school being at substantially lower risk for educational failure in future years.

These results appear to provide additional evidence in support of the efficacy of Direct Instruction as a core curriculum for all students. The school that employed the Direct Instruction program *Reading Mastery* as the core curriculum had significantly greater student gains over time and fewer students at risk for problems in the future. These results appeared with all students, both those in general education and those in special education.

In addition, for some of the measures the gains in achievement were especially strong for students in special education. As such, the results indicate that special education students may be better served in schools that employ Direct Instruction as a core curriculum rather than by simply use it as one of a variety of possible interventions in a "three tiered" model. That is, the systematic implementation of Direct Instruction within the entire school appears to be better for both the general population and the special education population of students.

Technical Appendix

Methodology

The discussion below describes the sample as well as the measures and analysis that were employed.

The Sample

Characteristics of the two schools, as of 2001-02, the year before the first year of data availability, are given in Table 1. The DI school included a broader span of grades (K through 8) than the control school (K through 5), and had slightly more students at each grade level (mean = 78) than the control school (mean = 68). In addition, the DI school had a slightly higher percentage of students on free and reduced lunch and lower scores on the statewide reading assessments (from students in grades 3 and 5). All of these factors could indicate that the DI school had a student body that was slightly more at risk. In contrast, however, the attendance rates of the two schools were almost identical, and the DI school had a lower percentage of minority students enrolled.

Table 1: Characteristics of the DI and Control School, 2001-02

	DI	Control
Grade Range	K-8	K-5
Enrollment, 2001-2	704	407
Free and reduced Lunch (%)	34.1	26.8
Minority enrollment (%)	9.8	17.9
Oregon State Assessment Reading Score, 2002	94.3	107.1
Attendance Rate, 2001-02 (%)	93.8	94.9

As noted in the text, the Direct Instruction School used *Reading Mastery* as the core curriculum in grades K-3. In fourth grade DI was used occasionally, but was not the core curriculum. In addition, in kindergarten, students in the Title I program received additional instruction with another phonics program (Optimize, now called Scott Foresman Reading). The control school did not implement Direct Instruction except for a few children who were judged as needing additional instruction. This was part of a "three tiered" plan where different levels of intervention were used depending upon the perceived needs of the students. Our analysis should be seen as an examination of the extent to which adopting DI as the core curriculum within a school can produce changes in the average levels of achievement and achievement gain over time.

In our analysis we were not able to ascertain which students received the supplemental instruction in Direct Instruction within the control school. Thus, the analysis should be seen as a conservative test of Direct Instruction, for some of the students within the control school, most likely those who were receiving special

education, may have had Direct Instruction. Any bias that would result from the few students in the control school that received DI would act against the possibility of significantly higher achievement in the DI schools. In addition, it would be difficult to argue that the bias that could be introduced by additional phonics instruction for a few students in the DI schools in kindergarten would persist through third grade, the end of our data collection period.

Data were available from 2 cohorts of students who were in the schools from kindergarten through third grade: 72 children who began kindergarten in 2002-03 and 97 who began kindergarten in 2003-04. Students were almost evenly split between the two schools with 85 in the control school and 84 in the DI school. Analyses with data within each school indicated no significant differences between the two cohorts on demographic characteristics, achievement measures or achievement growth. Thus the two cohorts are considered jointly.

Eighty-nine percent of the students were non-Hispanic whites, 29 percent received free or reduced lunch, and 20 percent were in special education. Differences between the students in the two schools were not statistically significant on any of these measures. That is, the students in the analysis from the two schools were very similar in race-ethnicity, poverty status, and probability of assignment to special education.

Measures

In both schools data were routinely gathered using a slightly modified version of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). The DIBELS measures are a set of standardized, individually administered indicators of early literacy development that take only about one minute to administer, can be used to regularly monitor the development of both pre-reading and early reading skills, and have very high reliability. All measures result in numeric scores that indicate the number of correct responses given in one minute. Benchmarks have been established to indicate where students should be at the beginning, middle, and end of a school year to be "at risk," at "some risk," or at "low risk" of not achieving established literacy goals.

Table 2 lists the measures that were used and the times of administration. A measure of letter naming fluency (LNF) was given three times in kindergarten and once in first grade. Because fluency in letter naming is not a necessary precursor to reading, it is not a core element of the DI curriculum in kindergarten. Thus, we did not expect to find any differences between the control and treatment school on this measure.

The other measures are more directly related to reading. Students exposed more consistently to the DI curriculum, as in the treatment school, would be expected to have higher scores and greater gains on these measures. The lowest level of these skills, recognizing the beginning sound of words, is measured by the onset recognition fluency (OnRF) measure, which was administered two times in the kindergarten year. Two measures were administered twice in kindergarten and three times in first grade: phonemic segmentation fluency (PSF), which measures the ability to separate words into

phonemes; and nonsense word fluency (NWF), which measures the ability to read phonetic nonsense words.

The most important measure is oral reading frequency (ORF), a measure of how fluently a child reads correctly. This measure was administered three times a year beginning in first grade. The ORF was the major focus of our analysis, although results are also presented for the other measures.

Table 2: DIBELS Measures Used in Analysis and Year of Administration

	Kindergarten	First Grade	Second Grade	Third Grade	Total
Letter naming fluency	3	1			4
Onset recognition fluency	2				2
Phonemic segmentation fluency	2	3			5
Nonsense word fluency	2	3			5
Oral reading fluency		3	3	3	9
Total number of measures	9	10	3	3	25

Note: Numbers in the table refer to the number of administrations that occurred for a given test in a given year.

Analysis

Students' scores on the first administration of each test were compared between those with and without free and reduced lunch, special education status, and minority status. The only characteristic consistently related to differences in the initial scores was special education status, with special education students always having lower scores. Thus, in the following analyses, students' special education status is included as a control variable. If DI is an effective treatment, it would be expected that it would impact both regular and special education students.

Two different techniques were used to analyze the data. The first is a repeated measures analysis of variance with the test scores as repeated measures and the school, or treatment group, as well as special education status as factors. We would expect a significant linear effect of treatment, with children gaining proficiency over time. If having DI as a core curriculum is more effective than the three tier model, we would also expect a significant two-way interaction between these over time changes and school, with greater gains for the treatment school. If school wide implementation of DI is more effective with special education students than the three tier model we would expect a significant three way interaction (school by special education by DIBELS measure). Mean scores and Cohen's d, as a measure of effect size, are included to provide descriptive understandings of the trends.

For the measure of oral reading fluency (ORF), arguably the most important indicator and the one for which we have the most data, we also calculated whether children's scores placed them "at risk" and whether they were "at low risk" for not meeting literacy goals. We compared the percentage of students in each category across the two schools using simple chi-square tests and also calculated the odds of children falling into these categories.

Results

The results with oral reading fluency (ORF) are presented first, followed by the results with the other measures.

Oral Reading Fluency

Table 3 gives the means and standard deviations for the ORF scores for each school and each administration for both the total sample and by special education status. Table 4 gives the results of the analysis of variance. The three-way interaction (ORF by school by special ed status) is statistically significant ($p < .001$), indicating that the changes over time in ORF differ both by school and by special education status.

Examining the means, one can see that this interaction results from the very strong gains of special education students in the DI school. The total group of students in the DI school had stronger gains, on average, than those in the control school, but these gains were even stronger among the special education students. On average, students in the control school gained 106 words per minute on the ORF from the beginning of first grade to the end of third grade, while those in the DI school gained 116 words (Panel A). Among special education students, those in the control school only gained 88 words, while those in the treatment school gained 119 words (Panel B). Similar results appeared with the non-special education students, although the differences weren't quite as dramatic: those in the control school gained an average of 111 words while those in the treatment school gained, on average, 117 words (Panel C).

Table 3: Means and Standard Deviations, Oral Reading Fluency Measures, by School and Grade, Total Sample and by Special Education Status

A. Total Sample	**Control School**			**DI School**			
	Mean	S.D.	N	Mean	S.D.	N	Cohen's D
1st grade, first of year	23	29	60	28	31	56	0.17
1st grade, mid year	34	31	84	49	34	83	0.46
1st grade, end of year	60	37	84	80	38	83	0.52
2nd grade, first of year	58	34	85	73	35	83	0.44
2nd grade, mid year	87	37	84	99	37	83	0.33
2nd grade, end of year	105	33	84	118	35	83	0.38
3rd grade, first of year	93	34	81	107	38	79	0.39
3rd grade, mid year	111	34	80	127	39	79	0.43
3rd grade, end of year	129	34	78	144	38	78	0.42
Difference 1-1 to 3-3	106			116			

B. Special Education Students

	Control School			DI School			
	Mean	S.D.	N	Mean	S.D.	N	Cohen's D
1st grade, first of year	13	12	8	11	9	7	-0.13
1st grade, mid year	21	23	16	36	27	16	0.59
1st grade, end of year	37	26	16	67	39	16	0.94
2nd grade, first of year	37	25	16	61	32	16	0.85
2nd grade, mid year	58	31	16	91	48	16	0.84
2nd grade, end of year	79	29	16	107	42	17	0.81
3rd grade, first of year	66	29	15	93	42	17	0.77
3rd grade, mid year	84	32	15	113	42	17	0.79
3rd grade, end of year	101	37	15	130	45	17	0.73
Difference 1-1 to 3-3	88			119			

C. Non-Special Education Students

	Control School			DI School			
	Mean	S.D.	N	Mean	S.D.	N	Cohen's D
1st grade, first of year	24	30	52	30	32	49	0.19
1st grade, mid year	37	32	68	52	34	67	0.45
1st grade, end of year	66	37	68	83	37	67	0.45
2nd grade, first of year	63	34	69	76	35	67	0.38
2nd grade, mid year	94	35	68	101	34	67	0.21
2nd grade, end of year	112	31	68	121	33	66	0.30
3rd grade, first of year	99	32	66	111	37	62	0.34
3rd grade, mid year	117	32	65	131	37	62	0.38
3rd grade, end of year	135	31	63	147	36	61	0.37
Difference 1-1 to 3-3	111			117			

The final column in Table 3 gives Cohen's d, a standard measure of effect size, calculated by dividing the difference between the means by the common standard deviation. Traditionally, d values of .25 or larger are considered educationally significant. The effect sizes are small at the first administration of the ORF in Grade 1, which would be expected given the general equivalence of the two groups. After that time, and through the end of third grade, the d values become quite large, ranging up to .94 for special education students at the end of first grade. Only one of the 24 effect sizes from the comparisons after the beginning of first grade fails to meet the .25 criterion (.21 for non-special education students at the middle of second grade).

Table 4: Analysis of Variance, ORF by School and Special Education Status

Within Subjects Effects	F	df	prob.
ORF	336.11	8	<.001
ORF by School	5.56	8	<.001
ORF by Special Ed Status	0.34	4	0.85
ORF by School by Special Ed	4.16	8	<.001
Between Subjects Effects			
School	5.82	1	0.02
Special Education Status	2.01	1	0.16
School by Special Ed	0.99	1	0.32
Error df		98	

Table 5 reports the percentage of students in each school who, based on the ORF scores, were "at risk" of not meeting literacy goals and the percentage at "low risk" of not meeting these goals. Values are given for each administration beginning with the middle of first grade. The students in the DI school outperform those in the control school in all of the comparisons. While not all of the individual comparisons are statistically significant, the chance of having all of the differences favor the treatment school is extremely small.

Examination of the magnitude of the percentages is telling. With most assessment points within the treatment school, fewer than five percent of the students are deemed at risk of not reaching the goals and students in the control school are at least twice as likely as those in the treatment school to be at risk. Similarly, students in the treatment school are much more likely to be deemed at low risk of not meeting literacy goals. The differences are even more striking with the special education students. In fact, over half of the special education students in the DI school are at low risk of failure.

The final column of Table 5 reports the odds ratios, another type of effect size. The "odds" simply refer to the relative chance of a child falling in one group or another. For instance, for students in the control school at the mid-year of first grade (the first row of data in panel B), 56% were deemed at low risk of not meeting achievement goals, while 44% did not fall in this category. Thus, the odds that a child in the control school would be at low risk were 56/44 = 1.27, close to even. In the treatment school, 82% of the students at the mid-year point of first grade were deemed at low risk of not meeting achievement goals. Thus, the odds that a child in the treatment school would be at low risk were 82/18 = 4.52, much better than even. The odds ratio is calculated by simply dividing these two values (4.52/1.27 = 3.56, the value in the last column of the first row of Panel B). This tells us that the odds that a treatment school child will be at low risk is over 3.5 times as great as the odds that a control school child will be at low risk.

Table 5: Percentage of Students "At Risk" and "At Low Risk" of Not Meeting Literacy Goals by School and Administration

A. Percent of Students At Risk by School

	Control	DI	Chi-square	prob.	n	Odds ratio
1st grade, mid year	5	2	0.67	0.41	167	0.49
1st grade, end of year	6	1	2.72	0.10	167	0.19
2nd grade, first of year	14	2	7.54	0.01	168	0.15
2nd grade, mid year	17	7	3.53	0.06	167	0.39
2nd grade, end of year	8	4	1.65	0.20	167	0.41
3rd grade, first of year	12	5	2.66	0.10	160	0.38
3rd grade, mid year	5	3	0.67	0.41	159	0.49
3rd grade, end of year	6	3	1.35	0.25	156	0.39

B. Percent of Students at Low Risk by School

	Control	DI	Chi-square	prob.	n	Odds Ratio
1st grade, mid year	56	82	13.14	<.001	167	3.56
1st grade, end of year	62	86	12.02	0.001	167	3.63
2nd grade, first of year	54	76	8.75	0.003	168	2.67
2nd grade, mid year	70	83	3.88	0.05	167	2.09
2nd grade, end of year	63	78	4.66	0.03	167	2.11
3rd grade, first of year	68	76	1.28	0.26	160	1.49
3rd grade, mid year	69	81	3.18	0.08	159	1.93
3rd grade, end of year	69	76	0.80	0.37	156	1.38

C. Percent of Special Education Students at Low Risk by School

	Control	DI	Chi-square	prob.	n	Odds Ratio
1st grade, mid year	31	63	3.14	0.08	32	3.68
1st grade, end of year	31	69	4.50	0.03	32	4.82
2nd grade, first of year	31	56	2.03	0.15	32	2.83
2nd grade, mid year	44	56	0.50	0.48	32	1.65
2nd grade, end of year	31	59	2.53	0.11	33	3.15
3rd grade, first of year	27	65	4.63	0.03	32	5.03
3rd grade, mid year	40	59	1.13	0.29	32	2.14
3rd grade, end of year	40	53	0.54	0.46	32	1.68

Note: Benchmarks are not available for the first administration in first grade. Special education students are included in the values in panels A and B. For Panel A the odds ratio describes the relative odds of a student in the DI school versus a student in the control school being at high risk of failure. For panels B and C the odds ratio describe the relative odds of being at low risk of failure.

In general, the odds ratios are striking. For all administrations, the odds of a treatment school child being at risk of failure are less than half – often far less than half – that of a child in the control school. (These are the values in Panel A of Table 5.) Similar results appear at the other end of the scale. The students in the treatment school have much higher odds of being at a low risk of academic failure, and these results appear for both the total group of students (Panel B) and for special education students (Panel C).

Other Indicators

Table 6 gives the analysis of variance results for the other DIBELS measures (Panel A) and the associated descriptive statistics for both the total group (Panel B) and special education students (Panel C). These other measures were only given through the end of first grade and thus should be seen as providing only supplementary information to that provided by the longer-term data on the ORF.

The first two columns have results for letter naming fluency (LNF), which was assessed three times in kindergarten and at the beginning of first grade. It should be recalled that we did not expect that students in the treatment school would have greater gains in letter naming fluency (LNF) because that skill is not explicitly taught until a later point in the DI curriculum. The data support this expectation. The results of the analysis of variance (Panel A) indicate significant changes over time ($p<.001$) as well as significant main effects of school ($p=.04$) and special education status ($p<.001$). Comparisons of the means (Table 7) indicate that students in the control school had higher scores than students in the treatment school and special education students had lower scores than non special education students at all administrations of the test.

Onset Recognition Fluency was tested at only the first and middle of the kindergarten year. The analysis of variance results indicate a significant effect only for the over time changes ($p<.001$), with a trend for the main effect of school ($p =.09$). Examination of the means indicates higher values for the treatment school for both administrations. The increase in scores is slightly higher for the treatment school, but the interaction effect does not reach traditional levels of significance ($p= .12$).

Results for Phonemic Segmentation Fluency, which was assessed from the middle of kindergarten through the end of first grade, show a significant three-way interaction effect ($p = .005$), as well as significant two-way interactions between gains and school ($p<.001$) and gains and special education status ($p=.04$). Examination of the means indicates that, as expected, the gains were stronger in the treatment school than the control school and that this was especially so with the special education students. The students in the DI school had scores at the first testing that were substantially lower than those of students in the control school, but had virtually identical scores by the last administration.

Finally, results with Nonsense Word Fluency, which was assessed from mid kindergarten through first grade, indicate significant two-way interactions of gains by school ($p=.002$) and gains by special education status ($p<.001$), but not a significant three-way interaction ($p=.38$). Examination of the means indicates that, as expected,

Table 6: Analysis of Variance Results for Letter Naming Fluency, Onset Recognition Fluency, Phonemic Segmentation Fluency, and Nonsense Word Fluency with Schools and Special Education Status as Factors

A: Analysis of Variance Results

	Letter Naming Fluency		Onset Recognition Fluency		Phonemic Segmentation Fluency		Nonsense Word Fluency	
Within Subjects Effects	F	prob.	F	prob.	F	prob.	F	prob.
Change over time	160.92	<.001	97.32	<.001	42.44	<.001	166.72	<.001
Change by School	1.57	0.20	2.43	0.12	7.53	<.001	4.35	0.002
Change by Special Ed Status	2.27	0.08	0.38	0.54	2.56	0.04	5.40	<.001
Change by School by Sp. Ed.	0.30	0.82	2.13	0.15	3.81	0.005	1.05	0.38
Between Subjects Effects								
School	4.12	0.04	2.93	0.09	2.88	0.09	0.07	0.79
Special Education Status	14.94	<.001	1.03	0.31	17.20	<.001	13.75	<.001
School by Special Education	0.14	0.71	0.06	0.81	0.00	0.99	0.02	0.89

B: Mean Values of LNF, OnRF, PSF, and NWF by Time of Administration and School, Total Group

	LNF		OnRF		PSF		NWF	
Time of Administration	Control	DI	Control	DI	Control	DI	Control	DI
K, first of year	18	16	15	17	-----	-----	-----	-----
K, mid year	32	26	29	34	45	37	30	26
K, end of year	47	39	-----	-----	50	50	42	41
1st grade, first of year	41	37	-----	-----	48	44	40	37
1st grade, mid year	-----	-----	-----	-----	53	52	67	69
1st grade, end of year	-----	-----	-----	-----	56	55	86	95
Gains	23	21	14	16	11	18	56	69

C: Mean Values of LNF, OnRF, PSF, and NWF by Time of Administration and School, Special Education Students

	LNF		OnRF		PSF		NWF	
Time of Administration	Control	DI	Control	DI	Control	DI	Control	DI
K, first of year	14	9	13	14	-----	-----	-----	-----
K, mid year	26	16	25	35	40	22	24	17
K, end of year	37	29	-----	-----	43	48	34	32
1st grade, first of year	31	27	-----	-----	42	42	29	23
1st grade, mid year	-----	-----	-----	-----	47	47	48	55
1st grade, end of year	-----	-----	-----	-----	51	50	56	76
Gains	18	18	12	21	12	29	33	60

Note: Degrees of freedom for within subject effects are 3, 156 for LNF, 1, 162 for OnRF, and 4, 154 for PSF and NWF. Degrees of freedom for between subject effects are 1, 156 for LNF, 1, 162 for OnRF, and 1,154 for PSF and NWF. "Gains" reported in Panels B and C are simply the difference between the last and the first administration of a test.

gains were much stronger in the treatment school than in the control school. The average scores of students in the DI school were lower than those of students in the control school at the first administration, but were substantially higher by the last administration.

From: jstockard@nifdi.org
Sent: Wednesday, November 19, 2008 12:24 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number:
1791130758

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: This study is appropriate for the beginning reading section. Please
feel free to ask if you have any questions.
Thank you.

From: WhatWorks
Sent: Wednesday, November 19, 2008 2:26 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse (WWCPC-873)
Dear Dr. Stockard,

Thank you for your submission. We received your attached PDF of the study.

Please note that studies submitted to the WWC could be considered for review under the next round of reviews as determined by IES. We are unable to provide a timeline for the review process at this time.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Wednesday, November 19, 2008 12:24 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number: 1791130758

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: jstockard@nifdi.org

Message: This study is appropriate for the beginning reading section. Please feel free to ask if you have any questions.
Thank you.

From: jstockard@nifdi.org
Sent: Monday, June 22, 2009 3:40 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Check on the Status of a Submission, Reference ID
Number: 109747337

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: jstockard@nifdi.org

Message: Might you be able to tell me the status of submission # 29487267,
please?
It was a study of Direct Instruction mathematics.
Thank you.
Jean Stockard, Ph.D.

From: WhatWorks
Sent: Wednesday, June 24, 2009 1:41 PM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse (WWCPC-1322)
Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). We have received your email about the study submission for Direct Instruction mathematics. WWC staff are reviewing your email and will prepare a response.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Monday, June 22, 2009 3:40 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Check on the Status of a Submission,
Reference ID Number: 109747337

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: jstockard@nifdi.org

Message: Might you be able to tell me the status of submission # 29487267, please?
It was a study of Direct Instruction mathematics.
Thank you.
Jean Stockard, Ph.D.

From: WhatWorks
Sent: Tuesday, June 30, 2009 12:18 PM
To: 'jstockard@nifdi.org'
Subject: RE: What Works Clearinghouse (WWCPC-1322)
Dear Dr. Stockard,

Thank you for submitting additional studies via the bibliography you provided in your last email. We will pass the bibliography along to the appropriate review team. Please be assured that the WWC will contact you if the review team needs any further information.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Monday, June 29, 2009 6:39 PM
To: What Works
Cc: cristy@nifdi.org
Subject: Re: What Works Clearinghouse (WWCPC-1322)

Thank you for your note regarding studies of Direct Instruction mathematics programs. I have attached a bibliography of 16 relevant studies. With the exception of one doctoral dissertation, this list only includes works published in refereed journals. Please let me know if I can provide additional information. I hope this is a sufficient list for review.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction

> Dear Dr. Stockard,
>
>
>
> Thank you for contacting the What Works Clearinghouse (WWC). Your
> submission
> of "Improving Elementary Level Mathematics Achievement in a Large Urban
> District: The Effects of NIFDI-Supported Implementation of Direct
> Instruction
> in the Baltimore City Public School System" was received on October 6,
> 2008.
>
>
>
> Please be advised that the order in which interventions are reviewed by
> the
> WWC is based, in part, on the availability of studies examining the

> effectiveness of an intervention. Thus, one of the reasons this study
> examining the effects of Direct Instruction mathematics has not yet been
> reviewed is that there are other interventions with a greater number of
> studies examining their effectiveness. If there are additional studies
> about
> the effectiveness of Direct Instruction mathematics, please submit them to
> the WWC.
>
>
>
> It is important to note that the WWC website does not list all
> products/interventions that are the subject of research submitted to the
> WWC,
> but rather interventions for which we have completed full literature
> searches
> and reviewed all eligible studies against evidence standards.
>
>
>
> We hope that you will find this information helpful in understanding the
> study submission and WWC review process.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source
> of scientific evidence of what works in education. For more information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/> .
>
>
>
> _____
>
> -----Original Message-----
>
> From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
>
> Sent: Monday, June 22, 2009 3:40 PM
>
> To: info@whatworks.ed.gov
>
> Subject: IES Website: Contact Us: Check on the Status of a Submission,
>
> Reference ID Number: 109747337
>
>
>
> info@whatworks.ed.gov, this email was automatically sent through the
> Contact
>

> link on the WWC website.
>
>
>
> From: jstockard@nifdi.org
>
>
>
> Message: Might you be able to tell me the status of submission # 29487267,
>
> please?
>
> It was a study of Direct Instruction mathematics.
>
> Thank you.
>
> Jean Stockard, Ph.D.
>
>
>
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

From: Jean Stockard [jstockard@nifdi.org]
Sent: Thursday, March 18, 2010 10:16 PM
To: What Works
Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
Subject: Re: WWC Adolescent Literacy Review
Attachments: March 18 2010 - letter to WWC in response to query.doc; Fourth Graders
Reading CBM for WWC March 18 2010.doc

Dear Dr. James-Burdumy:

Thank you very much for examining the study of Reading Mastery with fourth
grade students. I am happy to respond to your queries and have tried to follow your numbering system. The response is in the two attached documents. Please feel free to contact me if you need any additional information.

As noted in the response, I'm currently traveling and won't be able to fax
the signed form until March 29, at the earliest. However, I have answered all the queries in the attachment. I will also usually have access to e-mail and can also be reached by phone (541-556-2694).

Finally, I assume that I and my colleagues at the National Institute for Direct Instruction will be given the opportunity to review any reports that you may develop regarding Reading Mastery or other Direct Instruction
programs to help ensure that you have found all of the relevant literature
and that the reviews are accurate. As you may know there have been numerous problems with this process in the past, and we would like to help
assure that it works better in the future.

Again, thank you for examining this study. And please ask if there is any additional information you may need.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> The What Works Clearinghouse (WWC), an initiative of the U. S. Department
> of
> Education's Institute of Education Sciences, was established to provide

> educators, policymakers, researchers, and the public with a central and
> trusted source of scientific evidence of what works in education.
>
> The purpose of this letter is to notify you that as part of our review
of
> studies of Adolescent Literacy interventions, we are reviewing the
> following
> study for possible inclusion in WWC intervention reports:
>
>   Stockard, J. (2010). Fourth Graders' Growth in Reading Fluency: A
> Pretest-Posttest Randomized Control Study comparing Reading Mastery and
> Scott
> Foresman Basal Reading Program. National Institute for Direct
Instruction.
>
> In this letter, we ask you to clarify some information about your study
> and
> sign an agreement that allows us to use the information you provide in
our
> review.
>
> Sincerely,
>
> Susanne James-Burdumy
> Principal Investigator, WWC Adolescent Literacy review
> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>

From: WhatWorks
Sent: Tuesday, March 23, 2010 5:12 PM
To: 'jstockard@nifdi.org'
Subject: RE: WWC Adolescent Literacy Review (WWCPC 2015)
Dear Dr. Stockard,

We appreciate your prompt response to our questions about your study. Please send the signed author agreement and any additional information to us by April 1, ten business days from our original request.

We have located the article that you cited from the Journal of Behavioral Assessment and Intervention for Children. However, we are unable to locate a copy of "Promoting Reading Achievement and Countering the 'Fourth-Grade Slump': The Impact of Direct Instruction on Reading Achievement in Fifth Grade" since it is forthcoming. You can submit a copy of the article in response to this email.

If you have additional studies evaluating the effectiveness of Reading Mastery, please submit them to the WWC so we may consider them for our review.

Please note, in all cases, to be reviewed by the WWC, a study must be published research or research that is deemed "final" by the study authors and publicly available. "In press" studies are acceptable provided they are publicly available.

The WWC Procedures and Standards Handbook provides more detailed information about the WWC review process, including the WWC's policy regarding the production and release of intervention reports (http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocId=5#production).

Developers receive an embargoed copy of the intervention report 24 hours prior to its release on the WWC website. Other members of the public are able to review the report as soon as it is released by the Institute of Education Sciences on the WWC website. If you would like to be notified of a report's release, please respond to this e-mail letting us know.

We hope this information is helpful. If you have additional questions, please feel free to respond to this email.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----

From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Thursday, March 18, 2010 10:16 PM
To: What Works
Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
Subject: Re: WWC Adolescent Literacy Review

Dear Dr. James-Burdumy:

Thank you very much for examining the study of Reading Mastery with fourth grade students. I am happy to respond to your queries and have tried to follow your numbering system. The response is in the two attached documents. Please feel free to contact me if you need any additional information.

As noted in the response, I'm currently traveling and won't be able to fax the signed form until March 29, at the earliest. However, I have answered all the queries in the attachment. I will also usually have access to e-mail and can also be reached by phone (541-556-2694).

Finally, I assume that I and my colleagues at the National Institute for Direct Instruction will be given the opportunity to review any reports that you may develop regarding Reading Mastery or other Direct Instruction programs to help ensure that you have found all of the relevant literature and that the reviews are accurate. As you may know there have been numerous problems with this process in the past, and we would like to help assure that it works better in the future.

Again, thank you for examining this study. And please ask if there is any additional information you may need.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
> The What Works Clearinghouse (WWC), an initiative of the U. S. Department
> of
> Education's Institute of Education Sciences, was established to provide
> educators, policymakers, researchers, and the public with a central and
> trusted source of scientific evidence of what works in education.
>
> The purpose of this letter is to notify you that as part of our review of
> studies of Adolescent Literacy interventions, we are reviewing the
> following
> study for possible inclusion in WWC intervention reports:
>
> Stockard, J. (2010). Fourth Graders' Growth in Reading Fluency: A
> Pretest-Posttest Randomized Control Study comparing Reading Mastery and
> Scott

> Foresman Basal Reading Program. National Institute for Direct Instruction.
>
> In this letter, we ask you to clarify some information about your study
> and
> sign an agreement that allows us to use the information you provide in our
> review.
>
> Sincerely,
>
> Susanne James-Burdumy
> Principal Investigator, WWC Adolescent Literacy review
> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>

From: Jean Stockard [jstockard@nifdi.org]
Sent: Thursday, March 25, 2010 11:20 AM
To: WhatWorks
Cc: kengel@nifdi.org
Subject: RE: WWC Adolescent Literacy Review (WWCPC 2015)

Attachments: Fifth grade reading article revision submitted August 27 2009.doc; Appendix.doc

Dear Dr. James-Burdumy:

Thank you for your response. I have enclosed a copy of the article that is forthcoming in JESPAR. As explained in my previous communication I will fax the signed author agreement upon getting back to the office next week.
Thank you for allowing the full 10 business days for that transaction.

As regarding other articles, might it be possible for you to send me a list of all material that you are considering so that I can check it against our files to see what might be missing? As you may know, there are dozens of studies (actually well over 100) of Reading Mastery. We have spent many months trying to compile them all and are still finding new ones of which we were not aware. Only by looking through the list might we be sure that you have found all of the material.

Thank you for your statement regarding the receipt of an embargoed report.
Given the very large number of studies of RM I worry that the 24 hour turnaround for a review of the material that you described in your message is not sufficient. The process also does not seem to allow for the opportunity to correct errors. As you no doubt know, the WWC has received extensive criticism for a perceived lack of transparency and outside review. A previous WWC review of Reading Mastery had a very large number of errors, which have yet to be corrected. Allowing sufficient time for review of the reports and for corrections seems to be an essential element of ensuring that the public has accurate information.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> We appreciate your prompt response to our questions about your study.
> Please send the signed author agreement and any additional information
> to us by April 1, ten business days from our original request.
>
>

>
> We have located the article that you cited from the Journal of
> Behavioral Assessment and Intervention for Children. However, we are
> unable to locate a copy of "Promoting Reading Achievement and
> Countering the 'Fourth-Grade Slump': The Impact of Direct Instruction
> on Reading Achievement in Fifth Grade" since it is forthcoming. You
> can submit a copy of the article in response to this email.
>
>
>
> If you have additional studies evaluating the effectiveness of Reading
> Mastery, please submit them to the WWC so we may consider them for our
> review.
>
>
>
> Please note, in all cases, to be reviewed by the WWC, a study must be
> published research or research that is deemed "final" by the study
> authors and publicly available. "In press" studies are acceptable
> provided they are publicly available.
>
>
>
> The WWC Procedures and Standards Handbook provides more detailed
> information about the WWC review process, including the WWC's policy
> regarding the production and release of intervention reports
> (http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&to
> cI
> d=5#production).
>
>
>
> Developers receive an embargoed copy of the intervention report 24
> hours prior to its release on the WWC website. Other members of the
> public are able to review the report as soon as it is released by the
> Institute of Education Sciences on the WWC website. If you would like
> to be notified of a report's release, please respond to this e-mail
letting us know.
>
>
>
> We hope this information is helpful. If you have additional questions,
> please feel free to respond to this email.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more

> information, please visit http://ies.ed.gov/ncee/wwc/
> <http://ies.ed.gov/ncee/wwc/> .
>
>
>
>
>
> -----Original Message-----
>
> From: Jean Stockard [mailto:jstockard@nifdi.org]
>
> Sent: Thursday, March 18, 2010 10:16 PM
>
> To: What Works
>
> Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
>
> Subject: Re: WWC Adolescent Literacy Review
>
>
>
> Dear Dr. James-Burdumy:
>
>
>
> Thank you very much for examining the study of Reading Mastery with
> fourth
>
> grade students. I am happy to respond to your queries and have tried
> to
>
> follow your numbering system. The response is in the two attached
>
> documents. Please feel free to contact me if you need any additional
>
> information.
>
>
>
> As noted in the response, I'm currently traveling and won't be able to
> fax
>
> the signed form until March 29, at the earliest. However, I have
> answered
>
> all the queries in the attachment. I will also usually have access to
>
> e-mail and can also be reached by phone (541-556-2694).
>
>
>
> Finally, I assume that I and my colleagues at the National Institute
> for
>

> Direct Instruction will be given the opportunity to review any reports
>
> that you may develop regarding Reading Mastery or other Direct
> Instruction
>
> programs to help ensure that you have found all of the relevant
> literature
>
> and that the reviews are accurate. As you may know there have been
>
> numerous problems with this process in the past, and we would like to
> help
>
> assure that it works better in the future.
>
>
>
> Again, thank you for examining this study. And please ask if there is
> any
>
> additional information you may need.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>
>
>
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>> The What Works Clearinghouse (WWC), an initiative of the U. S.
> Department
>
>> of
>
>> Education's Institute of Education Sciences, was established to

> provide
>
>> educators, policymakers, researchers, and the public with a central
> and
>
>> trusted source of scientific evidence of what works in education.
>
>>
>
>> The purpose of this letter is to notify you that as part of our
>> review
> of
>
>> studies of Adolescent Literacy interventions, we are reviewing the
>
>> following
>
>> study for possible inclusion in WWC intervention reports:
>
>>
>
>> Stockard, J. (2010). Fourth Graders' Growth in Reading
> Fluency: A
>
>> Pretest-Posttest Randomized Control Study comparing Reading Mastery
> and
>
>> Scott
>
>> Foresman Basal Reading Program. National Institute for Direct
> Instruction.
>
>>
>
>> In this letter, we ask you to clarify some information about your
> study
>
>> and
>
>> sign an agreement that allows us to use the information you provide
>> in
> our
>
>> review.
>
>>
>
>> Sincerely,
>
>>
>
>> Susanne James-Burdumy
>
>> Principal Investigator, WWC Adolescent Literacy review

>
>> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
>
>> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>
>>
>
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

From: What Works
Sent: Friday, March 26, 2010 5:28 PM
To: 'jstockard@nifdi.org'
Subject: WWC Adolescent Literacy Review (WWCPC 2015)
Attachments: Direct Instruction Reading Mastery_References.doc

Dear Dr. Stockard,

Thank you again for contacting the What Works Clearinghouse (WWC). Attached is a reference list of studies for the Adolescent Literacy review of Reading Mastery. If you have additional studies evaluating the effectiveness of Reading Mastery, please submit them to us no later than April 1.

Unfortunately, we will be unable to provide you with a courtesy copy of the report as it is WWC policy to only release the courtesy copy of the report to the developer (SRA/McGraw Hill).

We appreciate your contacting the WWC and hope that this email has addressed your questions.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Thursday, March 25, 2010 11:20 AM
To: WhatWorks
Cc: kengel@nifdi.org
Subject: RE: WWC Adolescent Literacy Review (WWCPC 2015)

Dear Dr. James-Burdumy:

Thank you for your response. I have enclosed a copy of the article that is forthcoming in JESPAR. As explained in my previous communication I will fax the signed author agreement upon getting back to the office next week. Thank you for allowing the full 10 business days for that transaction.

As regarding other articles, might it be possible for you to send me a list of all material that you are considering so that I can check it against our files to see what might be missing? As you may know, there are dozens of studies (actually well over 100) of Reading Mastery. We have spent many months trying to compile them all and are still finding new ones of which we were not aware. Only by looking through the list might we be sure that you have found all of the material.

Thank you for your statement regarding the receipt of an embargoed report. Given the very large number of studies of RM I worry that the 24 hour turnaround for a review of the material that you described in your message

is not sufficient. The process also does not seem to allow for the opportunity to correct errors. As you no doubt know, the WWC has received extensive criticism for a perceived lack of transparency and outside review. A previous WWC review of Reading Mastery had a very large number of errors, which have yet to be corrected. Allowing sufficient time for review of the reports and for corrections seems to be an essential element of ensuring that the public has accurate information.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> We appreciate your prompt response to our questions about your study.
> Please send the signed author agreement and any additional information
> to us by April 1, ten business days from our original request.
>
>
>
> We have located the article that you cited from the Journal of
> Behavioral Assessment and Intervention for Children. However, we are
> unable to locate a copy of "Promoting Reading Achievement and Countering
> the 'Fourth-Grade Slump': The Impact of Direct Instruction on Reading
> Achievement in Fifth Grade" since it is forthcoming. You can submit a
> copy of the article in response to this email.
>
>
>
> If you have additional studies evaluating the effectiveness of Reading
> Mastery, please submit them to the WWC so we may consider them for our
> review.
>
>
>
> Please note, in all cases, to be reviewed by the WWC, a study must be
> published research or research that is deemed "final" by the study
> authors and publicly available. "In press" studies are acceptable
> provided they are publicly available.
>
>
>
> The WWC Procedures and Standards Handbook provides more detailed
> information about the WWC review process, including the WWC's policy
> regarding the production and release of intervention reports
> (http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocI
> d=5#production).
>
>

>
> Developers receive an embargoed copy of the intervention report 24 hours
> prior to its release on the WWC website. Other members of the public are
> able to review the report as soon as it is released by the Institute of
> Education Sciences on the WWC website. If you would like to be notified
> of a report's release, please respond to this e-mail letting us know.
>
>
>
> We hope this information is helpful. If you have additional questions,
> please feel free to respond to this email.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more
> information, please visit http://ies.ed.gov/ncee/wwc/
> <http://ies.ed.gov/ncee/wwc/> .
>
>
>
>
>
> -----Original Message-----
>
> From: Jean Stockard [mailto:jstockard@nifdi.org]
>
> Sent: Thursday, March 18, 2010 10:16 PM
>
> To: What Works
>
> Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
>
> Subject: Re: WWC Adolescent Literacy Review
>
>
>
> Dear Dr. James-Burdumy:
>
>
>
> Thank you very much for examining the study of Reading Mastery with
> fourth
>
> grade students. I am happy to respond to your queries and have tried to
>
> follow your numbering system. The response is in the two attached
>
> documents. Please feel free to contact me if you need any additional
>

> information.
>
>
>
> As noted in the response, I'm currently traveling and won't be able to
> fax
>
> the signed form until March 29, at the earliest. However, I have
> answered
>
> all the queries in the attachment. I will also usually have access to
>
> e-mail and can also be reached by phone (541-556-2694).
>
>
>
> Finally, I assume that I and my colleagues at the National Institute for
>
> Direct Instruction will be given the opportunity to review any reports
>
> that you may develop regarding Reading Mastery or other Direct
> Instruction
>
> programs to help ensure that you have found all of the relevant
> literature
>
> and that the reviews are accurate. As you may know there have been
>
> numerous problems with this process in the past, and we would like to
> help
>
> assure that it works better in the future.
>
>
>
> Again, thank you for examining this study. And please ask if there is
> any
>
> additional information you may need.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>

>
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>> The What Works Clearinghouse (WWC), an initiative of the U. S.
> Department
>
>> of
>
>> Education's Institute of Education Sciences, was established to
> provide
>
>> educators, policymakers, researchers, and the public with a central
> and
>
>> trusted source of scientific evidence of what works in education.
>
>>
>
>> The purpose of this letter is to notify you that as part of our review
> of
>
>> studies of Adolescent Literacy interventions, we are reviewing the
>
>> following
>
>> study for possible inclusion in WWC intervention reports:
>
>>
>
>> Stockard, J. (2010). Fourth Graders' Growth in Reading
> Fluency: A
>
>> Pretest-Posttest Randomized Control Study comparing Reading Mastery
> and
>
>> Scott
>
>> Foresman Basal Reading Program. National Institute for Direct
> Instruction.
>
>>
>
>> In this letter, we ask you to clarify some information about your
> study
>
>> and
>
>> sign an agreement that allows us to use the information you provide in
> our

>
>> review.
>
>>
>
>> Sincerely,
>
>>
>
>> Susanne James-Burdumy
>
>> Principal Investigator, WWC Adolescent Literacy review
>
>> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
>
>> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>
>>
>
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

From: Jean Stockard <jstockard@nifdi.org>
Sent: Sunday, March 28, 2010 10:58 AM
To: What Works
Cc: kengel@nifdi.org; ajohnston@nifdi.org; cristy@nifdi.org; (b)(6)
Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)

Dear Dr. James-Burdumy:

Thank you very much for sending the list of references. You appear to have done a much better job of finding material than those who did the previous review of Reading Mastery. Our research team will try to look through this list within the next few days and get back to you if we have any additions.

I did want to correct your information about the "developer" of reading mastery. SRA/McGraw Hill is the publisher, but the developer is the Engelmann-Becker Corporation. People in this group have been working for decades on both basic and applied research related to how children learn and how teaching can best be done. And, of course, in developing the curricula they do extensive field testing before sending it to the publisher. Thus, correspondence from WWC about RM should also be sent to EB Corporation.

I also wanted to again raise my concerns about what seems to be a lack of opportunity for any outside review of the report or a process by which errors can be corrected. The one-day, early notice procedure that you described in your earlier note does not appear to allow for such corrections. As you no doubt know, numerous scholars have raised very serious concerns regarding the lack of outside review of WWC reports and the extent to which they contain erroneous conclusions. While I know that you are part of a larger organization, I hope that you can use this review to show how allowing for a transparent process and the chance for correction of any problems can lead to more accurate products.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>

>
> Thank you again for contacting the What Works Clearinghouse (WWC).
> Attached
> is a reference list of studies for the Adolescent Literacy review of
> Reading
> Mastery. If you have additional studies evaluating the effectiveness of
> Reading Mastery, please submit them to us no later than April 1.
>
>
>
> Unfortunately, we will be unable to provide you with a courtesy copy of
> the
> report as it is WWC policy to only release the courtesy copy of the report
> to
> the developer (SRA/McGraw Hill).
>
>
>
> We appreciate your contacting the WWC and hope that this email has
> addressed
> your questions.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source
> of scientific evidence of what works in education. For more information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/> .
>
>
>
> _____
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Thursday, March 25, 2010 11:20 AM
> To: WhatWorks
> Cc: kengel@nifdi.org
> Subject: RE: WWC Adolescent Literacy Review (WWCPC 2015)
>
>
>

> Dear Dr. James-Burdumy:
>
>
>
> Thank you for your response. I have enclosed a copy of the article that is
>
> forthcoming in JESPAR. As explained in my previous communication I will
>
> fax the signed author agreement upon getting back to the office next week.
>
> Thank you for allowing the full 10 business days for that transaction.
>
>
>
> As regarding other articles, might it be possible for you to send me a
>
> list of all material that you are considering so that I can check it
>
> against our files to see what might be missing? As you may know, there are
>
> dozens of studies (actually well over 100) of Reading Mastery. We have
>
> spent many months trying to compile them all and are still finding new
>
> ones of which we were not aware. Only by looking through the list might we
>
> be sure that you have found all of the material.
>
>
>
> Thank you for your statement regarding the receipt of an embargoed report.
>
> Given the very large number of studies of RM I worry that the 24 hour
>
> turnaround for a review of the material that you described in your message
>
> is not sufficient. The process also does not seem to allow for the
>
> opportunity to correct errors. As you no doubt know, the WWC has received
>
> extensive criticism for a perceived lack of transparency and outside
>
> review. A previous WWC review of Reading Mastery had a very large number
>
> of errors, which have yet to be corrected. Allowing sufficient time for
>
> review of the reports and for corrections seems to be an essential element
>

> of ensuring that the public has accurate information.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>>
>
>>
>
>> We appreciate your prompt response to our questions about your study.
>
>> Please send the signed author agreement and any additional information
>
>> to us by April 1, ten business days from our original request.
>
>>
>
>>
>
>>
>
>> We have located the article that you cited from the Journal of
>
>> Behavioral Assessment and Intervention for Children. However, we are
>
>> unable to locate a copy of "Promoting Reading Achievement and Countering
>

>> the 'Fourth-Grade Slump': The Impact of Direct Instruction on Reading
>
>> Achievement in Fifth Grade" since it is forthcoming. You can submit a
>
>> copy of the article in response to this email.
>
>>
>
>>
>
>>
>
>> If you have additional studies evaluating the effectiveness of Reading
>
>> Mastery, please submit them to the WWC so we may consider them for our
>
>> review.
>
>>
>
>>
>
>>
>
>> Please note, in all cases, to be reviewed by the WWC, a study must be
>
>> published research or research that is deemed "final" by the study
>
>> authors and publicly available. "In press" studies are acceptable
>
>> provided they are publicly available.
>
>>
>
>>
>
>>
>
>> The WWC Procedures and Standards Handbook provides more detailed
>
>> information about the WWC review process, including the WWC's policy
>
>> regarding the production and release of intervention reports
>
>> (http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocI
>
>> d=5#production).
>

>>
>
>>
>
>>
>
>> Developers receive an embargoed copy of the intervention report 24 hours
>
>> prior to its release on the WWC website. Other members of the public are
>
>> able to review the report as soon as it is released by the Institute of
>
>> Education Sciences on the WWC website. If you would like to be notified
>
>> of a report's release, please respond to this e-mail letting us know.
>
>>
>
>>
>
>>
>
>> We hope this information is helpful. If you have additional questions,
>
>> please feel free to respond to this email.
>
>>
>
>>
>
>>
>
>> What Works Clearinghouse
>
>>
>
>>
>
>>
>
>> The What Works Clearinghouse was established by the U.S. Department of
>
>> Education's Institute of Education Sciences to provide educators,
>
>> policymakers, researchers, and the public with a central and trusted
>
>> source of scientific evidence of what works in education. For more
>

>> information, please visit http://ies.ed.gov/ncee/wwc/
>
>> <http://ies.ed.gov/ncee/wwc/> .
>
>>
>
>>
>
>>
>
>>
>
>>
>
>> -----Original Message-----
>
>>
>
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>
>>
>
>> Sent: Thursday, March 18, 2010 10:16 PM
>
>>
>
>> To: What Works
>
>>
>
>> Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
>
>>
>
>> Subject: Re: WWC Adolescent Literacy Review
>
>>
>
>>
>
>>
>
>> Dear Dr. James-Burdumy:
>
>>
>
>>
>
>>

>>
>
>> Thank you very much for examining the study of Reading Mastery with
>
>> fourth
>
>>
>
>> grade students. I am happy to respond to your queries and have tried to
>
>>
>
>> follow your numbering system. The response is in the two attached
>
>>
>
>> documents. Please feel free to contact me if you need any additional
>
>>
>
>> information.
>
>>
>
>>
>
>>
>
>> As noted in the response, I'm currently traveling and won't be able to
>
>> fax
>
>>
>
>> the signed form until March 29, at the earliest. However, I have
>
>> answered
>
>>
>
>> all the queries in the attachment. I will also usually have access to
>
>>
>
>> e-mail and can also be reached by phone (541-556-2694).
>
>>
>

>>
>
>>
>
>> Finally, I assume that I and my colleagues at the National Institute for
>
>>
>
>> Direct Instruction will be given the opportunity to review any reports
>
>>
>
>> that you may develop regarding Reading Mastery or other Direct
>
>> Instruction
>
>>
>
>> programs to help ensure that you have found all of the relevant
>
>> literature
>
>>
>
>> and that the reviews are accurate. As you may know there have been
>
>>
>
>> numerous problems with this process in the past, and we would like to
>
>> help
>
>>
>
>> assure that it works better in the future.
>
>>
>
>>
>
>>
>
>> Again, thank you for examining this study. And please ask if there is
>
>> any
>
>>
>

>> additional information you may need.
>
>>
>
>>
>
>>
>
>> Sincerely,
>
>>
>
>>
>
>>
>
>> Jean Stockard, Ph.D.
>
>>
>
>> Director of Research
>
>>
>
>> National Institute for Direct Instruction
>
>>
>
>> Eugene, Oregon
>
>>
>
>> phone: 541-485-1973 or toll-free 877-485-1973
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>
>

>>> Dear Dr. Stockard,
>
>>
>
>>>
>
>>
>
>>> The What Works Clearinghouse (WWC), an initiative of the U. S.
>
>> Department
>
>>
>
>>> of
>
>>
>
>>> Education's Institute of Education Sciences, was established to
>
>> provide
>
>>
>
>>> educators, policymakers, researchers, and the public with a central
>
>> and
>
>>
>
>>> trusted source of scientific evidence of what works in education.
>
>>
>
>>>
>
>>
>
>>> The purpose of this letter is to notify you that as part of our review
>
>> of
>
>>
>
>>> studies of Adolescent Literacy interventions, we are reviewing the
>
>>
>

>>> following
>
>>
>
>>> study for possible inclusion in WWC intervention reports:
>
>>
>
>>>
>
>>
>
>>> Stockard, J. (2010). Fourth Graders' Growth in Reading
>
>> Fluency: A
>
>>
>
>>> Pretest-Posttest Randomized Control Study comparing Reading Mastery
>
>> and
>
>>
>
>>> Scott
>
>>
>
>>> Foresman Basal Reading Program. National Institute for Direct
>
>> Instruction.
>
>>
>
>>>
>
>>
>
>>> In this letter, we ask you to clarify some information about your
>
>> study
>
>>
>
>>> and
>
>>
>

>>> sign an agreement that allows us to use the information you provide in
>
>> our
>
>>
>
>>> review.
>
>>
>
>>>
>
>>
>
>>> Sincerely,
>
>>
>
>>>
>
>>
>
>>> Susanne James-Burdumy
>
>>
>
>>> Principal Investigator, WWC Adolescent Literacy review
>
>>
>
>>> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
>
>>
>
>>> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>
>>
>
>>>
>
>>
>
>>
>
>>
>
>>
>

>
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

From: kengel@nifdi.org
Sent: Monday, March 29, 2010 8:41 PM
To: What Works
Cc: Jean Stockard; (b)(6) cristy@nifdi.org; ajohnston@nifdi.org
Subject: Fwd: Re: WWC Adolescent Literacy Review (WWCPC 2015)

Dr. James-Burdumy:

Dr, Stockard has kept me apprised of the correspondence between the National Institute for Direct Instruction (NIFDI) and the WWC, including WWC's assignment of SRA/McGraw Hill as the developer of Direct Instruction
(DI). In a recent email, you wrote that you will be unable to provide NIFDI with a courtesy copy of the report "as it is WWC policy to only release the courtesy copy of the report to the developer (SRA/McGraw Hill)."

WWC would be the first organization in the 40+ years of DI to recognize SRA/McGraw Hill as the developer of DI. American Institutes for Research and other organziations have recognized my father, Siegfried "Zig" Engelmann, senior author of all DI programs and the creator of DI. See http://psych.athabascau.ca/html/387/OpenModules/Engelmann/Engelmannbio.shtml
for a brief biography of my father. We can send you citations of his works describing the theoretical and experimental development of DI. He has also produced the Developer's Guidelines for schools implementing DI.

SRA/McGraw Hill is just one of the publishers of DI materials, albeit the main publisher. There are DI programs authored by my father that are published separately. Funnix, published by Funnix.com, and Direct Instruction Spoken English (DISE), published by Sopris West, are two examples of DI programs authored by my father and published separately.

I would be interested to know the criteria used for determining a developer. SRA/McGraw Hill in no way created or developed the DI theory or method. They have only helped in being the publishing outlet for the developers of DI.

I would welcome the opportunity to discuss this issue with you via our toll-free number, 1 877 485-1973. NIFDI has been created to act on the developer's behalf. Zig Engelmann can also be reached via that number of via zig@nifdi.org.

Regards,

KE

--
Kurt E. Engelmann, Ph.D., President
The National Institute for Direct Instruction (NIFDI)
www.nifdi.org * 541-485-1973 * FAX: 541-683-7543
P.O. Box 11248 * Eugene, OR 97440 * kengel@nifdi.org

---------- Original Message ----------
-
Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)
From: "Jean Stockard" <jstockard@nifdi.org>
Date: Mon, March 29, 2010 4:09 pm
To: "What Works" <whatworks@mathematica-mpr.com>
Cc: danielgjohnston@gmail.com
kengel@nifdi.org
ajohnston@nifdi.org
cristy@nifdi.org

-

Dear Dr. James-Burdumy:

I have attached the signed author agreement form that was requested in earlier e-mails.

We are in the midst of looking for additional studies of Reading Mastery that might fall in the age range that you are looking at and were not in the list that you sent. As you have probably found through your work in this area, Reading Mastery is part of the larger corpus of Direct Instruction (DI) curricular materials. While Reading Mastery is generally used with younger children, another DI program, Corrective Reading (CR), is usually used with older children who are still struggling with reading.
CR is based on the same large body of basic research regarding how children learn and how teaching is best done that led to the development of Reading Mastery.

I hope that you will soon turn your attention to the large body of work on
Corrective Reading, for it is no doubt much more suited to the age range that you are working with. If you need assistance finding articles related
to that curriculum as well, please let me know.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> Thank you again for contacting the What Works Clearinghouse (WWC).
> Attached
> is a reference list of studies for the Adolescent Literacy review of

> Reading
> Mastery. If you have additional studies evaluating the effectiveness of
> Reading Mastery, please submit them to us no later than April 1.
>
>
>
> Unfortunately, we will be unable to provide you with a courtesy copy of
> the
> report as it is WWC policy to only release the courtesy copy of the
report
> to
> the developer (SRA/McGraw Hill).
>
>
>
> We appreciate your contacting the WWC and hope that this email has
> addressed
> your questions.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source
> of scientific evidence of what works in education. For more
information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/>
.
>
>
>
> _____
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Thursday, March 25, 2010 11:20 AM
> To: WhatWorks
> Cc: kengel@nifdi.org
> Subject: RE: WWC Adolescent Literacy Review (WWCPC 2015)
>
>
>
> Dear Dr. James-Burdumy:
>
>
>
> Thank you for your response. I have enclosed a copy of the article that
is
>

> forthcoming in JESPAR. As explained in my previous communication I will
>
> fax the signed author agreement upon getting back to the office next
week.
>
> Thank you for allowing the full 10 business days for that transaction.
>
>
>
> As regarding other articles, might it be possible for you to send me a
>
> list of all material that you are considering so that I can check it
>
> against our files to see what might be missing? As you may know, there
are
>
> dozens of studies (actually well over 100) of Reading Mastery. We have
>
> spent many months trying to compile them all and are still finding new
>
> ones of which we were not aware. Only by looking through the list might
we
>
> be sure that you have found all of the material.
>
>
>
> Thank you for your statement regarding the receipt of an embargoed
report.
>
> Given the very large number of studies of RM I worry that the 24 hour
>
> turnaround for a review of the material that you described in your
message
>
> is not sufficient. The process also does not seem to allow for the
>
> opportunity to correct errors. As you no doubt know, the WWC has
received
>
> extensive criticism for a perceived lack of transparency and outside
>
> review. A previous WWC review of Reading Mastery had a very large
number
>
> of errors, which have yet to be corrected. Allowing sufficient time for
>
> review of the reports and for corrections seems to be an essential
element
>
> of ensuring that the public has accurate information.
>
>
>

> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>>
>
>>
>
>> We appreciate your prompt response to our questions about your study.
>
>> Please send the signed author agreement and any additional information
>
>> to us by April 1, ten business days from our original request.
>
>>
>
>>
>
>>
>
>> We have located the article that you cited from the Journal of
>
>> Behavioral Assessment and Intervention for Children. However, we are
>
>> unable to locate a copy of "Promoting Reading Achievement and
Countering
>
>> the 'Fourth-Grade Slump': The Impact of Direct Instruction on Reading
>
>> Achievement in Fifth Grade" since it is forthcoming. You can submit a
>
>> copy of the article in response to this email.
>
>>
>
>>

>
>>
>
>> If you have additional studies evaluating the effectiveness of Reading
>
>> Mastery, please submit them to the WWC so we may consider them for our
>
>> review.
>
>>
>
>>
>
>>
>
>> Please note, in all cases, to be reviewed by the WWC, a study must be
>
>> published research or research that is deemed "final" by the study
>
>> authors and publicly available. "In press" studies are acceptable
>
>> provided they are publicly available.
>
>>
>
>>
>
>>
>
>> The WWC Procedures and Standards Handbook provides more detailed
>
>> information about the WWC review process, including the WWC's policy
>
>> regarding the production and release of intervention reports
>
>>
(http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocI
>
>> d=5#production).
>
>>
>
>>
>
>>
>
>> Developers receive an embargoed copy of the intervention report 24
hours
>
>> prior to its release on the WWC website. Other members of the public
are
>
>> able to review the report as soon as it is released by the Institute
of

>
>> Education Sciences on the WWC website. If you would like to be
notified
>
>> of a report's release, please respond to this e-mail letting us know.
>
>>
>
>>
>
>>
>
>> We hope this information is helpful. If you have additional questions,
>
>> please feel free to respond to this email.
>
>>
>
>>
>
>>
>
>> What Works Clearinghouse
>
>>
>
>>
>
>>
>
>> The What Works Clearinghouse was established by the U.S. Department of
>
>> Education's Institute of Education Sciences to provide educators,
>
>> policymakers, researchers, and the public with a central and trusted
>
>> source of scientific evidence of what works in education. For more
>
>> information, please visit http://ies.ed.gov/ncee/wwc/
>
>> <http://ies.ed.gov/ncee/wwc/> .
>
>>
>
>>
>
>>
>
>>
>
>>
>
>> -----Original Message-----

>
>>
>
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>
>>
>
>> Sent: Thursday, March 18, 2010 10:16 PM
>
>>
>
>> To: What Works
>
>>
>
>> Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
>
>>
>
>> Subject: Re: WWC Adolescent Literacy Review
>
>>
>
>>
>
>>
>
>> Dear Dr. James-Burdumy:
>
>>
>
>>
>
>>
>
>> Thank you very much for examining the study of Reading Mastery with
>
>> fourth
>
>>
>
>> grade students. I am happy to respond to your queries and have tried
to
>
>>
>
>> follow your numbering system. The response is in the two attached
>
>>
>
>> documents. Please feel free to contact me if you need any additional
>
>>
>

>> information.
>
>>
>
>>
>
>>
>
>> As noted in the response, I'm currently traveling and won't be able to
>
>> fax
>
>>
>
>> the signed form until March 29, at the earliest. However, I have
>
>> answered
>
>>
>
>> all the queries in the attachment. I will also usually have access to
>
>>
>
>> e-mail and can also be reached by phone (541-556-2694).
>
>>
>
>>
>
>>
>
>> Finally, I assume that I and my colleagues at the National Institute
for
>
>>
>
>> Direct Instruction will be given the opportunity to review any reports
>
>>
>
>> that you may develop regarding Reading Mastery or other Direct
>
>> Instruction
>
>>
>
>> programs to help ensure that you have found all of the relevant
>
>> literature
>
>>
>
>> and that the reviews are accurate. As you may know there have been

>
>>
>
>> numerous problems with this process in the past, and we would like to
>
>> help
>
>>
>
>> assure that it works better in the future.
>
>>
>
>>
>
>>
>
>> Again, thank you for examining this study. And please ask if there is
>
>> any
>
>>
>
>> additional information you may need.
>
>>
>
>>
>
>>
>
>> Sincerely,
>
>>
>
>>
>
>>
>
>> Jean Stockard, Ph.D.
>
>>
>
>> Director of Research
>
>>
>
>> National Institute for Direct Instruction
>
>>
>
>> Eugene, Oregon
>
>>

>
>> phone: 541-485-1973 or toll-free 877-485-1973
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> The What Works Clearinghouse (WWC), an initiative of the U. S.
>
>> Department
>
>>
>
>>> of
>
>>
>
>>> Education's Institute of Education Sciences, was established to
>
>> provide
>
>>
>
>>> educators, policymakers, researchers, and the public with a central
>
>> and
>
>>
>
>>> trusted source of scientific evidence of what works in education.
>
>>
>
>>>
>
>>

>
>>> The purpose of this letter is to notify you that as part of our
review
>
>> of
>
>>
>
>>> studies of Adolescent Literacy interventions, we are reviewing the
>
>>
>
>>> following
>
>>
>
>>> study for possible inclusion in WWC intervention reports:
>
>>
>
>>>
>
>>
>
>>>   Stockard, J. (2010). Fourth Graders' Growth in Reading
>
>> Fluency: A
>
>>
>
>>> Pretest-Posttest Randomized Control Study comparing Reading Mastery
>
>> and
>
>>
>
>>> Scott
>
>>
>
>>> Foresman Basal Reading Program. National Institute for Direct
>
>> Instruction.
>
>>
>
>>>
>
>>
>
>>> In this letter, we ask you to clarify some information about your
>
>> study
>

>>
>
>>> and
>
>>
>
>>> sign an agreement that allows us to use the information you provide
in
>
>> our
>
>>
>
>>> review.
>
>>
>
>>>
>
>>
>
>>> Sincerely,
>
>>
>
>>>
>
>>
>
>>> Susanne James-Burdumy
>
>>
>
>>> Principal Investigator, WWC Adolescent Literacy review
>
>>
>
>>> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
>
>>
>
>>> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>
>>
>
>>>
>
>>
>
>>
>
>>
>
>>

> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>

From: What Works
Sent: Tuesday, May 04, 2010 1:47 PM
To: 'jstockard@nifdi.org'; 'kengel@nifdi.org'
Subject: WWC Adolescent Literacy Review (WWCPC 2015)

Dr. Stockard and Dr. Engelmann,

Each WWC Intervention Report provides educators with the contact information for the organization where they can purchase that intervention. For simplicity, we have referred to this as the "developer's" contact information. In instances where there is a difference between the original developer of an intervention and the intervention's current distributer, WWC policy has been to include only the distributer's contact information. (When the original developer is documented in the research, the WWC will also identify that original developer, as was done in the English Language Learners review of Reading Mastery.)

The WWC provides an embargoed courtesy copy to the organization whose contact information is included in the report. This courtesy copy is provided one day in advance of the release of the report, and is intended to alert the developer (or, in this case, the distributer) that they may be contacted when the report is released.

The concern you raise has led us to revisit our policy regarding the developer contact. For future reviews, when we initially contact the current distributer of the product with the reference list to review for completeness, we will ask the current distributer if there is a developer outside of their organization that they would like us to also send a copy of the reference list and ultimately a courtesy copy of the report. If the distributer provides a developer contact, the WWC will send those materials to the specified contact. We will collect this information from the distributer to ensure that the appropriate organization has been identified.

To ensure that we are applying our new policy consistently in the case of Reading Mastery as we will to other interventions, we will still require SRA/McGraw Hill to approve sending a courtesy copy to the National Institute for Direct Instruction. We will send SRA/McGraw Hill an inquiry explaining the new policy and asking whether they consider the National Institute for Direct Instruction an organization that can receive a courtesy copy.

It is important to note that sending out the courtesy copy of the intervention is not intended to be a review stage. The WWC does not delay the release of a report if a developer has comments on the courtesy copy. Instead, distributers, developers, researchers, or anyone else that has concerns is encouraged to submit the concerns through the WWC Help Desk (http://ies.ed.gov/ncee/wwc/help/webmail/). The WWC Quality Review Team will carefully investigate any concern, and if the team determines an error was made, a revised intervention report will be released.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: Jean Stockard [mailto:jstockard@nifdi.org]
Sent: Sunday, March 28, 2010 10:58 AM
To: What Works
Cc: kengel@nifdi.org; ajohnston@nifdi.org; cristy@nifdi.org;
(b)(6)
Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)

Dear Dr. James-Burdumy:

Thank you very much for sending the list of references. You appear to have done a much better job of finding material than those who did the previous review of Reading Mastery. Our research team will try to look through this list within the next few days and get back to you if we have any additions.

I did want to correct your information about the "developer" of reading mastery. SRA/McGraw Hill is the publisher, but the developer is the Engelmann-Becker Corporation. People in this group have been working for decades on both basic and applied research related to how children learn and how teaching can best be done. And, of course, in developing the curricula they do extensive field testing before sending it to the publisher. Thus, correspondence from WWC about RM should also be sent to EB Corporation.

I also wanted to again raise my concerns about what seems to be a lack of opportunity for any outside review of the report or a process by which errors can be corrected. The one-day, early notice procedure that you described in your earlier note does not appear to allow for such corrections. As you no doubt know, numerous scholars have raised very serious concerns regarding the lack of outside review of WWC reports and the extent to which they contain erroneous conclusions. While I know that you are part of a larger organization, I hope that you can use this review to show how allowing for a transparent process and the chance for correction of any problems can lead to more accurate products.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

> Dear Dr. Stockard,
>
>
>
> Thank you again for contacting the What Works Clearinghouse (WWC).
> Attached

> is a reference list of studies for the Adolescent Literacy review of
> Reading
> Mastery. If you have additional studies evaluating the effectiveness of
> Reading Mastery, please submit them to us no later than April 1.
>
>
>
> Unfortunately, we will be unable to provide you with a courtesy copy of
> the
> report as it is WWC policy to only release the courtesy copy of the report
> to
> the developer (SRA/McGraw Hill).
>
>
>
> We appreciate your contacting the WWC and hope that this email has
> addressed
> your questions.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source
> of scientific evidence of what works in education. For more information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/> .
>
>
>
> _____
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Thursday, March 25, 2010 11:20 AM
> To: WhatWorks
> Cc: kengel@nifdi.org
> Subject: RE: WWC Adolescent Literacy Review (WWCPC 2015)
>
>
>
> Dear Dr. James-Burdumy:
>
>
>
> Thank you for your response. I have enclosed a copy of the article that is
>
> forthcoming in JESPAR. As explained in my previous communication I will

> fax the signed author agreement upon getting back to the office next week.
>
> Thank you for allowing the full 10 business days for that transaction.
>
>
>
> As regarding other articles, might it be possible for you to send me a
>
> list of all material that you are considering so that I can check it
>
> against our files to see what might be missing? As you may know, there are
>
> dozens of studies (actually well over 100) of Reading Mastery. We have
>
> spent many months trying to compile them all and are still finding new
>
> ones of which we were not aware. Only by looking through the list might we
>
> be sure that you have found all of the material.
>
>
>
> Thank you for your statement regarding the receipt of an embargoed report.
>
> Given the very large number of studies of RM I worry that the 24 hour
>
> turnaround for a review of the material that you described in your message
>
> is not sufficient. The process also does not seem to allow for the
>
> opportunity to correct errors. As you no doubt know, the WWC has received
>
> extensive criticism for a perceived lack of transparency and outside
>
> review. A previous WWC review of Reading Mastery had a very large number
>
> of errors, which have yet to be corrected. Allowing sufficient time for
>
> review of the reports and for corrections seems to be an essential element
>
> of ensuring that the public has accurate information.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research

> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>>
>
>>
>
>> We appreciate your prompt response to our questions about your study.
>
>> Please send the signed author agreement and any additional information
>
>> to us by April 1, ten business days from our original request.
>
>>
>
>>
>
>>
>
>> We have located the article that you cited from the Journal of
>
>> Behavioral Assessment and Intervention for Children. However, we are
>
>> unable to locate a copy of "Promoting Reading Achievement and Countering
>
>> the 'Fourth-Grade Slump': The Impact of Direct Instruction on Reading
>
>> Achievement in Fifth Grade" since it is forthcoming. You can submit a
>
>> copy of the article in response to this email.
>
>>
>
>>
>
>>
>
>> If you have additional studies evaluating the effectiveness of Reading
>
>> Mastery, please submit them to the WWC so we may consider them for our

>
>> review.
>
>>
>
>>
>
>>
>
>> Please note, in all cases, to be reviewed by the WWC, a study must be
>
>> published research or research that is deemed "final" by the study
>
>> authors and publicly available. "In press" studies are acceptable
>
>> provided they are publicly available.
>
>>
>
>>
>
>>
>
>>
>
>> The WWC Procedures and Standards Handbook provides more detailed
>
>> information about the WWC review process, including the WWC's policy
>
>> regarding the production and release of intervention reports
>
>> (http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocI
>
>> d=5#production).
>
>>
>
>>
>
>>
>
>>
>
>> Developers receive an embargoed copy of the intervention report 24 hours
>
>> prior to its release on the WWC website. Other members of the public are
>
>> able to review the report as soon as it is released by the Institute of
>
>> Education Sciences on the WWC website. If you would like to be notified
>
>> of a report's release, please respond to this e-mail letting us know.
>
>>
>
>>

>
>>
>
>> We hope this information is helpful. If you have additional questions,
>
>> please feel free to respond to this email.
>
>>
>
>>
>
>>
>
>> What Works Clearinghouse
>
>>
>
>>
>
>>
>
>> The What Works Clearinghouse was established by the U.S. Department of
>
>> Education's Institute of Education Sciences to provide educators,
>
>> policymakers, researchers, and the public with a central and trusted
>
>> source of scientific evidence of what works in education. For more
>
>> information, please visit http://ies.ed.gov/ncee/wwc/
>
>> <http://ies.ed.gov/ncee/wwc/> .
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>
>
>> -----Original Message-----
>
>>
>
>>
>
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>
>>
>
>>
>
>> Sent: Thursday, March 18, 2010 10:16 PM

>
>>
>
>> To: What Works
>
>>
>
>> Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
>
>>
>
>> Subject: Re: WWC Adolescent Literacy Review
>
>>
>
>>
>
>>
>
>> Dear Dr. James-Burdumy:
>
>>
>
>>
>
>>
>
>> Thank you very much for examining the study of Reading Mastery with
>
>> fourth
>
>>
>
>> grade students. I am happy to respond to your queries and have tried to
>
>>
>
>> follow your numbering system. The response is in the two attached
>
>>
>
>> documents. Please feel free to contact me if you need any additional
>
>>
>
>> information.
>
>>
>
>>
>
>>

>
>> As noted in the response, I'm currently traveling and won't be able to
>
>> fax
>
>>
>
>> the signed form until March 29, at the earliest. However, I have
>
>> answered
>
>>
>
>> all the queries in the attachment. I will also usually have access to
>
>>
>
>> e-mail and can also be reached by phone (541-556-2694).
>
>>
>
>>
>
>>
>
>> Finally, I assume that I and my colleagues at the National Institute for
>
>>
>
>> Direct Instruction will be given the opportunity to review any reports
>
>>
>
>> that you may develop regarding Reading Mastery or other Direct
>
>> Instruction
>
>>
>
>> programs to help ensure that you have found all of the relevant
>
>> literature
>
>>
>
>> and that the reviews are accurate. As you may know there have been
>
>>
>
>> numerous problems with this process in the past, and we would like to
>
>> help

>
>>
>
>> assure that it works better in the future.
>
>>
>
>>
>
>>
>
>> Again, thank you for examining this study. And please ask if there is
>
>> any
>
>>
>
>> additional information you may need.
>
>>
>
>>
>
>>
>
>> Sincerely,
>
>>
>
>>
>
>>
>
>> Jean Stockard, Ph.D.
>
>>
>
>> Director of Research
>
>>
>
>> National Institute for Direct Instruction
>
>>
>
>> Eugene, Oregon
>
>>
>
>> phone: 541-485-1973 or toll-free 877-485-1973
>
>>

>
>>
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>> Dear Dr. Stockard,
>
>>
>
>>>
>
>>
>
>>> The What Works Clearinghouse (WWC), an initiative of the U. S.
>
>> Department
>
>>
>
>>> of
>
>>
>
>>> Education's Institute of Education Sciences, was established to
>
>> provide
>
>>
>
>>> educators, policymakers, researchers, and the public with a central
>
>> and
>
>>
>
>>> trusted source of scientific evidence of what works in education.
>
>>
>
>>>
>
>>
>
>>> The purpose of this letter is to notify you that as part of our review

>
>> of
>
>>
>
>>> studies of Adolescent Literacy interventions, we are reviewing the
>
>>
>
>>> following
>
>>
>
>>> study for possible inclusion in WWC intervention reports:
>
>>
>
>>>
>
>>
>
>>>   Stockard, J. (2010). Fourth Graders' Growth in Reading
>
>> Fluency: A
>
>>
>
>>> Pretest-Posttest Randomized Control Study comparing Reading Mastery
>
>> and
>
>>
>
>>> Scott
>
>>
>
>>> Foresman Basal Reading Program. National Institute for Direct
>
>> Instruction.
>
>>
>
>>>
>
>>
>
>>> In this letter, we ask you to clarify some information about your
>
>> study
>
>>

>
>>> and
>
>>
>
>>> sign an agreement that allows us to use the information you provide in
>
>> our
>
>>
>
>>> review.
>
>>
>
>>>
>
>>
>
>>> Sincerely,
>
>>
>
>>>
>
>>
>
>>> Susanne James-Burdumy
>
>>
>
>>> Principal Investigator, WWC Adolescent Literacy review
>
>>
>
>>> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
>
>>
>
>>> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>
>>
>
>>>
>
>>
>
>>
>
>>
>
>>
>
>>

>
>
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

From: Jean Stockard [jstockard@nifdi.org]
Sent: Monday, May 17, 2010 11:30 AM
To: What Works
Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)

Kurt - In light of the material below it might be wise to check with McGraw Hill folks about what they do when the WWC sends them things about RM and other programs. I'm not sure how that is done now and it looks like
we will only get something if McGraw Hill says it is okay.

I can just feel my blood pressure rising as I think about this!!

Jean

> Dr. Stockard and Dr. Engelmann,
>
>
>
> Each WWC Intervention Report provides educators with the contact
> information
> for the organization where they can purchase that intervention. For
> simplicity, we have referred to this as the "developer's" contact
> information. In instances where there is a difference between the original
> developer of an intervention and the intervention's current distributer,
> WWC
> policy has been to include only the distributer's contact information.
> (When
> the original developer is documented in the research, the WWC will also
> identify that original developer, as was done in the English Language
> Learners review of Reading Mastery.)
>
>
>
> The WWC provides an embargoed courtesy copy to the organization whose
> contact
> information is included in the report. This courtesy copy is provided one
> day
> in advance of the release of the report, and is intended to alert the
> developer (or, in this case, the distributer) that they may be contacted
> when
> the report is released.
>
>
>
> The concern you raise has led us to revisit our policy regarding the
> developer contact. For future reviews, when we initially contact the
> current
> distributer of the product with the reference list to review for

> completeness, we will ask the current distributer if there is a
developer
> outside of their organization that they would like us to also send a
copy
> of
> the reference list and ultimately a courtesy copy of the report. If
the
> distributer provides a developer contact, the WWC will send those
> materials
> to the specified contact. We will collect this information from the
> distributer to ensure that the appropriate organization has been
> identified.
>
>
>
> To ensure that we are applying our new policy consistently in the case
of
> Reading Mastery as we will to other interventions, we will still
require
> SRA/McGraw Hill to approve sending a courtesy copy to the National
> Institute
> for Direct Instruction. We will send SRA/McGraw Hill an inquiry
explaining
> the new policy and asking whether they consider the National Institute
for
> Direct Instruction an organization that can receive a courtesy copy.
>
>
>
> It is important to note that sending out the courtesy copy of the
> intervention is not intended to be a review stage. The WWC does not
delay
> the release of a report if a developer has comments on the courtesy
copy.
> Instead, distributers, developers, researchers, or anyone else that has
> concerns is encouraged to submit the concerns through the WWC Help Desk
> (http://ies.ed.gov/ncee/wwc/help/webmail/
> <http://ies.ed.gov/ncee/wwc/help/webmail/>). The WWC Quality Review
Team
> will carefully investigate any concern, and if the team determines an
> error
> was made, a revised intervention report will be released.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source

> of scientific evidence of what works in education. For more
information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/>
.
>
>
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Sunday, March 28, 2010 10:58 AM
> To: What Works
> Cc: kengel@nifdi.org; ajohnston@nifdi.org; cristy@nifdi.org;
> [redacted]
> Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)
>
>
>
> Dear Dr. James-Burdumy:
>
>
>
> Thank you very much for sending the list of references. You appear to
have
>
> done a much better job of finding material than those who did the
previous
>
> review of Reading Mastery. Our research team will try to look through
this
>
> list within the next few days and get back to you if we have any
>
> additions.
>
>
>
> I did want to correct your information about the "developer" of reading
>
> mastery. SRA/McGraw Hill is the publisher, but the developer is the
>
> Engelmann-Becker Corporation. People in this group have been working
for
>
> decades on both basic and applied research related to how children
learn
>
> and how teaching can best be done. And, of course, in developing the
>
> curricula they do extensive field testing before sending it to the
>
> publisher. Thus, correspondence from WWC about RM should also be sent
to
>
> EB Corporation.

>
>
>
> I also wanted to again raise my concerns about what seems to be a lack
of
>
> opportunity for any outside review of the report or a process by which
>
> errors can be corrected. The one-day, early notice procedure that you
>
> described in your earlier note does not appear to allow for such
>
> corrections. As you no doubt know, numerous scholars have raised very
>
> serious concerns regarding the lack of outside review of WWC reports
and
>
> the extent to which they contain erroneous conclusions. While I know
that
>
> you are part of a larger organization, I hope that you can use this
review
>
> to show how allowing for a transparent process and the chance for
>
> correction of any problems can lead to more accurate products.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>
>
>
>
>
>
>> Dear Dr. Stockard,
>
>>
>
>>

>
>>
>
>> Thank you again for contacting the What Works Clearinghouse (WWC).
>
>> Attached
>
>> is a reference list of studies for the Adolescent Literacy review of
>
>> Reading
>
>> Mastery. If you have additional studies evaluating the effectiveness
of
>
>> Reading Mastery, please submit them to us no later than April 1.
>
>>
>
>>
>
>>
>
>> Unfortunately, we will be unable to provide you with a courtesy copy
of
>
>> the
>
>> report as it is WWC policy to only release the courtesy copy of the
>> report
>
>> to
>
>> the developer (SRA/McGraw Hill).
>
>>
>
>>
>
>>
>
>> We appreciate your contacting the WWC and hope that this email has
>
>> addressed
>
>> your questions.
>
>>
>
>>
>
>>
>
>> What Works Clearinghouse
>

>>
>
>>
>
>>
>
>> The What Works Clearinghouse was established by the U.S. Department of
>
>> Education's Institute of Education Sciences to provide educators,
>
>> policymakers, researchers, and the public with a central and trusted
>
>> source
>
>> of scientific evidence of what works in education. For more
information,
>
>> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/>
.
>
>>
>
>>
>
>>
>
>> _____
>
>>
>
>> -----Original Message-----
>
>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>
>> Sent: Thursday, March 25, 2010 11:20 AM
>
>> To: WhatWorks
>
>> Cc: kengel@nifdi.org
>
>> Subject: RE: WWC Adolescent Literacy Review (WWCPC 2015)
>
>>
>
>>
>
>>
>
>> Dear Dr. James-Burdumy:
>
>>
>
>>
>
>

>>
>
>> Thank you for your response. I have enclosed a copy of the article
that
>> is
>
>>
>
>> forthcoming in JESPAR. As explained in my previous communication I
will
>
>>
>
>> fax the signed author agreement upon getting back to the office next
>> week.
>
>>
>
>> Thank you for allowing the full 10 business days for that transaction.
>
>>
>
>>
>
>>
>
>> As regarding other articles, might it be possible for you to send me a
>
>>
>
>> list of all material that you are considering so that I can check it
>
>>
>
>> against our files to see what might be missing? As you may know, there
>> are
>
>>
>
>> dozens of studies (actually well over 100) of Reading Mastery. We have
>
>>
>
>> spent many months trying to compile them all and are still finding new
>
>>
>
>> ones of which we were not aware. Only by looking through the list
might
>> we
>
>>
>
>> be sure that you have found all of the material.

>
>>
>
>>
>
>>
>
>> Thank you for your statement regarding the receipt of an embargoed
>> report.
>
>>
>
>> Given the very large number of studies of RM I worry that the 24 hour
>
>>
>
>> turnaround for a review of the material that you described in your
>> message
>
>>
>
>> is not sufficient. The process also does not seem to allow for the
>
>>
>
>> opportunity to correct errors. As you no doubt know, the WWC has
>> received
>
>>
>
>> extensive criticism for a perceived lack of transparency and outside
>
>>
>
>> review. A previous WWC review of Reading Mastery had a very large
number
>
>>
>
>> of errors, which have yet to be corrected. Allowing sufficient time
for
>
>>
>
>> review of the reports and for corrections seems to be an essential
>> element
>
>>
>
>> of ensuring that the public has accurate information.
>
>>
>
>>

>
>>
>
>> Sincerely,
>
>>
>
>>
>
>>
>
>> Jean Stockard, Ph.D.
>
>>
>
>> Director of Research
>
>>
>
>> National Institute for Direct Instruction
>
>>
>
>> Eugene, Oregon
>
>>
>
>> phone: 541-485-1973 or toll-free 877-485-1973
>
>>
>
>>
>
>>
>
>>
>
>>
>
>>> Dear Dr. Stockard,
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>

>
>>> We appreciate your prompt response to our questions about your study.
>
>>
>
>>> Please send the signed author agreement and any additional
information
>
>>
>
>>> to us by April 1, ten business days from our original request.
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> We have located the article that you cited from the Journal of
>
>>
>
>>> Behavioral Assessment and Intervention for Children. However, we are
>
>>
>
>>> unable to locate a copy of "Promoting Reading Achievement and
>>> Countering
>
>>
>
>>> the 'Fourth-Grade Slump': The Impact of Direct Instruction on Reading
>
>>
>
>>> Achievement in Fifth Grade" since it is forthcoming. You can submit a
>
>>
>
>>> copy of the article in response to this email.
>
>>
>
>>>
>
>>

>
>>>
>
>>
>
>>>
>
>>
>
>>> If you have additional studies evaluating the effectiveness of
Reading
>
>>
>
>>> Mastery, please submit them to the WWC so we may consider them for
our
>
>>
>
>>> review.
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> Please note, in all cases, to be reviewed by the WWC, a study must be
>
>>
>
>>> published research or research that is deemed "final" by the study
>
>>
>
>>> authors and publicly available. "In press" studies are acceptable
>
>>
>
>>> provided they are publicly available.
>
>>
>
>>>
>
>>

>
>>>
>
>>
>
>>>
>
>>
>
>>> The WWC Procedures and Standards Handbook provides more detailed
>
>>
>
>>> information about the WWC review process, including the WWC's policy
>
>>
>
>>> regarding the production and release of intervention reports
>
>>
>
>>>
(http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocI
>
>>
>
>>> d=5#production).
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> Developers receive an embargoed copy of the intervention report 24
>>> hours
>
>>
>
>>> prior to its release on the WWC website. Other members of the public
>>> are
>
>>
>
>>> able to review the report as soon as it is released by the Institute
of

>
>>
>
>>> Education Sciences on the WWC website. If you would like to be
notified
>
>>
>
>>> of a report's release, please respond to this e-mail letting us know.
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> We hope this information is helpful. If you have additional
questions,
>
>>
>
>>> please feel free to respond to this email.
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> What Works Clearinghouse
>
>>
>
>>>
>
>>
>
>>>

>
>>
>
>>>
>
>>
>
>>> The What Works Clearinghouse was established by the U.S. Department
of
>
>>
>
>>> Education's Institute of Education Sciences to provide educators,
>
>>
>
>>> policymakers, researchers, and the public with a central and trusted
>
>>
>
>>> source of scientific evidence of what works in education. For more
>
>>
>
>>> information, please visit http://ies.ed.gov/ncee/wwc/
>
>>
>
>>> <http://ies.ed.gov/ncee/wwc/> .
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> -----Original Message-----
>

>>
>
>>>
>
>>
>
>>> From: Jean Stockard [mailto:jstockard@nifdi.org]
>
>>
>
>>>
>
>>
>
>>> Sent: Thursday, March 18, 2010 10:16 PM
>
>>
>
>>>
>
>>
>
>>> To: What Works
>
>>
>
>>>
>
>>
>
>>> Cc: ajohnston@nifdi.org kengel@nifdi.org cristy@nifdi.org
>
>>
>
>>>
>
>>
>
>>> Subject: Re: WWC Adolescent Literacy Review
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>

>>> Dear Dr. James-Burdumy:
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> Thank you very much for examining the study of Reading Mastery with
>
>>
>
>>> fourth
>
>>
>
>>>
>
>>
>
>>> grade students. I am happy to respond to your queries and have tried
to
>
>>
>
>>>
>
>>
>
>>> follow your numbering system. The response is in the two attached
>
>>
>
>>>
>
>>
>
>>> documents. Please feel free to contact me if you need any additional
>
>>
>
>>>
>
>>
>
>>> information.

>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> As noted in the response, I'm currently traveling and won't be able
to
>
>>
>
>>> fax
>
>>
>
>>>
>
>>
>
>>> the signed form until March 29, at the earliest. However, I have
>
>>
>
>>> answered
>
>>
>
>>>
>
>>
>
>>> all the queries in the attachment. I will also usually have access to
>
>>
>
>>>
>
>>
>
>>> e-mail and can also be reached by phone (541-556-2694).
>
>>
>
>>>
>

>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> Finally, I assume that I and my colleagues at the National Institute
>>> for
>
>>
>
>>>
>
>>
>
>>> Direct Instruction will be given the opportunity to review any
reports
>
>>
>
>>>
>
>>
>
>>> that you may develop regarding Reading Mastery or other Direct
>
>>
>
>>> Instruction
>
>>
>
>>>
>
>>
>
>>> programs to help ensure that you have found all of the relevant
>
>>
>
>>> literature
>
>>
>
>>>
>
>>
>
>>> and that the reviews are accurate. As you may know there have been
>

>>
>
>>>
>
>>
>
>>> numerous problems with this process in the past, and we would like to
>
>>
>
>>> help
>
>>
>
>>>
>
>>
>
>>> assure that it works better in the future.
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> Again, thank you for examining this study. And please ask if there is
>
>>
>
>>> any
>
>>
>
>>>
>
>>
>
>>> additional information you may need.
>
>>
>
>>>
>
>>
>

>>>
>
>>
>
>>>
>
>>
>
>>> Sincerely,
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> Jean Stockard, Ph.D.
>
>>
>
>>>
>
>>
>
>>> Director of Research
>
>>
>
>>>
>
>>
>
>>> National Institute for Direct Instruction
>
>>
>
>>>
>
>>
>
>>> Eugene, Oregon
>
>>
>
>>>
>

>>
>
>>> phone: 541-485-1973 or toll-free 877-485-1973
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>
>>
>
>>>> The What Works Clearinghouse (WWC), an initiative of the U. S.
>
>>
>

>>> Department
>
>>
>
>>>
>
>>
>
>>>> of
>
>>
>
>>>
>
>>
>
>>>> Education's Institute of Education Sciences, was established to
>
>>
>
>>> provide
>
>>
>
>>>
>
>>
>
>>>> educators, policymakers, researchers, and the public with a central
>
>>
>
>>> and
>
>>
>
>>>
>
>>
>
>>>> trusted source of scientific evidence of what works in education.
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>

>>
>
>>>> The purpose of this letter is to notify you that as part of our
review
>
>>
>
>>> of
>
>>
>
>>>
>
>>
>
>>>> studies of Adolescent Literacy interventions, we are reviewing the
>
>>
>
>>>
>
>>
>
>>>> following
>
>>
>
>>>
>
>>
>
>>>> study for possible inclusion in WWC intervention reports:
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>
>>
>
>>>>   Stockard, J. (2010). Fourth Graders' Growth in Reading
>
>>
>
>>> Fluency: A
>
>>

>
>>>
>
>>
>
>>>> Pretest-Posttest Randomized Control Study comparing Reading Mastery
>
>>
>
>>> and
>
>>
>
>>>
>
>>
>
>>>> Scott
>
>>
>
>>>
>
>>
>
>>>> Foresman Basal Reading Program. National Institute for Direct
>
>>
>
>>> Instruction.
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>
>>
>
>>>> In this letter, we ask you to clarify some information about your
>
>>
>
>>> study
>
>>
>
>>>

>
>>
>
>>>> and
>
>>
>
>>>
>
>>
>
>>>> sign an agreement that allows us to use the information you provide
in
>
>>
>
>>> our
>
>>
>
>>>
>
>>
>
>>>> review.
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>
>>
>
>>>> Sincerely,
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>

>>
>
>>>> Susanne James-Burdumy
>
>>
>
>>>
>
>>
>
>>>> Principal Investigator, WWC Adolescent Literacy review
>
>>
>
>>>
>
>>
>
>>>> <<Author Request Table (AL017886).doc>> <<AL017886 Author Contact
>
>>
>
>>>
>
>>
>
>>>> Attachment.pdf>> <<AL017886 Author Contact Letter.pdf>>
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>

>>
>
>>
>
>>
>
>>
>
>> Jean Stockard, Ph.D.
>
>>
>
>> Director of Research
>
>>
>
>> National Institute for Direct Instruction
>
>>
>
>> Eugene, Oregon
>
>>
>
>> phone: 541-485-1973 or toll-free 877-485-1973
>
>>
>
>>
>
>
>
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

From: kengel@nifdi.org
Sent: Friday, May 21, 2010 1:37 PM
To: What Works
Cc: jstockard@nifdi.org; kengel@nifdi.org
Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)

Dear What Works:

We received your email on the clarification of the WWC's policy on courtesy copies of reports and the distinction between the developer and the distributor. In your email, you mention that "when the original developer is documented in the research, the WWC will also identify that original developer, as was done in the English Language Learners review of
Reading Mastery." It's true that Siegfried Engelmann is identified as the
developer in the English Language Learners review of Reading Mastery, but no contact information is given. Moreover, he is not identified as the developer in the Beginning Reading Intervention Report on Reading Mastery:

http://ies.ed.gov/ncee/wwc/pdf/WWC_ReadingMastery_081208.pdf

If the publisher is deemed by the WWC to be the organization where educators can purchase an intervention, then WWC has defined the intervention very narrowly as print or other physical products that can be
purchased. That contradicts the definition of the intervention in the research articles reviewed by WWC, which include training and other forms of support for teachers and administrators.

WWC doesn't apply the criterion of interventions being only physical products when defining other interventions. WWC analyses practices, such as Dialogic Reading, different reward systems and Applied Behavior Analysis. Is the DI method in effect being penalized because the developer articulated a part of the method into a print form, and now WWC is only recognizing the print form as the intervention?

Take the case of Enhanced Proactive Reading. The WWC Intervention Report http://ies.ed.gov/ncee/wwc/pdf/WWC_Proactive_Reading092806.pdf lists the developer as the Vaughn Gross Center for Reading and Language Arts and gives contact information:

College of Education SZB 228, University of Texas at Austin, 1 University Station D4900, Austin, TX 78712-0365. Email: info@texasreading.
org. Web: www.texasreading.org/utcrla/. Telephone: (512) 232-
2320.

Enhanced Proactive Reading involves curricular materials that are not available commercially and training provided through the Center. What if sometime in the future, the authors of the curricular materials reach an agreement to a publisher to distribute the materials? Will the WWC then de-list the Center and put in the publisher as the contact? That would then parallel the case of Reading Mastery.

It is a positive step that WWC will ask publishers whether there is a developer who should also be mentioned, but not providing the contact information of the developer really channels educators into thinking about
the intervention as merely a print product--a way of thinking that the developer of Reading Mastery has been fighting against for years. Purchasing the print program without proper support and understanding of the method behind the materials only leads to a low-level reproduction of the intervention, and not a reproduction similar to the model used during research.

I am copying Dr. Jean Stockard, NIFDI's Director of Research and Evaluation, for her information.

KE

--
Kurt E. Engelmann, Ph.D., President
The National Institute for Direct Instruction (NIFDI)
www.nifdi.org * 541-485-1973 * FAX: 541-683-7543
P.O. Box 11248 * Eugene, OR 97440 * kengel@nifdi.org

> Dr. Stockard and Dr. Engelmann,
>
>
>
> Each WWC Intervention Report provides educators with the contact
> information
> for the organization where they can purchase that intervention. For
> simplicity, we have referred to this as the "developer's" contact
> information. In instances where there is a difference between the original
> developer of an intervention and the intervention's current distributer,
> WWC
> policy has been to include only the distributer's contact information.
> (When
> the original developer is documented in the research, the WWC will also
> identify that original developer, as was done in the English Language
> Learners review of Reading Mastery.)
>
>
>
> The WWC provides an embargoed courtesy copy to the organization whose
> contact
> information is included in the report. This courtesy copy is provided one
> day
> in advance of the release of the report, and is intended to alert the
> developer (or, in this case, the distributer) that they may be contacted
> when

> the report is released.
>
>
>
> The concern you raise has led us to revisit our policy regarding the
> developer contact. For future reviews, when we initially contact the
> current
> distributer of the product with the reference list to review for
> completeness, we will ask the current distributer if there is a
developer
> outside of their organization that they would like us to also send a
copy
> of
> the reference list and ultimately a courtesy copy of the report. If
the
> distributer provides a developer contact, the WWC will send those
> materials
> to the specified contact. We will collect this information from the
> distributer to ensure that the appropriate organization has been
> identified.
>
>
>
> To ensure that we are applying our new policy consistently in the case
of
> Reading Mastery as we will to other interventions, we will still
require
> SRA/McGraw Hill to approve sending a courtesy copy to the National
> Institute
> for Direct Instruction. We will send SRA/McGraw Hill an inquiry
explaining
> the new policy and asking whether they consider the National Institute
for
> Direct Instruction an organization that can receive a courtesy copy.
>
>
>
> It is important to note that sending out the courtesy copy of the
> intervention is not intended to be a review stage. The WWC does not
delay
> the release of a report if a developer has comments on the courtesy
copy.
> Instead, distributers, developers, researchers, or anyone else that has
> concerns is encouraged to submit the concerns through the WWC Help Desk
> (http://ies.ed.gov/ncee/wwc/help/webmail/
> <http://ies.ed.gov/ncee/wwc/help/webmail/>). The WWC Quality Review
Team
> will carefully investigate any concern, and if the team determines an
> error
> was made, a revised intervention report will be released.
>
>
>
> What Works Clearinghouse

>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source
> of scientific evidence of what works in education. For more
information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/>
.
>
>
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Sunday, March 28, 2010 10:58 AM
> To: What Works
> Cc: kengel@nifdi.org; ajohnston@nifdi.org; cristy@nifdi.org;
> danielgjohnston@gmail.com
> Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)
>
>
>
> Dear Dr. James-Burdumy:
>
>
>
> Thank you very much for sending the list of references. You appear to
have
>
> done a much better job of finding material than those who did the
previous
>
> review of Reading Mastery. Our research team will try to look through
this
>
> list within the next few days and get back to you if we have any
>
> additions.
>
>
>
> I did want to correct your information about the "developer" of reading
>
> mastery. SRA/McGraw Hill is the publisher, but the developer is the
>
> Engelmann-Becker Corporation. People in this group have been working
for
>
> decades on both basic and applied research related to how children
learn
>
> and how teaching can best be done. And, of course, in developing the

> curricula they do extensive field testing before sending it to the publisher. Thus, correspondence from WWC about RM should also be sent to EB Corporation.

> I also wanted to again raise my concerns about what seems to be a lack of opportunity for any outside review of the report or a process by which errors can be corrected. The one-day, early notice procedure that you described in your earlier note does not appear to allow for such corrections. As you no doubt know, numerous scholars have raised very serious concerns regarding the lack of outside review of WWC reports and the extent to which they contain erroneous conclusions. While I know that you are part of a larger organization, I hope that you can use this review to show how allowing for a transparent process and the chance for correction of any problems can lead to more accurate products.

> Sincerely,

> Jean Stockard, Ph.D.
> Director of Research
> National Institute for Direct Instruction
> Eugene, Oregon
> phone: 541-485-1973 or toll-free 877-485-1973

From: WhatWorks
Sent: Monday, May 24, 2010 12:11 PM
To: 'kengel@nifdi.org'
Cc: 'jstockard@nifdi.org'; 'kengel@nifdi.org'
Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)
Dear Dr. Engelmann,

We received your email and will take your comments into consideration as we refine and implement the new policy.

Thank you,

What Works Clearinghouse

-----Original Message-----
From: kengel@nifdi.org [mailto:kengel@nifdi.org]
Sent: Friday, May 21, 2010 1:37 PM
To: What Works
Cc: jstockard@nifdi.org; kengel@nifdi.org
Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)

Dear What Works:

We received your email on the clarification of the WWC's policy on courtesy copies of reports and the distinction between the developer and the distributor. In your email, you mention that "when the original developer is documented in the research, the WWC will also identify that original developer, as was done in the English Language Learners review of Reading Mastery." It's true that Siegfried Engelmann is identified as the developer in the English Language Learners review of Reading Mastery, but no contact information is given. Moreover, he is not identified as the developer in the Beginning Reading Intervention Report on Reading Mastery:

http://ies.ed.gov/ncee/wwc/pdf/WWC_ReadingMastery_081208.pdf

If the publisher is deemed by the WWC to be the organization where educators can purchase an intervention, then WWC has defined the intervention very narrowly as print or other physical products that can be purchased. That contradicts the definition of the intervention in the research articles reviewed by WWC, which include training and other forms of support for teachers and administrators.

WWC doesn't apply the criterion of interventions being only physical products when defining other interventions. WWC analyses practices, such as Dialogic Reading, different reward systems and Applied Behavior Analysis. Is the DI method in effect being penalized because the developer articulated a part of the method into a print form, and now WWC is only recognizing the print form as the intervention?

Take the case of Enhanced Proactive Reading. The WWC Intervention Report http://ies.ed.gov/ncee/wwc/pdf/WWC_Proactive_Reading092806.pdf lists the developer as the Vaughn Gross Center for Reading and Language Arts and gives contact information:

College of Education SZB 228, University of Texas at Austin, 1 University Station D4900, Austin, TX 78712-0365. Email: info@texasreading.org. Web: www.texasreading.org/utcrla/. Telephone: (512) 232-2320.

Enhanced Proactive Reading involves curricular materials that are not available commercially and training provided through the Center. What if sometime in the future, the authors of the curricular materials reach an agreement to a publisher to distribute the materials? Will the WWC then de-list the Center and put in the publisher as the contact? That would then parallel the case of Reading Mastery.

It is a positive step that WWC will ask publishers whether there is a developer who should also be mentioned, but not providing the contact information of the developer really channels educators into thinking about the intervention as merely a print product--a way of thinking that the developer of Reading Mastery has been fighting against for years. Purchasing the print program without proper support and understanding of the method behind the materials only leads to a low-level reproduction of the intervention, and not a reproduction similar to the model used during research.

I am copying Dr. Jean Stockard, NIFDI's Director of Research and Evaluation, for her information.

KE

--
Kurt E. Engelmann, Ph.D., President
The National Institute for Direct Instruction (NIFDI)
www.nifdi.org * 541-485-1973 * FAX: 541-683-7543
P.O. Box 11248 * Eugene, OR 97440 * kengel@nifdi.org

> Dr. Stockard and Dr. Engelmann,
>
>
>
> Each WWC Intervention Report provides educators with the contact
> information
> for the organization where they can purchase that intervention. For
> simplicity, we have referred to this as the "developer's" contact
> information. In instances where there is a difference between the original
> developer of an intervention and the intervention's current distributer,
> WWC
> policy has been to include only the distributer's contact information.
> (When
> the original developer is documented in the research, the WWC will also
> identify that original developer, as was done in the English Language
> Learners review of Reading Mastery.)
>
>
>
> The WWC provides an embargoed courtesy copy to the organization whose
> contact
> information is included in the report. This courtesy copy is provided one
> day
> in advance of the release of the report, and is intended to alert the
> developer (or, in this case, the distributer) that they may be contacted
> when
> the report is released.
>
>

>
> The concern you raise has led us to revisit our policy regarding the
> developer contact. For future reviews, when we initially contact the
> current
> distributer of the product with the reference list to review for
> completeness, we will ask the current distributer if there is a developer
> outside of their organization that they would like us to also send a copy
> of
> the reference list and ultimately a courtesy copy of the report. If the
> distributer provides a developer contact, the WWC will send those
> materials
> to the specified contact. We will collect this information from the
> distributer to ensure that the appropriate organization has been
> identified.
>
>
>
> To ensure that we are applying our new policy consistently in the case of
> Reading Mastery as we will to other interventions, we will still require
> SRA/McGraw Hill to approve sending a courtesy copy to the National
> Institute
> for Direct Instruction. We will send SRA/McGraw Hill an inquiry explaining
> the new policy and asking whether they consider the National Institute for
> Direct Instruction an organization that can receive a courtesy copy.
>
>
>
> It is important to note that sending out the courtesy copy of the
> intervention is not intended to be a review stage. The WWC does not delay
> the release of a report if a developer has comments on the courtesy copy.
> Instead, distributers, developers, researchers, or anyone else that has
> concerns is encouraged to submit the concerns through the WWC Help Desk
> (http://ies.ed.gov/ncee/wwc/help/webmail/
> <http://ies.ed.gov/ncee/wwc/help/webmail/>). The WWC Quality Review Team
> will carefully investigate any concern, and if the team determines an
> error
> was made, a revised intervention report will be released.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source
> of scientific evidence of what works in education. For more information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/> .
>
>
>
> -----Original Message-----
> From: Jean Stockard [mailto:jstockard@nifdi.org]
> Sent: Sunday, March 28, 2010 10:58 AM

> To: What Works
> Cc: kengel@nifdi.org; ajohnston@nifdi.org; cristy@nifdi.org;
> (b)(6)
> Subject: Re: WWC Adolescent Literacy Review (WWCPC 2015)
>
>
>
> Dear Dr. James-Burdumy:
>
>
>
> Thank you very much for sending the list of references. You appear to have
>
> done a much better job of finding material than those who did the previous
>
> review of Reading Mastery. Our research team will try to look through this
>
> list within the next few days and get back to you if we have any
>
> additions.
>
>
>
> I did want to correct your information about the "developer" of reading
>
> mastery. SRA/McGraw Hill is the publisher, but the developer is the
>
> Engelmann-Becker Corporation. People in this group have been working for
>
> decades on both basic and applied research related to how children learn
>
> and how teaching can best be done. And, of course, in developing the
>
> curricula they do extensive field testing before sending it to the
>
> publisher. Thus, correspondence from WWC about RM should also be sent to
>
> EB Corporation.
>
>
>
> I also wanted to again raise my concerns about what seems to be a lack of
>
> opportunity for any outside review of the report or a process by which
>
> errors can be corrected. The one-day, early notice procedure that you
>
> described in your earlier note does not appear to allow for such
>
> corrections. As you no doubt know, numerous scholars have raised very
>
> serious concerns regarding the lack of outside review of WWC reports and
>
> the extent to which they contain erroneous conclusions. While I know that
>
> you are part of a larger organization, I hope that you can use this review

>
> to show how allowing for a transparent process and the chance for
>
> correction of any problems can lead to more accurate products.
>
>
>
> Sincerely,
>
>
>
> Jean Stockard, Ph.D.
>
> Director of Research
>
> National Institute for Direct Instruction
>
> Eugene, Oregon
>
> phone: 541-485-1973 or toll-free 877-485-1973
>
>

The Impact of Reading Mastery on Statewide Reading Assessment Scores in a Midwestern State:

A Cohort Control Group with Historical Comparisons Design

December 28, 2010

Jean Stockard, Ph.D.
Director of Research,
National Institute for Direct Instruction
Eugene, Oregon
jstockard@nifdi.org

The Impact of Reading Mastery on Statewide Reading Assessment Scores in a Midwestern State: A Cohort Control Group with Historical Comparisons Design

Purpose:
This study examined the relationship of exposure to Reading Mastery in early elementary school to scores on a statewide reading assessment test administered in fourth grade.

Sample and Design:
This study examined differences in scores on a state assessment of reading skills of students in a rural, Midwestern district who experienced different levels of exposure to the highly structured and explicit reading curriculum, Reading Mastery (RM). The study used a cohort control group (also called a recurrent institutional cycle) design, comparing cohorts of students with varying levels of exposure to RM. The research design literature suggests that this is a valid and useful alternative to randomized control trials in organizational settings:

> Many institutions experience regular turnover as one group "graduates" to another level and their place is taken by another group. Schools are an obvious example of this, as most children are promoted from one grade to the next each year….The term cohort designates the successive groups that go through processes such as these. Cohorts are particularly useful as control groups if (1) one cohort experiences a given treatment and earlier or later cohorts do not; (2) cohorts differ in only minor ways from their contiguous cohorts; (3) organizations insist that a treatment be given to everybody, thus precluding simultaneous controls and making possible only historical controls; and (4) an organization's archival records can be used for constructing and then comparing cohorts. (Shadish, Cook, and Campbell, 2002, pp. 148-149, emphasis in original,; see also Cook & Campbell, 1979, pp. 126-127 and Campbell & Stanley, 1963, pp. 56-61).

The data used in this analysis fit these criteria. All members of each cohort received the assigned treatment at the grades indicated in Table 1, and there were no apparent differences between the cohorts. The cohort that was in fourth grade in 2004-05 had no exposure to RM, while the cohort that was in fourth grade in 2008-09 had been exposed to the curriculum since kindergarten. To provide additional controls we include historical data for the state as a whole.

Measures and Analysis:
The percentage of fourth graders in the district who met or exceeded the established state standards, as determined by the state's assessment, were examined and compared to the percentage in the state as a whole. We computed an effect size describing the magnitude of the difference between the district performance and that of the state and tested the hypothesis that changes over time in the district were greater than in the state as a whole.

Results:
Table 1 reports the percentage of fourth graders who met the established state reading standards in the district and in the state as a whole from 2004-05 to 2008-09. In the earliest year none of the fourth graders had any exposure to RM, while in the last year the fourth graders had been exposed to the curriculum since beginning kindergarten. Thus the data allow us to compare performance on the statewide test for fourth grade cohorts before any exposure to the curriculum,

with cohorts who had partial exposure during their primary grades, and with a cohort with full exposure (having RM since kindergarten).

Results indicate that students in both the district and the state improved their performance over time, but the increase was greater for students in the targeted district than in the state as a whole. Before implementation of the curriculum the performance of the district's fourth graders was lower than that of students in the state as a whole (z = -.18), but by 2006-07, their performance was better than that of students state-wide (z = .14) and this higher level of performance remained for all subsequent cohorts in the analysis. We calculated a t-test examining the null hypothesis that the changes in the district over time, relative to the changes in the state, were equal to zero. The results indicate that these changes were statistically significant (t = 2.52, df = 259, p = .006 [one-tailed]), and the associated effect size of .31 would be considered educationally important. In other words, the change from 2004-05 to 2008-09 in the percentage of fourth grade students who passed the state exam was significantly greater than the change in this percentage in the state as a whole.

Conclusion

The results of this study replicate numerous other studies that have found that students receiving instruction in Reading Mastery have significantly higher reading achievement than students who study other curricula.

Table 5

Percentage of Fourth Graders Meeting State Reading Standards, by Exposure to Curriculum, District and State

Year	%-Dist	%-State	N	Z – score	Grades Exposed to RM	Type of Exposure
2004-05	79	85	117	-0.18	0	None
2005-06	87	88	147	-0.05	gr 3-4	Partial
2006-07	94	89	156	0.14	gr 2-4	Partial
2007-08	99	91	144	0.26	gr 1-4	Partial
2008-09	98	95	144	0.14	gr K-4	Full

References

Campbell, D.T. & Stanley, J.C. (1963). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

Cook, T.D. & Campbell, D.T. (1979). Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Shadish, W.R.. Cook, T.D. & Campbell, D.T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin.

From: WhatWorks
Sent: Monday, January 03, 2011 3:19 PM
To: 'jstockard@nifdi.org'; 'research@nifdi.org'
Subject: What Works Clearinghouse (WWC 2685)
Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). We received your email messages and one attached document, "The Impact of Reading Mastery on Statewide Reading Assessment Scores in a Midwestern State: A Cohort Control Group with Historical Comparisons Design." We apologize for any problems you experienced uploading documents to our website. Please respond to this email with the additional documents for your submission attached.

For additional information on the submission process, you may refer to the following link to the FAQ section on our website: http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=15&tocId=6#submitting5.

If you have any questions, please feel free to reply to this email.

Thank you,

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: research@nifdi.org [mailto:research@nifdi.org]
Sent: Friday, December 31, 2010 3:49 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number: 1334610581

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: research@nifdi.org

Message: I just tried to submit files and the site apparently did not work. (No studies would appear in the box. Please let me know when it is in working order again.
Thank you.
the reference number was 1894126256.
Please also send e-mail to jstockard@nifdi.org.
thank you.
Jean Stockard
Director of Research,
National Institute for Direct Instruction

-----Original Message-----
From: research@nifdi.org [mailto:research@nifdi.org]
Sent: Friday, December 31, 2010 3:42 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number: 1894126256

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: research@nifdi.org

Message: DATE: December 31, 2010

TO: What Works Clearinghouse Staff

FROM: Jean Stockard, PhD
Director of Research
National Institute for Direct Instruction

I am submitting 5 documents for inclusion in use in your reviews of the Direct Instruction curricula. Three are short research reports involving one site each. One is a randomized control study of fourth grade students reading comprehension. The others use cohort control group designs, deemed as most appropriate by the Campbell and Stanley literature for studies in intact organizations such as schools. The other two documents involve analyses of 18 different sites and over 60 individual comparisons. The extensive article explains the analysis procedures that were used and the appendix provides additional explanation for the analyses within each site.

Please send any questions to me at jstockard@nifdi.org and also to research@nifdi.org so that others on our research team may respond as needed. We look forward to seeing the corrections needed in your earlier reports on the various Direct Instruction programs.

From: Jean Stockard <jstockard@nifdi.org>
Sent: Wednesday, January 05, 2011 7:17 PM
To: WhatWorks
Cc: ajohnston@nifdi.org
Subject: Re: What Works Clearinghouse (WWC 2685)
Attachments: Rural Midwestern - RM in K on later reading achievement.pdf; Fourth Graders Reading Comprehension for WWC 12.31.10.pdf; Merging accountability and scientific research article 11.05.10 with appendices.pdf; Appendix B 12.30.10.pdf

Thank you very much for your reply. I have attached the four remaining
papers to this message.
Jean

> Dear Dr. Stockard,
>
> Thank you for contacting the What Works Clearinghouse (WWC). We received
> your email messages and one attached document, "The Impact of Reading
> Mastery on Statewide Reading Assessment Scores in a Midwestern State: A
> Cohort Control Group with Historical Comparisons Design." We apologize for
> any problems you experienced uploading documents to our website. Please
> respond to this email with the additional documents for your submission
> attached.
>
> For additional information on the submission process, you may refer to the
> following link to the FAQ section on our website:
> http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=15&tocId=6#submitting5.
>
> If you have any questions, please feel free to reply to this email.
>
> Thank you,
>
> What Works Clearinghouse
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source of scientific evidence of what works in education. For more
> information, please visit http://ies.ed.gov/ncee/wwc/.
>
> -----Original Message-----
> From: research@nifdi.org [mailto:research@nifdi.org]
> Sent: Friday, December 31, 2010 3:49 PM
> To: info@whatworks.ed.gov
> Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID
> Number: 1334610581
>
> info@whatworks.ed.gov, this email was automatically sent through the

> Contact link on the WWC website.
>
> From: research@nifdi.org
>
> Message: I just tried to submit files and the site apparently did not
> work. (No studies would appear in the box. Please let me know when it is
> in working order again.
> Thank you.
> the reference number was 1894126256.
> Please also send e-mail to jstockard@nifdi.org.
> thank you.
> Jean Stockard
> Director of Research,
> National Institute for Direct Instruction
>
>
>
>
>
> -----Original Message-----
>
> From: research@nifdi.org [mailto:research@nifdi.org]
>
> Sent: Friday, December 31, 2010 3:42 PM
>
> To: info@whatworks.ed.gov
>
> Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID
>
> Number: 1894126256
>
>
>
> info@whatworks.ed.gov, this email was automatically sent through the
> Contact
>
> link on the WWC website.
>
>
>
> From: research@nifdi.org
>
>
>
> Message: DATE: December 31, 2010
>
>
>

> TO: What Works Clearinghouse Staff
>
>
>
> FROM: Jean Stockard, PhD
>
> Director of Research
>
> National Institute for Direct Instruction
>
>
>
> I am submitting 5 documents for inclusion in use in your reviews of the
> Direct Instruction curricula. Three are short research reports involving
> one site each. One is a randomized control study of fourth grade students
> reading
>
> comprehension. The others use cohort control group designs, deemed as most
> appropriate by the Campbell and Stanley literature for studies in intact
> organizations such as schools. The other two documents involve analyses of
> 18 different sites and over 60 individual comparisons. The extensive
> article explains the analysis procedures that were used and the appendix
> provides additional explanation for the analyses within each site.
>
>
>
> Please send any questions to me at jstockard@nifdi.org and also to
> research@nifdi.org so that others on our research team may respond as
> needed. We look forward to seeing the corrections needed in your earlier
> reports on
>
> the various Direct Instruction programs.
>
>
>
>
>
>
>
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

From: research@nifdi.org
Sent: Friday, December 31, 2010 3:42 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number:
1894126256

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: research@nifdi.org

Message: DATE: December 31, 2010

TO: What Works Clearinghouse Staff

FROM: Jean Stockard, PhD
Director of Research
National Institute for Direct Instruction

I am submitting 5 documents for inclusion in use in your reviews of the Direct
Instruction curricula. Three are short research reports involving one site
each. One is a randomized control study of fourth grade students reading comprehension. The others use cohort control group designs, deemed as most
appropriate by the Campbell and Stanley literature for studies in intact organizations such as schools. The other two documents involve analyses of 18
different sites and over 60 individual comparisons. The extensive article explains the analysis procedures that were used and the appendix provides additional explanation for the analyses within each site.

Please send any questions to me at jstockard@nifdi.org and also to research@nifdi.org so that others on our research team may respond as needed.
We look forward to seeing the corrections needed in your earlier reports on
the various Direct Instruction programs.

From: research@nifdi.org
Sent: Friday, December 31, 2010 3:49 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Submit a Study to Review, Reference ID Number:
1334610581

info@whatworks.ed.gov, this email was automatically sent through the Contact
link on the WWC website.

From: research@nifdi.org

Message: I just tried to submit files and the site apparently did not work.
(No studies would appear in the box.
Please let me know when it is in working order again.
Thank you.
the reference number was 1894126256.
Please also send e-mail to jstockard@nifdi.org.
thank you.
Jean Stockard
Director of Research,
National Institute for Direct Instruction

Merging the Accountability and Scientific Research Requirements of the No Child Left Behind Act: Using Cohort Control Groups

Jean Stockard
Professor Emerita, University of Oregon
and
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

November 5, 2010

Please send correspondence to Jean Stockard, Department of Planning Public Policy and Management, University of Oregon, Eugene, Oregon 97403; 541-346-5005; jeans@uoregon.edu, jstockard@nifdi.org

The author thanks Douglas Carnine, Kurt Engelmann, Zig Engelmann, Dan Johnston, and Robert O'Brien for helpful comments on earlier drafts. Any errors are the sole responsibility of the author.

Merging the Accountability and Scientific Research Requirements of the No Child Left Behind Act: Using Cohort Control Groups

Abstract

This paper shows how assessment data such as that mandated by NCLB can be used to examine the effectiveness of educational interventions and meet the Act's mandate for "scientifically based research." Based on the classic research design literature a cohort control group (CCG) and a cohort control group with historical comparisons (CCG-H) design are suggested as internally valid analyses. The logic of the "grounded theory of generalized causal inference" is used to develop externally valid results. The procedure is illustrated with published data regarding the Reading Mastery curriculum. Empirical results are comparable to those obtained in meta-analyses of the curriculum, with effect sizes surpassing the usual criterion for educational importance. Implications for school officials and policy makers are discussed.

Merging the Accountability and Scientific Research Requirements of the No Child Left Behind Act: Using Cohort Control Groups

Few would dispute that the No Child Left Behind (NCLB) Act has had an extraordinary influence on education in the United States. School administrators, teachers, and students have probably been most affected by the Act's focus on accountability and using standardized assessments to identify schools needing improvement. A key element of the Act was the requirement that states measure progress in reading and mathematics and compare students' scores to established benchmarks. Educational researchers have probably been more affected by the Act's requirement for "scientifically based research" and the call for "evidence-based practice."

To a large extent, the work resulting from these two aspects of the Act appears to have developed in parallel, and generally non-overlapping, avenues. The most powerful voices in the educational research community have defined "scientifically based research" in ways that emphasize small, randomized control trials within schools or artificial settings. Specialists in assessment have designed elaborate instruments to measure basic skills, gather data from schools throughout the country, and compare results to established benchmarks defining adequate progress. Yet, the two strands of work are largely unconnected.

This paper suggests that the data schools obtain as part of their routine state and federally mandated assessment programs can be used to examine the effectiveness of educational curricula. More importantly, based on the logic developed in the classic research design literature, we suggest that appropriate analysis of these data can approximate the quality of results that could be obtained through randomized control trials of the same curricula. The analysis approach presented is simple, but, we assert, logically valid and "scientifically-based." It

could easily be used by local school officials wanting to assess the impact of curricular changes. Thus, the approach directly addresses an issue of utmost concern to school authorities – examining the extent to which changes at their school produce desired results.

The first section of this paper reviews the classic literature regarding how research designs assess causal relationships (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). The second section builds upon this discussion by showing how assessment data can be used to study the effectiveness of educational interventions in ways that provide high internal validity. The third section expands upon this analysis by reporting the results of several dozen analyses from eighteen different sites, following the Campbell, et al. tradition of using multiple studies to develop generalizations about causal relationships and thus promote external validity. The final section summarizes the work and discusses potential implications for practitioners and researchers.

All of our examples involve applications of the Reading Mastery curriculum. This program is especially appropriate for this discussion because its publisher routinely disseminates reports of assessment scores of schools using the materials. These scores provided the basis of our analysis. The logic of our approach could, of course, be used to evaluate other types of interventions and curricula in institutional settings.

Research Designs to Assess Effective Programs

Lawrence Sherman (2003) suggested that the "political use" of "evidence-led policy" began in the U.K. in the 1990s. The term first appeared in medicine, apparently in reaction to studies finding that the vast majority of medical treatments in everyday use had never been scientifically tested (Millenson 1997, p. 4, cited by Sherman, 2003, p. 7). The concern for such evidence quickly spread to other policy arenas. In education, the response has focused on calls

for a greater reliance on experimental designs, particularly those involving randomized assignment (Cook 2002, 2003; Sherman, 2003; Towne, Wise, & Winters, 2005). In contrast, in medicine the response appears to have also included calls for systematic reviews of accumulated evidence (e.g., Noble, 2006). The sections below discuss both of these foci utilizing the classical discussions regarding research design from the social science literature.

Experimental Designs to Establish Causality

In recent years, building upon the NCLB's mandate for "scientifically based research," the most influential proponent of randomized control trials in education has no doubt been the Institute of Education Sciences (IES), the "research arm" of the U.S. Department of Education. It has defined randomized controlled trials as the "gold standard" for educational studies, stating that "randomized trials are the only sure method for determining the effectiveness of education program and practices" (Whitehurst, 2003, p. 6). The use of these procedures has a strong influence on the ranking of curricula in literature reviews and in determining research funding (Julnes & Rog 2007; NMAP, 2008; Schwandt, 2005; St. Pierre, 2006; U.S. Department of Education, 2002; U. S. Department of Education, 2006; What Works Clearinghouse, 2008).

The logical argument for randomized control trials is familiar to any undergraduate social science major who has studied experimental design. The literature builds on John Stuart Mill's delineation of three important criteria for inferring causation: co-variation between a cause and an effect, temporal precedence of the cause to the effect, and ruling out other possible causal variables. Expanding upon that logic, Campbell and Stanley (1963) and their successors (Cook & Campbell, 1979, and Shadish, Cook, & Campbell, 2002, which are the later, expanded editions of the 1963 monograph) provided extensive discussions of the variety of designs that could be used in testing causal relationships. These writings have long been considered the standard

reference for research design in the social sciences. In the pages that follow we will refer to the work as the Campbell, Cook, Shadish, Stanley (CCSS) approach.

The notion of validity is central to writings of CCSS. This term refers to the accuracy of inferences or conclusions that can be made from the results of a study. The most recent developments distinguish four types of validity:

1. Statistical conclusion validity, regarding the association (correlation) between receiving a treatment, such as exposure to a curriculum, and an outcome, such as achievement;
2. Internal validity, regarding whether the observed association between the treatment and outcome represents a causal relationship;
3. Construct validity, regarding the extent to which inferences from the study can be related to the theory on which it was based; and
4. External validity, the extent to which the results can be extended to other settings and conditions.

The arguments for using randomized control trials in educational research have focused on internal validity, because these designs logically deal with each of the classically defined "threats to internal validity." Because subjects are randomly assigned to either an experimental or control condition, they theoretically experience similar maturational processes, historical events, reactions to testing, etc. Thus, any differences accruing to the experimental group that are different than those accruing to the control group may be attributed to the experimental intervention. Like the general CCSS tradition, the argument stresses the vital importance of study design in producing results that should be used as the basis of social policy.

While accepting the logical argument of this approach, numerous authors have suggested that restricting analyses of educational interventions to randomized control trials may not be

appropriate and could result in misleading conclusions. Using the language of experimental design, they suggest that randomized trials, when implemented in real-life settings, may have "many more potential threats to internal validity than would highly controlled quasi-experiments" (McMillan, 2007, p. 1).[1] The concerns involve numerous issues ranging from how subjects are assigned to treatment when students must be grouped in classes or schools, to how fidelity of treatment can be ensured in real-life settings where both the implementers (e.g. teachers) and the subjects (e.g. students) may routinely interact.

The classic "gold standard" experiment in a field such as medicine is "double blind," with neither the interventionists (e.g. doctors) nor the subjects (e.g. patients) knowing which group they are in. Yet, such a situation is generally impossible to attain in a field setting such as a school or even a school district with multiple schools. Teachers certainly know the situation and students no doubt also soon understand what is happening. Generations of research have documented the "Hawthorne" and "reverse Hawthorne" or "John Henry" effects, regarding how knowledge of experimental designs influences subject behavior (e.g. Zdep & Irvine, 1970). Most importantly, such knowledge and the potential of its effect immediately adds another variable to the design. There is, logically, no way for an analyst to determine if differences between the control and experimental group are due to the intervention or to other types of actions by the subjects based on their knowledge of the situation (Cook, Scriven, Coryn, & Evergreen, 2009; McMillan, 2007; Raudenbush, 2008; Scriven, 2008, Scriven, n.d.; Slavin, 2008; see also Cartwright, 2007). In summarizing these concerns, Michael Scriven concluded that "quasi-experimental designs…are alternative ways to establish conclusions, often better ways in particular circumstances" (Cook, et al, 2009, pp. 4-5).[2]

A close reading of the classical CCSS writings on experimental designs, as well as more recent commentaries by Cook and Shadish, indicate that the essence of this conclusion also appears in that material, especially in the discussion of "recurrent institutional cycle" or "cohort control group" designs. In this discussion they propose a useful alternative to randomized control trials in organizational settings:

> Many institutions experience regular turnover as one group "graduates" to another level and their place is taken by another group. Schools are an obvious example of this, as most children are promoted from one grade to the next each year….The term cohort designates the successive groups that go through processes such as these. Cohorts are particularly useful as control groups if (1) one cohort experiences a given treatment and earlier or later cohorts do not; (2) cohorts differ in only minor ways from their contiguous cohorts; (3) organizations insist that a treatment be given to everybody, thus precluding simultaneous controls and making possible only historical controls; and (4) an organization's archival records can be used for constructing and then comparing cohorts. (Shadish, Cook, and Campbell, 2002, pp. 148-149, emphasis in original,; see also Cook & Campbell, 1979, pp. 126-127 and Campbell & Stanley, 1963, pp. 56-61).

Thus, the classic CCSS writings on research design explicitly note that there are valid alternatives to random assignment, especially in field settings, and that these designs can counter the "reactive effects" endemic to employing random assignment in institutional settings (Campbell & Stanley, 1963, p. 57).[3]

As described more fully below, this cohort control group design could be used by schools to examine the efficacy of new curricular programs for their own particular context. Routinely collected assessment data could be used to compare the achievement of students in cohorts that experienced the new curriculum with those that did not. In contrast to randomized control trials, which may reflect results on relatively few students in settings often quite dissimilar from a particular school district, this design could be explicitly applied to the actual context of a given school or district and use data that were already routinely collected (cf. Engelmann, 2009). This concern relates to the notion of generalization.

The Cumulative Nature of Science and Generalizations

In a lengthy discussion of the role of experimental designs in establishing causality, Shadish and associates noted that

> the strength of experimentation is its ability to illuminate causal inference. The weakness of experimentation is doubt about the extent to which that causal relationship generalizes. … [E]xperiments yield hypothetical and fallible knowledge that is often dependent on context and imbued with many unstated theoretical assumptions (Shadish, et al., pp. 18, 29).

The CCSS discussion of construct validity and external validity directly relates to their concerns with generalization. As noted above, construct validity refers to the relationship between a research design and the theory it is designed to test. For instance, a study of the impact of a particular mathematics curriculum on children's arithmetic skills might utilize assessment operations that involve "story problems," thus assuming certain reading skills. One could argue that the study might lack construct validity, as it applied to concerns with mathematics achievement, for it is not clear if math or reading skills were being assessed. External validity refers to the extent to which a causal relationship holds across a variety of settings and populations. As Shadish and colleagues put it, "external validity…[regards] the extent to which the effect holds over variations in persons, settings, treatments, or outcomes" (Shadish, et al., 2002, p. 22).

CCSS stressed that sampling methods "are insufficient to solve either problem of generalization" (construct or external validity) (Shadish, et al., 2002, p. 24). In an extended discussion they concluded that even relatively complex models such as random selection of units from a population followed by random assignment to treatments

> cannot be advocated as the model for generalized causal inference….Though we unambiguously advocate it when it is feasible, we obviously cannot rely on it as an all-purpose theory of generalized causal inference. So researchers must use

> other theories and tools to explore generalized causal inferences of this type (2002, p. 348, emphasis in original).[4]

They proposed a "grounded theory of generalized causal inference," which they suggested "is more practical than random sampling for daily scientific work," noting that it builds on numerous conceptualizations developed over the last half century (2002, p. 348, citing Brunswik, 1956, Campbell, 1986, Cook, 1990, 1991; Cronbach, 1982; Cronbach & Meehl, 1955). The theory embodies five general principles that they suggested scientists routinely utilize as they make causal generalizations. (See pp. 353-354 of Shadish, et al., 2002, for a summary.) The first is called the principle of "surface similarity," which involves "judging the apparent similarities between the things that they studied and the targets of generalization" (p. 353). For instance, an educational researcher reviewing numerous studies regarding a curricular approach might find that the same results obtain across applications in mathematics and reading. The second principle, "ruling out irrelevancies," involves determining what types of factors do not seem to make a difference, as for instance finding that conclusions hold in a variety of settings (rural versus urban) or with populations of different racial-ethnic backgrounds. In contrast, the third principle, "making discriminations," involves finding discriminations that limit generalization, as in determining that an intervention appears to be more effective with younger children than with older children. The fourth principle, "interpolation and extrapolation," refers to the ways in which scientists may extend findings to groups that are within a sampled range but not yet studied (e.g. children in grade three when evidence is from those in grades two and four) or, more difficult to do, extending to those outside the sample range (e.g. children in grades three to four when the evidence relates to those in grades one to two). Finally, the fifth principle, termed "causal explanation," refers to the development and testing of explanatory theories about the target of generalization, whether they are persons, settings, treatments, or outcomes.

Shadish and associates then described how scientists can develop generalized causal inference, discussing techniques for single studies and multiple studies, or programs of research. For those engaging in a single study, they stressed the centrality of "purposive sampling." One approach is to purposively sample "typical instances," choosing samples of people, settings, times, treatments, and/or outcomes to which one wants to generalize. For instance, a school might be particularly interested in the extent to which a given curriculum affects the reading achievement of students in particular grades in their own localities using test instruments that they have previously found to be highly predictive of later success. Another alternative is to purposively sample "heterogeneous instances," defining the range of persons, settings, treatments or outcomes to which one wants to generalize and creating a sample that reflects that heterogeneity. For instance, a school district might wish to examine the extent to which an intervention impacts the achievement of children from different economic backgrounds and would thus ensure that the study sample included the full range of possibilities.

It is, of course, very difficult for one study to encompass the range of persons, settings, times, treatments, and outcome measures needed to develop causal generalizations, but multiple studies can provide this diversity. Sometimes multiple studies involve the gradual development of understandings and causal inferences, through a "phased model," typically moving from rather small and highly focused controlled experimental designs to tests with more varied settings, subjects, and outcomes. Through purposive sampling these series of studies can address the five principles of causal explanation described above. The Direct Instruction curriculum, which includes our targeted program of Reading Mastery and was first developed in the 1960s, illustrates this approach. The curriculum developed through lengthy and detailed experiments with different populations of students and in different settings. Extensive work also validated the

principles of learning and instruction that provide its theoretical base. Field testing with large and varied populations of students and teachers examined the extent to which it was effective across heterogeneous populations (Collins & Carnine, 1988; Engelmann, 2007; Engelman & Carnine, 1982; Huitt, Monetti, & Hummel, 2009).

The phased model of research is both time-consuming and expensive, and, thus systematic literature reviews are a much more common approach to using multiple studies to develop causal generalizations. In recent years, meta-analyses using quantitative techniques to summarize the average results of studies have become increasingly common. Typically these studies convert the outcomes from each study to a common "effect size" metric. Using the logic embedded in the five principles outlined above, meta-analysts can look for similarities and variations across studies, find gaps in the literature, and test for moderating effects. While meta-analyses are not immune from potential problems related to both internal and external validity (Shadish, et al, 2002, pp. 446-454), it is clear that multiple examinations of a research problem are superior to single tests. As Cook and Campbell put it, "we stress the need for many tests to determine whether a causal proposition has or has not withstood falsification; such determinations cannot be made on one or two failures to achieve predicted results" (1979, p. 31, emphasis in original).

In the later volume Shadish, Cook, and Campbell noted that

> Among scientists, belief in the experiment as the only means to settle disputes about causation is gone, though it is still the preferred method in many circumstances. Gone, too, is the belief that the power experimental methods often displayed in the laboratory would transfer easily to applications in field settings (2002, p. 30, emphasis in original).

They described their work as

> about improving the yield from experiments that take place in complex field settings, both the quality of causal inferences they yield and our ability to

> generalize these inferences to constructs and over variations in persons, settings, treatments, and outcomes (Shadish, et al, 2002, p. 32).

Our analysis attempts to follow in this tradition by showing how the data routinely gathered by schools in response to the accountability and assessment movement can be used to test the effectiveness of curricular interventions. We use the concepts developed within the CCSS tradition to show how these analyses can provide valid tests of causal relationships. We first discuss two designs that could be used in single studies. Then, in the following major section, we summarize results from multiple analyses that used these designs, following the logic of CCSS regarding causal generalizations.

Using Assessment Data to Examine the Impact of Curricular Change

In this section we discuss two ways in which the assessment data often available to school personnel can be analyzed to examine the impact of implementing a new curriculum. Both of these designs use cohort groups. The difference is in the use of additional comparative data, with one using a simple cohort control group design and the other employing time series data over the same historical period as the cohort design as an additional control. In each of the sections we discuss how the designs deal with validity issues described in the CCSS tradition. The data for our examples came from an SRA/McGraw-Hill publication entitled Results with Reading Mastery, which reported results that schools around the country had with this program. For this section we use data from an elementary (K-5) school in rural Alaska, with slightly more than 300 students. Almost three-quarters of the students were Caucasian and one-quarter were of Native American or Alaskan Native descent.

Cohort Control Group (CCG) Design

One set of data from this school was changes in Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores of kindergarten students from the 2002-03 year, before Reading

Mastery (RM) was implemented to the 2006-07 year, after four years of implementation. The percentage of students classified as at "low risk," "some risk," and "at risk" for poor language and reading outcomes in both years based on the established DIBELS benchmarks were given. (See Table 1.) Based on their overall scores, 22 percent of the students in 2002-03 were classified as at risk, but only two percent had this classification in 2006-07. At the other end of the spectrum, less than half of the kindergarten students were at low risk of future difficulties before RM was implemented, but 87 percent were in this category four years later.

[Table 1 About Here]

To analyze the changes in DIBELS scores over time we can employ a cohort control group design (Cook & Campbell, 1979, pp. 126-133; Shadish, et al, 2002, p. 137 and called "recurrent institutional cycle design" by Campbell and Stanley, 1963, pp. 57-60), the logic of which is diagrammed in Figure 1.[5] Following the custom used by Campbell and Stanley (1963) and successors, we use X to refer to the intervention (the curriculum that the school used – Reading Mastery (RM)) and O_i to refer to the observation or assessment (DIBELS). Each line represents a cohort in the school. For this analysis Cohort 1 attended kindergarten before RM was implemented (the 2002-03 school year), while Cohort 2 attended kindergarten after the program was established (2006-07). Thus, data are available for a cohort that had not been exposed to the experimental intervention (O_1) and for a cohort with such exposure (O_2), and the data from O_1 and O_2 can be compared to determine the impact of the intervention (X).

[Figure 1 About Here]

A comparison of proportions test can be used to examine the statistical significance of the differences between these two years. The results for each measure are in the fourth and fifth columns of Table 1 and indicate that all the differences were statistically significant. Also

included is an effect size for each comparison (Cohen's d), and each of these surpassed the usual criterion of educational importance, ranging from .48 to .87. Finally, the lower limit of a one-tailed confidence interval is reported. This may be interpreted as the lower limit of an expected effect size at 95 percent confidence. (e.g. for the "at risk" comparison, $P[D > .25] = .95$, where D is the effect size associated with the difference in at risk status between the two cohorts in a hypothetical larger population of cohorts.) The mathematics involved in these comparisons is simple and can be completed with an excel spreadsheet or an on-line calculator. Details on computations are in Appendix A.

To what extent is this design valid? As noted above, the CCSS literature distinguishes four types of validity with which researchers should be concerned: statistical conclusion, construct, external, and internal validity. As described above, statistical conclusion validity refers "to the appropriate use of statistics to infer whether the presumed independent and dependent variables covary" (Shadish, et al, 2002, p. 37). Most researchers now agree that the best way to handle these issues is to report both inferential and descriptive statistics (such as effect sizes) and to report complete details of the inferential results, as was done above with both the z value and the associated exact probability. Construct validity regards the extent to which inferences from the study can relate to the underlying theory. For this analysis the key question is no doubt the extent to which the DIBELS measure accurately reflects and measures the school's concern with increasing reading skills. Numerous studies indicate the predictive validity of DIBELS measures, both in terms of later reading achievement and comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good, Simmons, & Kame'enui, 2001). External validity refers to the extent the results can be applied to other settings and conditions. While it is unclear to what extent the findings from this small community in Alaska could apply to other settings, the issue for the district itself is

probably more limited. For school officials, the question of external validity probably involves the extent to which the findings can apply to later cohorts of students. This issue of comparability of cohorts is central to determining the final type of validity – internal validity.

No doubt the trickiest element of the cohort control group design revolves around internal validity, inferences regarding whether the observed relationship represents a causal relationship. In response to misunderstandings regarding his original use of this term, Campbell (1957) suggested relabeling the concept as "local moral causal validity." Building on this suggestion, Shadish et al. stressed that "internal validity is about whether a complex and inevitably multivariate treatment package caused a difference in some variable-as-it-was-measured within the particular setting, time frames, and kinds of units that were sampled in a study" (2002, p. 54). In other words, the important question is, within the particular setting of this Alaskan community, as RM was implemented with these two cohorts, did this implementation result in changing reading skills?

The list of potential threats to internal validity should be familiar to all social scientists: 1) ambiguous temporal precedence, 2) selection, 3) history, 4) maturation, 5) regression, 6) attrition, 7) testing, 8) instrumentation, and 9) additive and interactive effects of the preceding threats (Shadish et al., 2002, p. 55). Several of these threats would appear to be unlikely with the cohort control group design. For instance, temporal precedence issues would appear quite unlikely, as long as the new curriculum was implemented after Cohort 1 finished kindergarten. Following the standard testing protocols would help to minimize testing and instrumentation effects. Each of the other threats, however, relates directly to the extent to which the cohorts involved are comparable. In other words, with reference to this example, did kindergarten students in 2006-07 differ systematically from kindergarten students in 2002-03?

As noted above, CCSS provided explicit guidance on this issue. We suggest that the design easily meets three of their four criteria: 1) the earlier cohort did not have the treatment (RM), while the later cohort did; 3) the treatment given to the later cohort applied to all students, and 4) archival records provided comparable data (the DIBELS scores). The most difficult potential issue is their second criterion, ensuring that the "cohorts differ in only minor ways from their contiguous cohorts" (Shadish et al, 2002, p. 149). For instance, changes in reading skills could reflect changes in the demographic characteristics of cohorts if the community experienced dramatic changes in the composition of its residents or if the school had a dramatically altered student body over time. School officials would, of course, be in the best position to judge such changes (and see the discussion in the next section for examples). In addition, however, empirical data could be used to address this question. First, schools routinely gather data on the percentage of enrolled students who receive free and reduced lunch and are members of racial-ethnic minorities, common measures of "at-risk" status, and these data could be examined to determine any differences in demographic characteristics. Second, most schools now have excellent data on the mobility and turn-over of student cohorts and the extent to which members of cohorts have had a "full dosage" of a given intervention. Third, and applicable to this example, the DIBELS system provides measures of skills at the start of schooling, and comparisons of these scores between cohorts would help to establish any differences. If school officials could be convinced that the cohorts were essentially equivalent, then one could logically argue that threats from selection, maturation, regression, and attrition would be minimized.

A more difficult question involves threats to internal validity from history, the impact of events that occur naturally or concurrently with the treatment. Might there be circumstances or events within the broader social-historical context that contributed to these changes other than

changes in the school curriculum? Given the strong legislative and political impetus for school improvement around the nation it would be important for a school to know how they were faring relative to other districts. In other words, even if achievement in this Alaskan school improved after implementing RM, is this improvement greater than what was observed in other schools over the same historical period? This question is addressed by our second design: the cohort control group with comparative historical data (CCG-H) design.

Cohort Control Group Design with Comparative Historical Data

The logic of the Cohort Control Group with Comparative Historical Data Design (CCG-H) is diagrammed in Figure 2. Like the Cohort Control Group (CCG) Design, this design compares outcomes between cohorts with differential exposure to the treatment. Two cohorts from the same setting are compared, with one cohort not exposed to the treatment and the other receiving the treatment. In addition, however, to provide a control for historical change, comparisons are made with cohorts from a larger group that provides an appropriate reference. (This logic is analogous to the modification of the cohort control group design described by Cook and Campbell, 1979, pp. 130 and Shadish, et al., 2002, pp. 149-150.) For schools, this larger group could logically be students within the same state or district, who would presumably be affected by similar opportunities and restrictions related to political and funding issues. This additional comparison group is shown in the bottom lines of Figure 2, with observations paralleling those obtained from both of the cohorts. The logic involved in determining the equivalence of school cohorts for the CCG design would apply here. The addition of the larger comparison group, however, allows a control for history effects. To be considered significant relative to changes in the larger historical context, differences between O_2 and O_1 would have to

be greater than differences between O_4 and O_3, the comparable time points for these observations for the comparison group.

[Figure 2 About Here]

As an example we again use data from the Alaskan school, looking at the percentage of third graders who scored at the proficient or advanced level in reading and writing in the state's annual assessment test. We compared the change in third graders' scores from before the school began to use Reading Mastery (spring, 2003) to 2007, when third graders continuing in the school would have been exposed to the curriculum since kindergarten, to changes over this time period within the state as a whole. We test the hypothesis that the changes over time in the community were greater than the changes over time in the state or, equivalently, that the difference between scores of students in the community and those in the state of Alaska was the same in 2003 as in 2007. In other words, we examine the extent to which third graders' reading achievement increased in our target school after the implementation of the curriculum relative to increases for all third graders within the state of Alaska.

The first panel of Table 2 reports data provided by the SRA report and the comparable state assessment data, which we obtained from the web site for the Alaska State Department of Education. In 2003 the percentage of the school's third graders scoring proficient or advanced was very similar to the percentage for the state in both reading (74-75%) and writing (60%). However, by 2007, the percentages in the school were higher than those in the state (86% versus 80% for reading and 89% versus 77% in writing). The final column of the first panel gives z-scores (standard scores) comparing the proportions in the school to those for all third graders in the state. As would be expected given the greater increase in scores in the school, these values were near zero for 2003, but positive for 2007.

[Table 2 About Here]

The second panel of Table 2 reports effect sizes describing the change in scores from 2003 to 2007 in the school relative to the change in scores in the state. While both effect sizes were positive, only the one associated with writing reached the usual level used to denote educational importance. Also included are t-values testing the null hypothesis that the changes from 2003 to 2007 in the school, relative to changes in the state, equaled zero. The probability level associated with the changes in writing scores approached significance, while that associated with the change in reading scores did not. As would be expected given these t-values, the lower bounds of the uni-dimensional 95 percent confidence intervals were less than zero. (Details on computations are in Appendix A.)

Note that the design used in this analysis explicitly controls for the "history" threat to internal validity. By comparing changes in the community to those that occurred within the state as a whole we adjusted for the changes that were occurring within the state. As described more fully in Appendix A, because achievement scores were increasing in both the community and the state, the estimates of change in the community would have been too high if this control were not included. If scores in the state had been declining during this time period omitting the control would have under-estimated the magnitude of change in the school.

Summary

This section demonstrated analyses with two research designs that schools can use to test the impact of curricular interventions on student achievement with assessment data that they routinely gather. Both designs compare the achievement of student cohorts, an approach that the classical literature suggests is appropriate in institutional settings such as schools. Most important, in these settings these designs could provide more internal validity than randomized

control trials because they may minimize the possibility of reactivity, a serious concern when conducting experiments within intact organizations. We also showed how, with some types of data, this design can be expanded to include comparative data from similar cohorts in a larger entity representative of the sample groups, such as their state or perhaps other schools in their district. This cohort control group with comparative historical data (CCG-H) design thus helps promote internal validity by helping to control for historical changes that might be affecting the cohorts. In short, we suggest that these designs provide valid ways to assess the impact of curricular innovations and avoid the possible reactive effects that can occur by attempting to randomly assign students within intact groups.

However, as described previously, the CCSS research design literature stresses that the results of one study are far from sufficient to determine a causal relationship. While individual schools and districts may be convinced by the results of their own intervention and analysis, other schools and districts, as well as researchers, would want additional evidence. They would be concerned with "external validity," the extent to which results of one study may generalize to other students, settings, and outcomes. We suggest that the logic involved in Shadish, Cook, and Campbell's (2002) "grounded theory of generalized causal inference" can be used to develop this understanding. This can occur by examining the results of multiple observations, numerous tests of the efficacy of a curriculum or intervention. In the next section we use this logic to examine several dozen analyses of the implementation of RM from around the country using the CCG and CCG-H designs.

Generalizing from Replications

Our analyses come from data in the report disseminated by SRA/McGraw Hill, the publisher of Reading Mastery. We use the logic outlined by Shadish, Cook, and Campbell (2002)

to compare results obtained across different settings and assessment techniques, examining the extent to which they were similar across schools with different student characteristics and modes of assessment. Further details on the results for each site are given in Appendix B.

Of the 26 reports that were reviewed, eight were deemed to fall outside the parameters of the CCG or CCG-H design because they did not meet the criteria outlined above.[6] Our analysis below focuses on the remaining 18 reports. The second column of Table 3 reports the study design that was determined appropriate for the data. Five of the reports had data appropriate for the CCG design, eleven had data appropriate for analysis with the CCG-H design, and two provided data appropriate for both designs.

[Table 3 About Here]

Table 3 also summarizes the characteristics of the 18 reports that were reviewed, including geographic location, the grades served by the school or district, the number of students, and the type of assessment used. It can be seen that the reports came from all regions of the country and involved data from individual schools as well as several schools within one district. The schools varied in size from just over 100 to over 1,000. Assessments described included curriculum-based measures (DIBELS in all instances), norm-referenced tests such as the ITBS and CTBS, and state assessments. Data were given for students in grades K-6 as well as special education students (grades K-8), but most often involved third and fourth graders, perhaps because that is often the age at which state assessments are first given. All of the schools implemented RM, and reading achievement was the focus of the assessments. Seven of the sites also used various other Direct Instruction programs for areas such as language, spelling, and/or interventions for older children.[7]

Most of the reports provided information about several grades and/or schools and thus yielded a number of possible comparisons (60 in all). However, comments within two of the reports indicated that some comparisons would be less valid than others, specifically addressing the second criterion listed by Shadish and associates – that the "cohorts differ in only minor ways from their contiguous cohorts" (Shadish, et al, 2002, p. 149). In these situations, the school officials noted that one of the cohorts for which data were presented differed from other cohorts due to an unexpected influx of new students. To account for this possible non-equivalence, we completed analyses with and without these potentially incomparable cohorts. The sample that omits the incomparable cohorts is referred to below as the "reduced sample."

Table 4 provides summary descriptive statistics on the characteristics of the schools (or, in some cases, districts) that the students attended and the effect sizes associated with the analyses. The top part of the table reports data when individual comparisons were used as the unit of analysis, while the bottom part reports data using the report (site) as the unit of analysis. Thus the information in the top part of the table can represent multiple analyses from a school or district, while the information in the bottom reflects one summary number for each site. The data indicate that the analyses involved schools and districts with a wide range of characteristics, thus providing the large number and variety of comparisons required by the logic of the grounded theory of generalized causal inference. For instance, they ranged from schools where less than a quarter of students qualified for free or reduced lunch to those where all students qualified. Similarly, they varied in racial-ethnic composition, from schools that were predominantly minority – Hispanic, African American, or Native American – to those that were virtually all Caucasian in composition. They also varied in the percentage of students who were English Language learners. On average, the sites had implemented RM for four to five years, but ranged

from as little as two years to nine. The comparisons involved students who had begun their instruction in RM in kindergarten through ones involving students who began the program only in the upper elementary grades. Some involved only a few students (a minimum of 34), while others involved well more than a thousand. The average was slightly less than 200.

[Table 4 About Here]

The last two lines of each panel of Table 4 give summary statistics regarding the effect sizes for both the total sample and the reduced sample, which omitted the cohorts reported to be non-comparable. The effect sizes ranged from -.26 to 1.66 for the single analyses and .04 to 1.14 for the analyses aggregated to the site/report level. As would be expected, the minimum values were slightly higher in the reduced samples. However, the means changed only slightly and ranged from .47 to .57 across the four sets of analyses, well above the level usually considered educationally important.

The grounded theory of generalized causal inference described by Shadish, Cook, and Campbell (2002) and summarized above calls for multiple comparisons of results, looking for similarities and differences across numerous studies, searching for factors that might suggest moderating influences or conditions that limit the generalizability of findings. The wide range of settings in the SRA reports allowed the comparison of findings across schools of different size, with different socio-demographic student characteristics and varying experience with the curriculum, including the years of implementation and utilization of other programs. They also allowed us to examine the relation of design characteristics to the results. Table 5 provides a summary of these comparisons. It reports the average effect size calculated for different sets of comparisons, using the four sample groups used in Table 4. The last rows of the table report the maximum and minimum average values across the subgroups.

[Table 5 About Here]

The results in Table 5 indicate a good deal of consistency across the various analyses. In all of the sub-groups the average effect sizes surpassed the usual criterion for educationally important effects. At the same time there were some variations. The largest differences appeared to be those related to the design and assessments used. As would be expected, effects with the CCG-H design, which explicitly controls for historical changes, were smaller than those with the CCG design. The analyses using state assessments, virtually all of which used the CCG-H design, also yielded slightly smaller effect sizes. In addition, effect sizes tended to be higher in analyses of smaller schools and sites. There were very small differences in effect sizes of schools with different socio-demographic characteristics or with those with different levels of experience with the curriculum. In other words, using terms from the CCSS research design literature, these multiple comparisons suggested that the results regarding the efficacy of the RM curriculum appear to generalize to schools with different compositions and settings and with different types of assessments.

Summary and Discussion

This paper used the classic research design literature to demonstrate how assessment data that are routinely collected by schools can be used to analyze the impact of curricular interventions on student achievement. This work shows how two areas of work fostered by the No Child Left Behind Act – the requirements of regular assessment of student achievement and the call for "scientifically based research" – can be merged in a way that provides useful information to school officials and researchers in a cost-effective manner. Following the suggestions of Campbell and Stanley (1963), Cook and Campbell (1979), and Shadish, Cook, and Campbell (2002), the classic textbooks on experimental design, we showed how

comparisons of intact cohorts within schools can result in designs that help counter the problems of reactivity that are likely to appear when randomized trials are implemented within institutional settings and thus promote internal validity. In addition, we used the logic of Shadish, Cook, and Campbell's grounded theory of causal inference to systematically compare several dozen tests of the efficacy of one curriculum across settings with widely differing characteristics, promoting external validity or generalization of results. Below, we briefly discuss how our work relates to discussions of methodologies that should be employed by educational researchers, potential limitations of this approach and how these could be addressed, and implications of this work for practitioners and policy makers.

Methodology in Education Research

As noted above, the educational establishment has indicated a strong preference for randomized control trials as a way to establish "scientific validity" of research results. Several authors, including those in the CCSS tradition (Cook, 2002, 2003) have chastised educational researchers for appearing to avoid randomized assignment (see also Sherman, 2003; Towne, Wise, & Winters, 2005) This preference for randomization as a "gold standard" is no doubt well-meaning and based on the general outlines of the CCSS tradition. However, as noted above, when viewed in its entirety, the CCSS literature is quite clear in noting the limits of randomized assignment in field settings, such as schools. More important, as discussed above, it offers alternative designs that promote internal validity, can utilize the type of data schools routinely collect, and can answer the types of questions with which schools may be most often concerned – how the implementation of a new curriculum in their setting, with their teachers and their students, has affected students' achievement.

Thus, our work supplements the growing body of scholarship that suggests that there has been an over-emphasis on randomized control trials in educational research, as well as other field settings. One element of this literature is empirical in nature and has compared the results obtained with randomized trials and quasi-experimental designs. Some have found systematic differences between these types of studies (e.g. Agordini & Dynarski, 2004; Glazerman, Levy, & Myers, 2003; Weisburd, Lum, & Petrosino's 2001), while others suggest that the results are minimal when appropriate statistical controls are employed (Heinsman & Shadish, 1996; Shadish, Clark, & Steiner, 2008; Slavin, 2008).[8] Some scholars have also focused on theoretical issues regarding research designs and, especially, those that are appropriate given the "social structure of instruction" and ways in which instruction is nested within classrooms and schools, resulting in multiple influences on student outcomes (Raudenbush, 2008; Sloane, 2008a,b).

An alternative approach that has appeared within the critical literature, echoing Shadish et al's (2002) call for a "phased model" of work, is programmatic research, series of studies that address an issue, such as the impact of a curriculum. Scholars urging this approach stress that programs of research should use a variety of methodological approaches to develop a fuller and complete picture related to a research question (Odom, Brantlinger, Gersten, Horner, Thompson, & Harris, 2005. Robinson, 2004). Sloane (2008a, b) argued that a programmatic approach more closely parallels the model used in drug efficacy research, which begins with basic research and small feasibility studies and proceeds to small-scale efficacy tests often using randomized trials. These are followed by larger scale experimental studies to test effectiveness in broader populations and implementation, scaling, and sustainability research with much larger groups, often using quasi-experimental designs. All of these researchers stress that randomized control trials are more appropriate at some stages of the research program than others.

Interestingly, research studies of Direct Instruction curriculum, including RM, illustrate this progression. In a recent review of the literature Coughlin (forthcoming) has found that randomized control trials were most common in the 1970s through the early 1990s. After that time, as would be suggested by Sloane's description, the literature moved to larger implementation trials, studies of how the curriculum could be used in whole systems (scaling studies) and how it could be sustained over time (sustainability research, the last element in Sloane's description). We suggest that the analytic model described in this paper could be seen as reflecting the later stages in Sloane's description of research phases, especially implementation trials, large-scale quasi-experiments "to determine the effectiveness of the … intervention under real world conditions" (Sloane, 2008a, p. 628).

Limitations and Potential Modifications

There are of course, several potential limitations to the approach described in this document. First, assessment data are often reported as the percentage of students meeting certain benchmarks. Compared to analyses with scale scores, these percentages have less variability and are more apt to reflect ceiling and floor effects. To avoid this problem, schools could use, whenever possible, the actual scale scores. This could also counteract the issue of states setting unusually high or low benchmarks, such as a "Lake Woebegone" effect, with cut-offs that result in the vast majority of students being "above average."

Second, the designs could be criticized for not controlling for the threat to internal validity that comes from regression toward the mean. This threat is most likely to occur when a sample represents groups that are at extremes of the distribution (higher or lower than the overall mean) and chance variation produces scores closer to the overall average. The use of the historical control group in the CCG-H design and the use of several years of data for baseline

and follow-up measures can help control for this possibility. In addition, analysts could directly address this concern by selecting a sub-sample of comparable schools from a larger set of comparative data. For instance, using data sets often made available by state departments of education, they could choose other schools with similar levels of baseline achievement as a comparison group.

Third, these designs ignore the issue of treatment fidelity and the extent to which the programs have been implemented as designed. The analyses outlined in this paper describe the effect of programs as they were implemented within a given setting. Variation in fidelity from the optimum would logically affect both mean levels of achievement as well as variations across classrooms and/or schools. Thus, one way to examine variations in fidelity could be to directly study variations in outcomes across classrooms and schools and potential factors that might be contributing to these differences.

Fourth, even though the approach described in this paper helps to control for reactivity that might be related to random assignment within a school, it does not remove the possibility of other sources of this phenomenon. It is not unreasonable to expect that the ways in which teachers taught and students learned in one year would influence the ways that they taught and learned in subsequent years, even if a new curriculum were implemented. This is one reason that experts in implementation suggest that a substantial amount of time may be needed before new programs can be fully established, or stabilized, within a school (Engelmann & Engelmann, 2004). One way in which researchers might handle this issue with the current designs is to wait to examine results until a program has been in place for several years.

Fifth, the use of state-level data as a control in the CCG-H design, while cheap and efficient, is, of course, far from precise, for it is possible that other schools in a state were also

using the curriculum under study. Fortunately, however, this possibility would lead to smaller differences between the two groups, thus providing a conservative test of the curriculum's impact.

Finally, it is possible that the set of data analyzed in this document could be highly selective, for the publisher might have chosen only to report data from successful sites. As a result, the reported effect sizes might over-state the impact of the curriculum. Interestingly, however, the effect sizes found in our comparative analysis are similar to those reported in meta-analyses of studies of Direct Instruction programs. For instance, Hattie's recent meta-analysis of meta-analyses incorporated results from 304 studies of DI programs, with 597 effects and over 42,000 students, and concluded that the average effect size associated with DI was .59 (1999, pp. 206-207). This value is slightly higher than the values reported above. Importantly, as with the other studies, our analysis found no difference in the effects across a variety of settings.

Implications for Practitioners and Policy Makers

The approach described in this paper can be seen as a way to empower school practitioners. They could easily employ the techniques described above. The calculations are easy to complete, and on-line calculators can usually be used to compute the statistics. In addition, the approach is cheap and efficient. The calculations use readily available data, which are, for the most part, in the public domain and/or already gathered in response to government regulations. Perhaps most important, because the analyses involve locally gathered data, school officials can be confident that they apply to their setting. In short, the analyses described in this report can empower local school officials – giving them the ability to determine, with relative accuracy and confidence, the extent to which curricular interventions produced desired results within their local setting.

The discussion above could also be used to support calls for those who control educational research agendas, both research funding and the reviews of research findings, to move away from a single-minded focus on randomized control trials. While this focus is justified as a way to guarantee "scientific" research, the CCSS literature is clear in its suggestion that other types of designs are often more appropriate in field settings such as schools and that it is much more important to examine results from a series of investigations in a range of settings rather than to seek out only a very few studies that meet limited and strict criteria.[9]

It is perhaps reasonable to expect that at least some educational researchers will resist the suggestions of this report, for conducting randomized control trials has become big business within that sector, employing dozens of researchers and their assistants, providing support for large research institutes, and elevating scientific reputations of the "experts" vis-à-vis practitioners. In the long run, however, one could suggest that students – those whom we most want to benefit – would be best served by using assessment and analysis procedures that are cost-effective and also provide the greatest internal and external validity. To do anything less could be termed educational, as well as scientific, malpractice.

Figure 1
Cohort Control Group Design

Cohort 1	O_1		
Cohort 2		X	O_2

Figure 2
Recurrent Institutional Cycle Design (Matched Cohort Design) with Comparative Historical Data

Cohort 1	O_1		
Cohort 2		X_2	O_2
Comparison Group Cohort 1-C	O_3		
Comparison Group Cohort 2-C			O_4

Table 1

At Risk Status Based on End of Year Kindergarten DIBELS Scores, 2002-03 and 2006-07, Rural Alaskan Elementary School

	2002-03	2006-07	z	p (one-tail)	Effect Size	Min. C.I.
At Risk	22	2	2.77	0.006	0.62	0.25
Some risk	32	12	2.17	0.002	0.48	0.12
Low risk	46	87	4.15	<.0001	0.87	0.52

Note: Details on calculations are given in Appendix A. Reading Mastery had not been implemented in 2003-03, but had been implemented for 4 years by the end of 2006-07. One-tail probabilities were used to match the question of theoretical interest.

Table 2: Changes in Third Graders Reading and Writing Scores on State Assessments, 2003-2007, An Alaskan Elementary School and State of Alaska

Proportion of Students Reaching Proficient or Advanced Benchmarks			
Reading	School	Alaska	Z-score
2003	0.75	0.74	0.01
2007	0.86	0.80	0.15
Writing			
2003	0.60	0.60	0.00
2007	0.89	0.77	0.29

Change 2003 to 2007		
	Reading	Writing
Effect Size	0.13	0.28
t-value	0.73	1.55
Df	102	102
prob.	0.24	0.06
Min. C.I.	-0.17	-0.02

Note: There were 41 third graders in 2003 and 63 in 2007. In 2003 the third graders had no exposure to Reading Mastery; in 2007 third graders in the school had been exposed to RM since kindergarten. Probabilities are one-tail, reflecting the hypothesis that exposure to RM will increase achievement.

Table 3
Characteristics of SRA/McGraw Hill Reports on RM Interventions

Report #	Design	# of Comparisons	State	Grades in School	# of Students	Assessment	Grades for Analysis
1	CCG & CCG-H	5	Alaska	K-5	300	CBM and SA	K, 3
2	CCG-H	4	Alabama	K-6	470	SA	3- 6
3	CCG	1	Arizona	PreK-3	813	NR	3
5	CCGH	6	Colorado	K-5	135	SA	3- 5
6	CCGH	1	Delaware	PreK-7	160	SA	3
7	CCGH	11	Florida	Pre-K-5	723	SA	3 - 5
8	CCG	5	Florida	K-6	593	NR	3 - 6
10	CCGH	1	Florida	K-5	1577	SA	4
11	CCGH	1	Florida	PreK-5	554	SA	4
13	CCGH	1	Kentucky	PreK-5	250	SA	4
14	CCG	5	Kent. (5 schools)	Pre-k - 6	318 (average)	NR	3
15	CCG	4	Minnesota	K-12	216	CBM	K-3
17	CCGH	6	N. Carolina (2 schools)	gr 3 - 5	675	SA	3 - 5
18	CCGH	2	Ohio	K-6	208	SA	4
19	CCGH	3	Oregon	K-12	2059 (district)	SA	3
21	CCG	1	Tennessee	K-12	9918 (district)	SA	3-8, SPED only
23	CCG & CCGH	3	Washington	K-6	652	NR	3 - 4
24	CCGH	1	Wisconsin	PreK-5	374	SA	4

Note: Reports numbered 7 and 19 also provided information that allowed comparison of students with disadvantaged status. Report numbered 17 also gave data for a third school, but provided no information for the period before implementation of RM, so those data were not used. Numbers are not contiguous because some of the 26 reports that were reviewed did not have data appropriate for the CCG or CCG-H analysis. CBM refers to curriculum-based measures, SA refers to state assessments, and NR refers to norm-referenced tests.

Table 4

Descriptive Statistics, Case Study Reports

Comparisons as Unit of Analysis (n = 60)				
	Min.	Max.	Mean	S.D.
School/District Size	135	9918	713.6	1282.9
FRL %	21	100	68.3	21.0
African Amer. %	0	94	17.1	22.3
Caucasian %	0	97	54.5	30.0
Hispanic %	0	90	17.4	20.9
Asian %	0	7	1.7	2.7
Native American %	0	100	9.2	25.4
ELL %	0	58	14.8	20.1
Years RM implemented	2	9	3.7	1.7
Grade last cohort began RM	0	5	1.1	1.5
N in comparison	34	1800	192	251
Effect size (d)	-.26	1.66	.47	.38
Effect size (d), reduced sample (n=58)	.00	1.66	.48	.37
Site as Unit of Analysis (n = 18)				
	Min.	Max.	Mean	s.d.
School/District Size	135	9918	1108.3	2253.9
FRL %	21	100	67.8	25.4
African Amer. %	.0	94.0	23.1	31.8
Caucasian %	.0	97.0	51.4	34.0
Hispanic %	.0	90.0	16.4	24.4
Asian %	.0	7.0	1.1	2.0
Native American %	.0	100.0	7.6	23.8
ELL %	.0	58.0	11.1	20.1
Years RM implemented	2.0	9.0	4.40	2.35
Grade last cohort began RM	.0	3.5	.99	1.31
N in comparison	36	3236	649	824
Effect size (d)	.04	1.14	.56	.34
Effect size (d), reduced sample	.14	1.14	.57	.34

Note: The data with site as the unit of analysis represent the average value across the individual comparisons within a site. The reduced sample omits the cohorts noted in the SRA report as being non-comparable to other cohorts. N = 58 for the comparisons for the reduced sample, and n remains at 18 for the site statistics.

Table 5

Effect Sizes by Sample and Sub-Groups

	Comparisons as Unit of Analysis - All Cases		Comparisons as Unit of Analysis - Reduced Sample		Sites as Unit of Analysis - All Cases		Sites as Unit of Analysis - Reduced Sample	
School/Site Size	Mean	N	Mean	N	Mean	N	Mean	N
135 to 469	0.64	27	0.65	26	0.74	8	0.76	8
469 to 9918	0.32	33	0.34	32	0.41	10	0.42	10
Free and Reduced Lunch %								
21 to 71%	0.47	26	0.47	26	0.50	8	0.50	8
72 to 100%	0.47	34	0.48	32	0.61	10	0.63	10
% African American								
0 to 29%	0.48	51	0.48	51	0.55	13	0.55	13
30 to 94%	0.42	9	0.49	7	0.57	5	0.63	5
% Caucasian								
0 to 49%	0.47	34	0.48	32	0.61	10	0.63	10
50 to 97%	0.47	26	0.47	26	0.50	8	0.50	8
% Hispanic								
0 to 29%	0.50	39	0.52	37	0.55	14	0.57	14
30 to 90%	0.40	21	0.4	21	0.58	4	0.58	4
% ELL								
0 to 29%	0.52	37	0.54	35	0.55	14	0.57	14
30 to 58%	0.40	21	0.40	21	0.58	4	0.58	4
Comparison cohort began RM in Kindergarten								
No	0.39	27	0.41	26	0.53	7	0.54	7
Yes	0.51	32	0.51	31	0.48	9	0.50	9
School Implemented RM for 4 or more years								
No	0.48	31	0.50	30	0.64	9	0.65	9
Yes	0.45	29	0.45	28	0.48	9	0.50	9
Number of students in comparisons								
22 to 100	0.76	18	0.77	17	0.92	2	0.92	2
101 to 1800	0.34	42	0.35	41	0.51	16	0.53	16
Design Type								
CCG	0.68	20	0.68	20	-----	-----	-----	-----
CCG=H	0.36	40	0.37	38	-----	-----	-----	-----
Type of Assessment								
CBM	0.79	7	0.79	7	-----	-----	-----	-----
Norm refer.	0.58	12	0.58	12	-----	-----	-----	-----
State Assess.	0.38	41	0.39	39	-----	-----	-----	-----
Used Other DI Programs								
Only RM	0.49	42	0.51	41	0.55	11	0.56	11

Other DI	0.41	18	0.40	17	0.57	7	0.59	37
Summary								
Minimum	0.32		0.34		0.41		0.42	
Maximum	0.79		0.79		0.92		0.92	

Note: Averages for categories of design type and type of assessment could not be computed when the site was the unit of analysis because some sites had multiple designs and assessment types.

References

Agaodino, R. & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs. The Review of Economics and Statistics, 86, 180-194.

Boruch, R. (2005). Better evaluation for evidence-based policy: Place randomized trials in education, criminology, welfare, and health. Annals, AAPS, 599 (May), 6-18.

Brunswick, E. (1956). Perception and the representative design of psychological experiments (2nd ed.) Berkeley: University of California Press.

Campbell, D.T. & Stanley, J.C. (1963). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

Campbell, D.T. (1986). Relabeling internal and external validity for applied social scientists. In W.M.K. Trochim (Ed.) Advances in quasi-experimental design and analysis (pp. 67-77). San Francisco: Jossey-Bass.

Cartwright, N. (2007). Are RCTs the gold standard? BioSocieties, 2, 11-20.

Clay, R.A. (2010). More than one way to measure," Monitor on Psychology, 41 (September), 52-55.

Cook, T.D. (1990). The generalization of causal connections: Multiple theories in search of clear practice. In L. Sechrest, E. Perrin, & J. Bunker (Eds.), Research methodology: Strengthening causal interpretations of nonexperimental data (DHHS Publication No. PHS 90-3454, pp. 9-31). Rockville, MD: Department of Health and Human Services.

Cook, T.D. (1991). Clarifying the warrant for generalized causal inferences in quasi-experimentation. In M.W. McLaughlin & D.C. Phillips (Eds.), Evaluation and education: At quarter-century (pp. 115-144). Chicago: National Society for the Study of Education.

Cook, T.D. (2005). Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. Annals, AAPS, 599 (May), 176-198.

Cook, T.D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. Educational Evaluation and Policy Analysis, 24, 175-199.

Cook, T.D. (2003). Why have educational evaluators chosen not to do randomized experiments, Annals AAPS, 589, 114-149.

Cook, T.D. & Campbell, D.T. (1979). Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Cook, T.D. & Steiner, P.M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. Psychological Methods, 15, 56-58.

Cook, T.D., Scriven, M., Coryn, C.L.S., & Evergreen, S.D.H. (2009). Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven. American Journal of Evaluation, published online December 4, 2009.

Collins, M. & Carnine, D. (1988). Evaluating the field test revision process by comparing two versions of a reasoning skills CAI program. Journal of Learning Disabilities, 21, 375-379.

Couglin, C. (forthcoming). A review of the Direct Instruction Literature: A Four Decade Program of Research. John Lloyd (ed.) Direct Instruction and Evidence-Based Practice. Eugene, Oregon: ADI Press.

Cronbach, L.J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.
Engelmann, S. (2007). Teaching needy kids in our backward system: 42 Years of trying. Eugene, Oregon: ADI Press.
Engelmann, S.E. & Carnine, D. (1982). Theory of instruction: Principles and applications. New York: Irvington Publishers.
Engelmann, S.E. & Engelmann, K.E. (2004). Impediments to scaling up effective comprehensive school reform models. In T.K. Glennan, Jr., S.J. Bodilly, J.R. Galegher, K.A. Kerr (Eds.). Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions (pp. 107-133). Santa Monica, CA: Rand.
Engelmann, Z. (2009). Socrates on Gold Standard Experiments. http://www.zigsite.com/PDFs/socrates5mm2.pdf
Fayer, H. (2005). Place randomized trials: Experimental tests of public policy. Annals, AAPS, 599 (May), 272-291.
Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Scientific Studies of Reading, 5(3), 239-256.
Glazerman, S., Levy, D.M., & Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. Annals, AAPSS, 589, 63-93.
Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. Scientific Studies of Reading, 5 (3), 257-288.
Heinsman, D.T. & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? Psychological Methods, 1, 154-169.
Huitt, W.G., Monetti, D.M., & Hummel, J.H. (2009). Direct approach to instruction. In C.Reigeluth & A. Carr-Chellman (Eds.), Instructional-Design Theories and Models: Volume III, Building a Common Knowledge Base (pp. 73-98). Mahwah, NJ: Lawrence Erlbaum.
Julnes, G. & Rog, D.J. (2007). Current federal policies and controversies over methodology in evaluation. New Directions for Evaluation, 113 (spring), 1-12.
Maxwell, S.E. (2010). Introduction to the special section on Campbell's and Rubin's conceptualizations of causality, Psychological Methods, 15, 1-2.
McMillan, J.H. (2007). Randomized field trials and internal validity: Not so fast my friend. Practical Assessment, Research & Evaluation, 12 (15). http://pareonline.net/pdf/v12n15.pdf
Millenson, M. (1997) Demanding medical excellence: Doctors and accountability in the information age. Chicago: University of Chicago Press.
Miron, G. (2005). The constructive use of existing data and research for evaluating charter schools, Paper prepared for the Symposium on the Use of School-Level Data for Evaluating Federal Education Programs, December 8-9, 2005, Washington, D.C.
Morgan, S.L. & Winship, C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research. New York: Cambridge University Press.
National Mathematics Advisory Panel.(NMAP) (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, D.C.: U.S. Department of Education.

Noble, J.H., Jr. (2006). Meta-analysis: Methods, strengths, weaknesses, and political uses. Journal of Laboratory and Clinical Medicine, 147: 7-20.
Odom, S.L., Brantlinger, E., Gersten, R., Horner, R.H., Thompson, B., & Harris, K.R. (2005). Research in special education: Scientific methods and evidence-based practices. Exceptional Children, 71, 137-148.
Raudenbush, S.W. (2008). Advancing educational policy by advancing research on instruction. American Educational Research Journal, 45, 206-230.
Robinson, D.H. (2004). Scientific research is programmatic. Pp. 121-128 in Scientifically-based Education Research and Federal Funding Agencies: The Case of the No Child Left Behind Legislation. Information Age Publishing.
Rubin, D.B. (2010). Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010). Psychological Methods, 15, 38-46.
St. Pierre, E.A. (2006). Scientifically based research in education: Epistemology and Ethics. Adult Education Quarterly, 56, 239-266.
Sampson, R.J. (2010). Gold standard myths: Observations on the experimental turn in quantitative criminology. Journal of Quantitative Criminology, published online September 2010.
Schwandt, T.A. (2005). A diagnostic reading of scientifically based research for education. Educational Theory, 55, 285-305.
Scriven, Michael. (n.d.). The Logic of Causal Investigations. Unpublished paper, Western Michigan University.
Scriven, M. (2005). Can we infer causation from cross-sectional data? National Academy of Sciences. http://www7.nationalacademies.org/bota/School-Level%20Data_Michael%20Scriven-Paper.pdf.
Scriven, M. (2008). A summative evaluation of RCT methodology: And an alternative approach to causal research. Journal of MultiDisciplinary Evaluation, 5, 11-24.
Shadish, W.R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. Psychological Methods, 15, 3-17.
Shadish, W.R. & Cook, T.D. (2009). The renaissance of field experimentation in evaluating interventions. Annual Review of Psychology, 60, 607-629.
Shadish, W.R., Clark, M.H., & Steiner, P.M. (2008) Can randomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. Journal of the American Statistical Association, 103, 1334-1343.
Shadish, W.R.. Cook, T.D. & Campbell, D.T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin.
Sherman, L.W. (2003). Misleading evidence and evidence-led policy: Making social science more experimental. Annals, AAPSS, 589 (September), 6-19.
Slavin, R.E. (2008). What works? Issues in synthesizing educational program evaluations. Educational Research, 37, 5-14.
Sloane, F. (2008a). Randomized trials in mathematics education: Recalibrating the proposed high watermark. Educational Researcher, 37, 624-630.
Sloane, F. (2008b). Through the looking glass: Experiments, quasi-experiments, and the medical model. Educational Research, 37, 41-46.
Towne, L., Wise, L.L., & Winters, T.M. (2005). Advancing Scientific Research in Education. Washington, D.C.: National Academies Press.

U. S. Department of Education. (2006). Improving Teacher Quality State Grants, ESEA Title II, Part A, Non-Regulatory Guidance. Washington, D.C.: U.S. Department of Education, October 5, 2006, http://www2.ed.gov/programs/teacherqual/guidance.pdf

Weisburd, D., Lum, C.M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? Annals, AAPSS, 578, 50-70.

West, S.G. & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. Psychological Methods, 15, 18-37.

What Works Clearinghouse (2008). Procedures and Standards Handbook (Version 2.0). http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf

Whitehurst, G.J. (2003). The Institute of Education Sciences: New wine, new bottles. Paper presented at the 2003 Annual Meeting of the American Educational Research Association, April 22, 2003.

Williams, B.A, (2010). Perils of evidence-based medicine. Perspectives in Biology and Medicine, 53, 106-120.

Winship, C. & Morgan, S.L. (1999). The estimation of causal effects from observational data. Annual Review of Sociology, 25: 659-706.

Yin, R.K. & Davis, D. (2007). Adding new dimensions to case study evaluations: The case of evaluating comprehensive reforms. New Directions for Evaluation, 10, 75-93

Zdep, S.M. & Irvine, S.H. (1970). A reverse Hawthorne effect in educational evaluation. Journal of School Psychology, 8, 89-95.

Appendix A
Statistical Calculations

This appendix describes the statistical calculations used to assess data in the cohort control group (CCG) and the cohort control group with historical comparisons (CCG-H) designs. The data presented are from the Alaskan school discussed in the second major section of the paper.

Analyzing the Cohort Control Group (CCG) Design

As described in the text, the question of interest with the data in Table 1 is the extent to which spring DIBELS scores of the 2007 cohort differed from those of the 2003 cohort. The cohorts can be considered statistically independent and thus a simple difference of proportions test can be used. (If mean scores were available then a difference of means test would be appropriate.) The hypotheses to be tested are:

Ho: $p_{2003} = p_{2007}$ versus

H_1: $p_{2003} < p_{2007}$.

That is, we wish to test the hypothesis that the proportion of students meeting the benchmarks (in the case of being at low risk) is greater in 2007 than in 2003.

Such tests can be easily accomplished with on-line calculators. (Traditionalists could, of course, use simple pencil and paper, calculator, or excel-aided computations; however, the on-line calculators provide important safeguards against computational errors.) For the results in this paper we used the calculator available at Dimension Research, Inc. (http://www.dimensionresearch.com/resources/ resources_overview.html), and all results obtained were equivalent, within rounding error, to those calculated by hand. To provide more exact estimates of the associated probabilities we used a different on-line calculator

(http://faculty.vassar.edu/lowry/tabs.html#t). To adjust for variations in sample size we used the usual formula for comparisons [df = $(n_1 + n_2) - 2$].

We used Cohen's d as a measure of effect size, defined as the difference between the proportions or means divided by the common standard deviation. In contrast, the z or t scores used to test hypotheses about the differences between these values is calculated by dividing the difference by an estimate of the standard error. (The standard error is inversely related to the sample size, and is equal to the standard deviation divided by a function of n.) To minimize computational error, we again used an on-line calculator, choosing to employ one that converts the z-scores to effect sizes (http://www.uccs.edu/~faculty/lbecker/). Again, the results obtained were equivalent, within sampling error, to those obtained with pencil and paper calculations.

Finally, we used these data to compute uni-dimensional confidence intervals, given in the last column of Table 1. These confidence intervals tell the minimum value of the effect size, at a given level of confidence, which we have arbitrarily taken at .95. Using the logic of confidence intervals, if the sample value of the effect size is used as an estimate of the population value,

$$P [D > d - (1.645)(s.e.)] = .95 \quad (1),$$

where d is the sample value of the effect size, D is the population value of the effect size, and s.e. refers to the standard error. The standard error can easily be calculated from the values of d and z as in

$$s.e. = d/z \quad (2).$$

Thus, for the proportions of students at risk,

$$s.e. = d/z = .62/2.77 = .22, \text{ and} \quad (3)$$

$$P[D > (.62 - (1.645*.22)] = P [D > .25] = .95 \quad (4).$$

That is, we can be 95 percent confident that the effect size in a hypothetical larger population is greater than .25. Note also that to maintain comparability, all effect sizes and minimal values of the confidence intervals were coded so that positive values indicate that the students with the RM program had higher achievement and negative values indicate lower achievement.

Analyzing the Cohort Control Group with Historical Comparisons (CCG-H) Design

The basic logic of the cohort control group with historical comparisons design was summarized in Figure 2 in the text. In this design the differences between cohorts within a school are compared with differences between cohorts in a larger group to which the cohorts belong. In the examples used in this report, because the dependent measure was third graders' scores on the state assessment tests, the larger comparison group was third grade students in the state. Table 2 summarized the data used in the analysis, showing the percentage of third graders in the school and in the state who had scores on the reading and writing state assessments that reached or surpassed the "proficient" level in 2003 and in 2007. The school implemented RM after the 2003 testing and thus the change from 2003 to 2007 demonstrates the difference that occurred after that implementation. During that period schools throughout the state were trying to improve achievement, and the increase in the state scores illustrates that change. To get an accurate picture of the change in the school that might be attributed to RM it is important to control for these more general changes in the state. Our analysis involved three steps: 1) Comparing the school values to those for the state using simple standard deviation or z-scores; 2) Using these scores to calculate effect sizes that describe the change over time in the school, controlling for change within the state; and 3) Calculating t-ratios that can be used to compute the statistical significance of the change and confidence intervals regarding its magnitude. Each of these steps is described below.

Comparing school and state values using standard scores

Standard scores, or z-scores, were used to compare the scores of a sample with a population in the first step of the analysis. The formula for a z-score is

$Z = (M_i - \mu)/ \sigma$ (5),

where M_i is the sample mean, μ is the population mean and σ is the population standard deviation. The resulting value tells the magnitude of the difference between a sample and a population in standard deviation terms. Thus, a z value of 1.0 indicates a difference of one standard deviation; a value of .50 indicates a difference of one-half of a standard deviation, etc.

The fourth column of Table 2 reports z-scores comparing the results for the school to those for the state as a whole for each year. Because we wish to compare the values for the school (a sample) to the state as a whole (the population), we need to compute the standard deviation for the state. For proportions (or percentages) this can be done with the binomial distribution. Translating the percentages to proportions (by simply dividing by 100),

$\sigma = \sqrt{}$, (6),

where p_u = the proportion in the population, and $q_u = (1- p_u)$.

As an example, consider the computations for 2003. For this year, pu = .739, qu = .261.

Thus σ = sq root [.739*.261] = sqrt [.193] = .439 (7).

Substituting in the formula (5) for standard (z) scores, where .745 is the sample value M (or p_s) (for the district) and .739 is μ (or p_u),the value for the population,

z = (.745 - .739)/.439 = .006/.44 = .014. (8)

This value (rounded to two significant digits) is given in the first line of Table 2 of the text. It indicates that in 2003 the percentage of third graders in the school who scored at the proficient level or above on the reading test was .014 of a standard deviation greater than in the state.

Similar calculations were completed for 2007 and for scores on the writing test and are reported in Table 2. They indicate that by 2007 the percentage of the school's third graders scoring at or above the proficient level in reading was .15 of a standard deviation greater than in the state as a whole. For writing, there were no differences between the school and the state in 2003, but by 2007 the percentage for the school was .29 of a standard deviation higher than in the state.

Effect Sizes for Changes over Time in a District Controlling for Changes in the State

Educational researchers often use Cohen's d, a measure of effect size, to describe the magnitude of an effect. Cohen's d is calculated as

$(M_1 - M_2)/$ s.d. (9),

where M_i = mean of a group and s.d. = the common standard deviation.

The resulting value tells the magnitude of the difference between two groups in standard deviation terms. A value of 1.0 indicates that the means differ by an entire standard deviation; a value of .50 indicates that they differ by one-half of a standard deviation. Note that this interpretation of an effect size is precisely the same as interpretations of z-scores – a difference in standard deviation terms. Thus, one can see z-scores as effect sizes; they are equivalent to Cohen's use of standard deviation units as an effect size for differences between means, but involve comparing a sample mean to a population mean.

While the results given in Table 2 provide a snapshot of achievement in each year relative to the state, the question of greater interest is the extent to which changes occurred over time in the school. To be most accurate we also need to control for changes that occurred within the state as a whole. Specifically, we want to know the extent to which a school's or district's performance changed over time relative to the performance of the state. A simple way to

describe these changes is to compare the z-scores from one year to another. In other words, we can simply calculate the change in the standard deviation scores. Again, we build on the standard formula for Cohen's d, where

$$d = (M_1 - M_2)/ s.d \quad (10).$$

Because the standard deviation for z-scores is, by definition, 1.0,

$$d_z = (Z_1 - Z_2) \quad (11).$$

Thus, the effect size that describes the change in a district relative to changes in the state can be calculated simply by comparing the z-scores using equation (11).

Again the data in Table 2 can illustrate. For reading scores in 2003, the z-score comparing the proportion of third graders scoring at the proficient level or higher in the school with the proportion in the state was .014, indicating that the proportion for the school was .014 standard deviations above the proportion for the state as a whole. In 2007 the z score for this comparison was +.147, indicating that the proportion was .147 of a standard deviation higher than the score for the state.

The difference of these scores can be easily calculated

$$d_z = .147 - (.014) = .134 \quad (12).$$

From 2003 to 2007, the proportion of third graders scoring at the proficient level or higher increased by .13 of a standard deviation relative to changes in the scores of third graders throughout the state. The comparable result for writing scores was

$$d_z = .289 - .004 = .285 \quad (13).$$

One could, of course, simply look at the change in scores over time and compute an effect size with this information (using the formula $d = (M_{t1} - M_{t2})/s.d.$). This is the type of calculation used with the cohort control group design and described above. However, as

described in the text, one potential problem with this design is that it fails to control for possible historical changes. Comparing the changes in the district with those in the state as a whole, as occurs with formula (11), provides a control for such historical changes.

The data in Table A-1 illustrate the importance of including such a comparison. The data report the results of a simulation involving changes in achievement over time. We assume that the proportion of students meeting criteria in an imaginary district changed from .30 in year 1 to .70 in year 10. If only these data were considered, the effect size reflecting the change would be .87 [=(.70 - .30)/.46 = .40/.46]. This indicates a large change, one that would be considered educationally significant.

The data within the body of Table A-3 provide information on different possible changes within the state over the ten year time period. For instance, the first line reports the situation where .50 of the students in the state met criteria in Year 1 and the same proportion met criteria in Year 10. The third and fourth columns report the population standard deviation for years 1 and 10 (using equation (8) above); and the fifth and sixth columns give the z-scores comparing district values with the state values for each year, using the formula in equation (6) above. For Year 1, the district proportion of .3 compared to a state proportion of .5 results in a z-value of -.40; for year 10, the district proportion of .7 compared to a state proportion still at .5 results in a z-value of +.40. In other words, in year 1 the district proportion was .4 of a standard deviation below the state value, but in year 10 the value was .4 of a standard deviation above the state value. Using formula (11) above, the effect size for the district change, while controlling for the state values $d_z = .40 - (-.40) = .80$. This value is quite close to the unadjusted value of .87, which would be expected given the lack of change in the state. (See the last column of Table A-1,

which reports the difference of d_z and the value of d without controlling for changes in the state; that is, the bias from not controlling for changes in the state.)

[Table A-1 About Here]

The second, third and fourth lines of data depict a situation where there were positive changes within the state. The second line shows a change from .40 to .60 meeting criteria, slightly less than in the district; the third line has a change that exactly matches that within the district, from .30 to .70; and the fourth line depicts a situation with a larger change, .20 to .80. In other words, these are situations where the district improved over time, but so did all students within the state. In one situation the change was slightly greater in the district than in the state (line 2); in the next situation (line 3) the change was exactly the same as in the state; and in the third situation (line 4) the change in the district was actually less than in the state. Clearly the effect size of .87 does not accurately portray what really happened in the district relative to what was happening in the state as a whole. However, the effect size calculated with equation (11) (in the next to last column of the table) accurately reflects these changes. The d_z in line 2, with district changes that are slightly greater than in the state is .41. This is still positive, but smaller than with the unadjusted value. The d_z in line three with changes equal to the state is, as one would expect, equal to zero. Finally, the d_z in line four, where the state had greater positive changes than the district, is negative. This is appropriate because the changes in the district, although positive, were less than in the state as a whole. Relative to the state as a whole the students in our hypothetical district lost ground over time.

The final three lines in Table A-1 depict a situation where the students in the state became less likely to meet criteria over time. In these situations the effect size that only considers district data (d = .87) underestimates the actual magnitude of change. For instance, the

fifth line of data simulates a change in the state from .60 meeting criteria in Year 1 to .40 in year 10. The effect size when these state level changes are considered is 1.22, substantially larger than the value of .87 calculated without this control.

Similar results occur with other patterns of change. The essential point is that, if one wants to examine the amount of change relative to some type of larger comparison group, data for the comparison group need to be considered. Examining changes in the z-scores relative to this larger population provides a simple, accurate way to describe these changes.

Inferential Tests to Examine Changes over Time in a District Controlling for Changes in the State

While the effect size computed above (d_z) provides a descriptive measure of the magnitude of change, practitioners, policy makers, and researchers are often interested in whether or not the changes might have occurred by chance. To answer this question researchers use simple hypothesis tests. For the data presented in this paper we can use t-tests. As mentioned above, the data used in the cohort control group designs, comparing cohorts of students from one year to another, involve samples that may be seen as independent. The vast majority of children in a grade in one year would not be in that grade in a subsequent year.

The null hypothesis tested is that there was no change in a district's results, relative to the state, over time. In other words, the null hypothesis is that the z-scores in year a equal the z-scores in year b. If a curriculum had no effect we would expect that the z-scores relative to the state would be the same in both years. This is our null hypothesis: the difference between the two z-scores is zero.

$$H_o: z_b - z_a = 0 \qquad (14).$$

Alternatively, if there were a positive effect, we would expect there to be fewer students at risk or with a deficit and more students at low risk in the later years. (This is our alternative hypothesis.)

$H_1: Z_b - Z_a > 0$ (15)

To test this hypothesis we can do a simple comparison of means test using the t-distribution, treating the z-scores for each year as the means and using the standard formula for a t-test

$t = (M_2 - M_1)/s.e._{2-1}$ (16),

The standard error is simply a function of the standard deviation and the sample size.

s.e. = $\sqrt{\qquad\qquad\qquad\qquad}$. (17).

By definition, the z-scores have standard deviations of 1.0.

Thus, s.e. = $\sqrt{\qquad\qquad\qquad}$. (18)

To illustrate these calculations, consider the data for 2003 and 2007 from the Alaskan school reported in Table 2 in the text. For 2003, n = 41 and the z-score comparing the value to the state = .014; for 2007, n = 63, and the z-score comparing the value to the state = +.147.

s.e. = $\sqrt{\qquad\qquad\qquad}$ = .184. (19)

$t = (Z_2 - Z_1)/s.e. = (.147 - .014)/.184 = 0.726.$ (20)

The degrees of freedom associated with this test are

$df = n_1 + n_2 - 2 = 41 + 63 - 2 = 102.$ (21)

Using a standard t-table (see an on-line calculator at http://faculty.vassar.edu/lowry/tabs.html#t), it can be found that the probability of getting a t-value of .726 by chance with samples of this size is .2348 (one-tail).

The corresponding values for writing are

$t = (Z_2 - Z_1)/s.e. = d/s.e = .285/.184 = 1.55$ (22)

and the associated probability is .0621.

These values can also be used to place confidence intervals around the effect size. Because schools would logically be interested only in interventions that enhanced student achievement, one tail hypothesis tests, as well as unidirectional confidence intervals, would be most appropriate. For such a confidence interval one would assume that the mean of the sampling distribution equals a sample value (the effect sizes calculated above) and then calculate the value above which 95 percent of the sample values would fall. In other words, based on information from this analysis, what is the minimal effect size that we could expect with 95 percent confidence. In a typical normal curve this would correspond to values at 1.645 standard errors below the mean. This corresponds to the following formula:

$P[D > (d - (1.645)(s.e.))] = .95$ (23),

where D is the estimate of the effect size within the population.

For the Alaska site, the application of these formulas leads to

$P [D > (.134 - (1.645)(.184))] = P[D > (.134 - .303)] = P [D > -.169] = .95$ for reading and

$P [D > (.254 - (1.645)(.184))] = P[D > (.285 - .303)] = P [D > -.018] = .95$ for writing.

Other values could, of course be used for more narrow or wide confidence intervals.

Table A-1

Example of Calculating Cohen's Effect Size, d, with and without Controlling for Changes in the State

State Year 1	State Year 10	s.d. yr. 1	s.d. yr. 10	z year 1	z year 10	d_z	Bias
0.5	0.5	0.50	0.50	-0.40	0.40	0.80	0.07
0.4	0.6	0.49	0.49	-0.20	0.20	0.41	0.46
0.3	0.7	0.46	0.46	0.00	0.00	0.00	0.87
0.2	0.8	0.40	0.40	0.25	-0.25	-0.50	1.37
0.6	0.4	0.49	0.49	-0.61	0.61	1.22	-0.35
0.7	0.3	0.46	0.46	-0.87	0.87	1.75	-0.88
0.8	0.2	0.40	0.40	-1.25	1.25	2.50	-1.63

Note: It is assumed that the values for the district were p = .30 for year 1 and p = .70 for year 10. Thus the effect size, d, without controlling for any changes within the state, would be .87.

Appendix B

Statistical Analyses of Case Studies

This appendix summarizes the results of the separate analyses used in the third section of the paper, which illustrates the grounded theory of generalized causal inference. The first section gives the results for analyses that used the cohort control group (CCG) design, and the second gives results for analyses that used the cohort control group with historical comparisons (CCGH) design.

Analyses Using the Cohort Control Group Design

Table 3 in the text described the sites included in the SRA report. Data for seven sites were deemed appropriate for analysis with the CCG design. They represented all parts of the country, schools with varying grade levels and focused on students at different grades. Table B-1 summarizes the socio-demographic characteristics of the schools, and this also indicates substantial variation. While five of the seven sites had a majority of Caucasian students, Hispanic students comprised 90 percent of one school and Native American students comprised 100 percent of another. Data from these sites were available for 21 separate analyses.

Table B-2 summarizes the information obtained in the individual analyses. (Details on each separate analysis are available on request from the author.) The first panel of the table gives statistics for each of the 21 analyses, and the bottom panel aggregates the information across each of the seven sites. As would be expected, the range of effect sizes was larger with the group of individual comparisons (0.08 to 1.66) than with the values averaged to the site level (.23 to 1.14). The average effect size was, however, slightly larger with the analysis with sites as the unit of analysis (.66 versus .71). On average, the minimum value for the 95 percent

(unidimensional) confidence interval was .29 for the set of individual comparisons and .39 for the site level analysis.

Cohort Control Group Design with Historical Comparisons (CCG-H)

Table 3 in the text also summarized characteristics of the sites that had data appropriate for analysis with the cohort control group with historical comparisons (CCG-H) design. There were 13 sites and 40 separate comparisons. As with the CCG analyses, these sites represented schools and districts from around the country and schools with varying grades. Most of the analyses involved students in grades three and four, with a few also including students in grades 5 and 6. This no doubt reflects the grades at which state assessments are first routinely given. Table B-3 summarizes the socio-demographic characteristics of the students. There was substantial variation across the sites in racial-ethnic composition, in the percentage of students eligible for free or reduced lunch, and in the percentage of ELL students.

Table B-4 summarizes the results. The first panel of the table gives statistics for each of the 40 separate analyses, and the bottom panel aggregates the information across the thirteen sites. As would be expected, the range of effect sizes is larger with the group of individual comparisons (-.26 to 1.04) than with the values averaged to the site level (.03 to 1.04). As with the analyses with the CCG designs, the average effect size is slightly larger with the analysis with sites as the unit of analysis (.43 versus .36). On average, the minimum value for the 95 percent (unidimensional) confidence interval was .07 for the set of individual comparisons and .13 for the site level analysis.

Table B-1

Socio-Demographic Characteristics of Schools in CCG Comparisons

Report #	Free or Reduced Lunch	African American	Caucasian	Hispanic	Asian-American	Native Amer./Al.Native	ELL
1	64	0	75	0	0	25	0
3	95	4	3	90	1	2	52
8	47	1	97	1	0	1	1
14	42	1	97	1	1	0	0
15	100	0	0	0	0	100	0
21	47	2	91	6	1	0	5
23	30	4	89	2	5	0	0

Note: All numbers are percentage of students listed in the report as having a given characteristic.

Table B-2

Summary of Results with Analyses Using Cohort Control Group Designs, 7 Sites and 21 Comparisons

Site #	Grade	Year began RM	Years DI at School	effect size	z	prob.	Min. for C.I.
1	K	K	4	0.62	2.77	0.006	0.25
1	K	K	4	0.48	2.17	0.002	0.12
1	K	K	4	0.87	4.15	<.0001	0.53
3	3	1	3	0.85	7.64	<.0001	0.67
8	2	K	3	0.58	3.8	0.0001	0.33
8	3	1	3	0.08	0.5	0.31	-0.18
8	4	2	3	0.08	0.49	0.31	-0.19
8	5	3	3	0.23	1.49	0.07	-0.02
8	6	4	3	0.18	1.15	0.13	-0.08
14	3	K	5	0.71	2.94	0.0023	0.31
14	3	K	5	0.81	3.91	<.0001	0.47
14	3	K	5	1.66	7.24	<.0001	1.28
14	3	K	5	0.66	4.23	<.0001	0.40
14	3	K	5	0.71	3.19	0.001	0.34
15	K	K	2	1.22	3.21	0.002	0.59
15	1	K	2	1.43	3.86	0.003	0.82
15	2	1	2	0.30	0.53	0.3	-0.63
15	3	2	2	0.64	1.52	0.07	-0.05

21	3 - 8 – SPED		2	1.14	20.83	<.0001	1.05
23	3	K	7	0.16	1.04	0.15	-0.09
23	3	K	6	0.35	2.40	0.01	0.11
Min.				0.08			-0.63
Max.				1.66			1.28
Aver.				0.66			0.29

Site-Level Summaries

Site #	Average effect size	Aver. Min. for C.I.
1	0.66	0.30
3	0.85	0.67
8	0.23	-0.03
14	0.91	0.56
15	0.90	0.18
21	1.14	1.05
23	0.25	0.01
Min.	0.23	-0.03
Max.	1.14	1.05
Aver.	0.71	0.39

Note: The three analyses for site 1 are those reported in the text regarding the three different DIBELS measures. The five analyses for site 14 reflect analyses for different schools. Other cases with multiple analyses for a site reflect the different grades examined as listed in the second column of the table.

Table B-3

Socio-Demographic Characteristics of Schools in CCG-H Comparisons

Report #	Free or Reduced Lunch	African American	Caucasian	Hispanic	Asian-American	Native Amer./Al.Native	ELL
1	64	0	75	0	0	25	0
2	72	55	42	3	0	0	0
5	92	4	47	49	0	0	44
6	85	92	2	6	0	0	0
7	76	26	32	34	7	1	32
10	21	3	73	17	1	6	4
11	91	94	1	3	1	1	3
13	92	61	38	1	0	0	0
17	51	23	70	6	0	0	0
18	100	38	47	12	0	0	0
19	72	0	49	50	0	0	58
23	30	4	89	2	5	0	0
24	44	8	73	15	3	1	0

min	21	0	1	0	0	0	0
max	100	94	89	50	7	25	58
mean	68.5	31.4	49.1	15.2	1.3	2.6	10.8

Note: All numbers are percentage of students listed in the report as having a given characteristic.

Table B-4

Summary of Results with Analyses Using Cohort Control Group with Historical Comparisons Designs, 13 Sites and 40 Comparisons

Site #	Grade and Comparison	Year began RM	Years DI at School	effect size	z	prob.	Min. for C.I.
1	3 - reading	K	4	0.13	0.73	0.24	-0.17
1	3 - writing	K	4	0.28	1.55	0.06	-0.02
2	3	2	2	0.11	0.65	0.26	-0.17
2 (a)	4	3	2	-0.26	-1.35	0.91	-0.58
2	5	4	2	0.21	1.26	0.10	-0.06
2	6	5	2	0.09	0.52	0.30	-0.2
5	3 - proficient +	2	2	0.63	2.08	0.02	0.13
5	4 - proficient +	3	2	0	-0.02	0.51	-0.5
5	5 - proficient +	4	2	0.35	1.17	0.13	-0.14
5(b)	3 - unsatisfactory	2	2	1.04	3.45	0.001	0.55
5(b)	4 - unsatisfactory	3	2	0.59	1.94	0.03	0.09
5(b)	5 - unsatisfactory	4	2	0.27	0.9	0.19	-0.23
6	3	K	7	0.79	2.36	0.01	0.24
7	3 to 5, AA to AA	k	4	0.29	1.99	0.02	0.05
7	3 to 5, Hisp. to Hisp.	k	4	0.2	1.55	0.06	-0.01
7	3 to 5, Low Inc. to LI	k	4	0.18	1.99	0.02	0.03
7	3 to 5, Excp. to Excp.	k	4	0.27	1.46	0.07	-0.03
7	3 to 5, Cauc. to Cauc.	k	4	0.23	1.74	0.04	0.01
7	3 to 5, total to total	k	4	0.09	1.25	0.11	-0.03
7	3 to 5, AA to Total	k	4	0.34	2.33	0.01	0.1
7	3 to 5, Hisp. to Total	k	4	0.28	2.2	0.01	0.07
7	3 to 5, LI to Total	k	4	0.22	2.36	0.01	0.07
7	3 to 5, Excp. to Total	k	4	0.23	1.25	0.11	-0.07

7	3 to 5, Cauc. to Total	k	4	0.18	1.39	0.08	-0.03
10	4	K	9	0.19	2.18	0.01	0.05
11	4	K	7+	0.22	1.37	0.09	-0.04
13	4	3	2	1.04	4.40	<.0001	0.65
17	3 - sch. A	1	3	0.53	5.57	<.0001	0.37
17	4 - sch. A	2	3	0.35	3.71	0.0001	0.19
17	5 - sch. A	3	3	0.22	2.36	0.01	0.07
17	3 - sch. B	1	3	0.36	2.06	0.02	0.07
17	4 - sch. B	2	3	0.33	1.86	0.03	0.04
17	5 - sch. B	3	3	0.01	0.05	0.67	-0.28
18 (a)	4	k	8	0.63	3.46	0.0005	0.33
18	4 (alternate years)	k	7	0.94	5.17	<.0001	0.64
19	3 - Total to total	1	3	0.65	5.72	<.0001	0.47
19	3 - Hisp. To Hisp.	1	3	0.80	4.96	<.0001	0.53
19	3 - Hisp. To Total	1	3	0.81	5.03	<.0001	0.55
23	4	K	7	0.03	0.20	0.42	-0.12
24	4	K	7	0.56	2.87	0.0025	0.24
Min.				-0.26			-0.58
Max.				1.04			0.65
Aver.				0.36			0.07

Site-Level Summaries

Site #	Average effect size	Aver. Min. for C.I.
1	0.21	-0.09
2	0.04	-0.25
5	0.48	-0.02
6	0.79	0.24
7	0.23	0.01
10	0.19	0.05
11	0.22	-0.04
13	1.04	0.65
17	0.30	0.08
18	0.79	0.49
19	0.75	0.52
23	0.03	-0.12
24	0.56	0.24
Min.	0.03	-0.25
Max.	1.04	0.65
Aver.	0.43	0.13

Note: Comparisons designated with "a" in the site column are those for which the SRA report indicated a non-comparable cohort. Statistics associated with comparisons designated with a b were reversed to preserve the positive sign associated with findings in favor of the DI programs.

Endnotes

[1] Concerns have also been voiced regarding the use of randomized control trials in medicine (e.g. Clay, 2010; Williams, 2010) and criminology (Sampson, 2010).

[2] While the psychological (and education) literature seems to generally use the notion of "reactivity," building on the CCSS tradition, the economics and sociological literature appears to also discuss violations of the "stable unit treatment value assumption" (SUTVA), using the logic of counterfactuals and the tradition of causal analysis developed by Donald Rubin. (See Cook & Steiner, 2010; Maxwell, 2010; Rubin, 2010; Shadish, 2010; and West & Thoemmes, 2010, for discussions of the differences of the two approaches.). Writers using these terms note that the SUTVA assumption might be met with drug trials where the impact of a dose on one patient is independent of the dose that another patient receives. In schools and classrooms, however, administration of a treatment (such as instruction) can involve different forms of applications for each unit (students) given the impact of the multiple influences (other students and variability in teachers) in their environment (Raudenbush, 2008). Put another way, Sloane (2008b), an educator and using experimental design terms, suggests that the nesting of data within schools results in variability in both fixed effects (differences in means of the dependent variable) and random effects (variability in the dependent variable). The traditional randomized control trial model only considers the former, potentially leading to misleading results and difficulties in generalizing to other contexts. (See also Yin and Davis, 2007.) A large literature, largely within sociology and economics, has addressed methodological issues of causal inference when using observational data and the logic of the counterfactual approach. (See Morgan & Winship, 2007 and Winship & Morgan, 1999 for especially readable discussions of this area.)

[3] These criteria are very similar to those discussed in a recent article by Shadish and Cook (2009).

[4] The call made by some authors for place-based randomization in educational contexts (e.g. Boruch 2005; Cook, 2005; Fayer, 2005) could be an example of the techniques to which Shadish and colleagues were referring.

[5] Miron (2005) briefly described a similar design that he called "successive cohorts," but did not cite the CCSS tradition or use the analysis approach that we describe below.

[6] Three reports examined the growth over one year of small numbers of special education students, rather than achievement of larger cohorts over multiple years. One report reported data for one cohort over only one year, which precluded comparisons between cohorts and years. Two sites implemented the programs with only a sub-set of students, but reported data for the entire school, and thus it was not possible to match the intervention with the available data. Finally, two districts had been using the programs for an extended period of time and did not provide data from before the implementation began.

[7] Sites 8 and 21 also used Corrective Reading, an intervention program for older students. Site 17 used Corrective Reading for older students and Language for Learning with younger students. Site 18 used Language for Learning , Site 24 used Language for Learning and Language for Thinking, Sites 6 and 23 used Spelling programs, and Site 6 also used Reasoning and Writing. Our analysis reported below examined differences in effects between these sites and others and found minimal differences.

[8] The reviews suggesting differences between the techniques examined studies on criminal justice, welfare and job training programs, and dropout preventions. Those suggesting more minimal effects examined issues related to student achievement and learning, arguably the substantive area that would be of more concern to educational researchers.

[9] The What Works Clearinghouse is typical of this approach, often examining dozens of studies of curricula but finding that only a handful, almost always fewer than five, are worthy of review (Slavin, 2008).

From: Jean Stockard <jstockard@nifdi.org>
Sent: Monday, June 29, 2009 6:39 PM
To: What Works
Cc: cristy@nifdi.org
Subject: Re: What Works Clearinghouse (WWCPC-1322)
Attachments: DI Mathematics Articles June 30, 2009.doc

Thank you for your note regarding studies of Direct Instruction mathematics programs. I have attached a bibliography of 16 relevant studies. With the exception of one doctoral dissertation, this list only includes works published in refereed journals. Please let me know if I can provide additional information. I hope this is a sufficient list for review.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction

> Dear Dr. Stockard,
>
>
>
> Thank you for contacting the What Works Clearinghouse (WWC). Your
> submission
> of "Improving Elementary Level Mathematics Achievement in a Large Urban
> District: The Effects of NIFDI-Supported Implementation of Direct
> Instruction
> in the Baltimore City Public School System" was received on October 6,
> 2008.
>
>
>
> Please be advised that the order in which interventions are reviewed by
> the
> WWC is based, in part, on the availability of studies examining the
> effectiveness of an intervention. Thus, one of the reasons this study
> examining the effects of Direct Instruction mathematics has not yet been
> reviewed is that there are other interventions with a greater number of
> studies examining their effectiveness. If there are additional studies
> about
> the effectiveness of Direct Instruction mathematics, please submit them to
> the WWC.
>
>
>
> It is important to note that the WWC website does not list all

> products/interventions that are the subject of research submitted to the
> WWC,
> but rather interventions for which we have completed full literature
> searches
> and reviewed all eligible studies against evidence standards.
>
>
>
> We hope that you will find this information helpful in understanding the
> study submission and WWC review process.
>
>
>
> What Works Clearinghouse
>
>
>
> The What Works Clearinghouse was established by the U.S. Department of
> Education's Institute of Education Sciences to provide educators,
> policymakers, researchers, and the public with a central and trusted
> source
> of scientific evidence of what works in education. For more information,
> please visit http://ies.ed.gov/ncee/wwc/ <http://ies.ed.gov/ncee/wwc/> .
>
>
>
> _____
>
> -----Original Message-----
>
> From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
>
> Sent: Monday, June 22, 2009 3:40 PM
>
> To: info@whatworks.ed.gov
>
> Subject: IES Website: Contact Us: Check on the Status of a Submission,
>
> Reference ID Number: 109747337
>
>
>
> info@whatworks.ed.gov, this email was automatically sent through the
> Contact
>
> link on the WWC website.
>
>

>
> From: jstockard@nifdi.org
>
>
>
> Message: Might you be able to tell me the status of submission # 29487267,
>
> please?
>
> It was a study of Direct Instruction mathematics.
>
> Thank you.
>
> Jean Stockard, Ph.D.
>
>
>
>
>
>

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon
phone: 541-485-1973 or toll-free 877-485-1973

Improving First Grade Reaching Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Jean Stockard
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

September 2008

© 2008 National Institute for Direct Instruction
PO Box 11248 • Eugene, OR 97440
1-877-485-1973

Table of Contents

	Page
Executive Summary	iv
List of Tables	v
List of Figures	vi
Report	1
Methodology	1
Results	3
1. Did students who received Direct Instruction have higher achievement scores than students in the other schools?	3
2. Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in the other schools?	4
3. Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools were equalized?	5
4. Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?	6
5. What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?	10
Summary	12
Appendix I: Sample and Measures	14
Sample	14
Measuring Achievement	15
Measuring School Context	15
Appendix II: Detailed Statistical Results	19
Methodology	19

Research Question 1: Did students who received Direct Instruction have higher achievement scores than students in the other schools? 20

Research Question 2:"Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in the other groups?" 22

Research Question 3: Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized? 23

Research Question 4: Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school? 30

Research Question 5: What is the Magnitude of the Effect of NIFDI-supported Direct Instruction? 35

References 38

Improving First Grade Reaching Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Executive Summary

In the mid 1990s, in response to very low reading achievement scores, the Baltimore City Public School System (BCPSS) implemented curricular reforms. Sixteen schools used Direct Instruction. All the Direct Instruction schools used the *Reading Mastery Classic, Language for Learning, Language for Thinking,* and *Reasoning and Writing* curricula. Eleven of these schools received technical support from the National Institute for Direct Instruction throughout the time period of this study, while the others ceased support or used an alternative provider. The Comprehensive Test of Basic Skills (CTBS) was administered to all first graders in the spring of 6 school years: 1997-1998 through 2002-2003. Data were obtained from over 40,000 students on reading vocabulary, comprehension, and a composite score.

The study compared students' achievement in 1) schools with NIFDI support, 2) schools that implemented Direct Instruction without NIFDI support (other DI schools), and 3) schools without Direct Instruction (the Control schools). Schools in the three groups had similar demographic characteristics, although one of the Other DI schools had a markedly lower rate of poverty. Key findings are listed below.

- Combining data across all years (1998-2003) students in schools that implemented Direct Instruction had significantly higher achievement than students in other schools.
- Combining data across all years (1998-2003) students in NIFDI-supported schools had significantly higher achievement than students in the Control schools on all three measures of achievement and significantly higher comprehension scores than students in the Other DI schools.
- When the socio-economic characteristics of the schools were controlled, the students in the NIFDI-supported schools had significantly higher achievement scores than students in both the Other DI schools and the control schools on all measures of achievement.
- First grade students in all 3 groups of schools had higher achievement scores in 2003 than in 1998, but the increase was significantly larger for students in NIFDI-supported schools than for students in the other schools. On average, first grade composite reading achievement scores in the NIFDI-supported schools increased by 113 percent from 1998 to 2003, while those in the control schools and the Other DI schools increased by 56 percent or less.
- The magnitude of the effect on first grade achievement from attending a NIFDI-supported school was statistically significant and substantively large. An effect size of .25 has traditionally been considered educationally important. At the end of the study period (2003) the effect of attending a NIFDI-supported school versus attending an Other DI school on composite achievement was .63. The effect of attending a NIFDI-supported school versus a Control school was .82.

List of Tables

Title	Page
A-1. Schools in Treatment Groups and Start Dates of Treatment	14
A-2. Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels, Total and by Treatment Group	16
A-3. Factor Analysis of School Characteristics, Elementary Schools, BCPSS, 1998-2003	17
A-4. Average Reading Achievement Scores of Students with and without Direct Instruction, All Years, First Graders, BCPSS	21
A-5. Average Reading Achievement Scores by Group, All Years, First Graders	22
A-6. Average Reading Achievement Scores by Group, All Years, First Graders, BCPSS, Four Groups	23
A-7: Variables in Models and Measures of Fit, Analysis of First Grade Reading Achievement, BCPSS, 1998-2003	25
A-8: Coefficients Associated with Mixed Model Regressions of First Grade Reading Achievement Scores on School and Treatment Factors, BCPSS, 1998-2003	29
A-9: Predicted Achievement Scores Assuming Equal School Contexts	30
A-10 First Grade Reading Achievement Scores by Group and Years of Implementation, Adjusting for School Advantage, BCPSS	32
A-11: Reading Achievement Scores by Year and Group, First Graders, BCPSS, 1998-2003	33
A-12: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Reading Achievement, BCPSS, 1998-2003	36

List of Figures

Page

1. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average First Grade Student, DI and Control Schools, BCPSS, All Years 3

2. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average First Grade Student, NIFDI-Supported, Other DI, and Control Schools, BCPSS, All Years 4

3. Reading Percentile Achievement Scores of an Average First Grade Student, Adjusting for School Context, NIFDI-Supported, Other DI, and Control Schools, BCPSS, All Years 5

4. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average First Grade Student, BCPSS, 1998-2003 6

5. Composite Reading Achievement Percentile Score of an Average First Grade Student, NIFDI-Supported, Other DI, and Control Schools, BCPSS, First Year and Seventh Year of Implementation, Adjusting for School SES 7

6. Reading Comprehension Percentile Scores of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 8

7. Vocabulary Percentile Scores of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 8

8. Vocabulary Percentile Scores of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 9

9. Percentage Change in Achievement of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 10

10. Effect Sizes, Comparisons of Composite Reading Achievement Scores of First Grade Students in NIFDI-Supported, Other DI, and Control Schools, BCPSS, All Years and 2003 11

11. Effect Sizes, Comparisons of Composite Reading Achievement Scores of First Grade Students in NIFDI-Supported, Other DI, and Control Schools, Adjusted for School SES, All Years and 2003 12

Improving First Grade Reaching Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in Baltimore City Public Schools

Low reading achievement of elementary school students has been a major concern of school districts throughout the country. The Baltimore City Public School System (BCPSS) is similar to many other large city school districts that serve students with high levels of poverty and struggle with low levels of achievement. In the late 1990s, curricular reforms were implemented in the BCPSS elementary schools to address this low achievement.

Sixteen of the schools used Direct Instruction for reading instruction, primarily the *Reading Mastery Classic, Language for Learning, Language for Thinking,* and *Reasoning and Writing* curricula. Other schools in the district adopted the Open Court Reading curriculum. In addition, while all of the schools using Direct Instruction started implementation with the National Institute for Direct Instruction (NIFDI), five of these schools either ceased involvement with NIFDI soon after implementation or began to work with another provider. These variations provide the basis for what could be seen as a natural experiment, providing the opportunity to compare student achievement with different types of curricula and different sources of implementation support.[1]

A very large body of literature, including well designed meta-analyses, has demonstrated the effectiveness of Direct Instruction in promoting high reading achievement. A slightly smaller body of literature has highlighted the importance of consistent implementation of the program in achieving the best results. Using data on the achievement of first graders in the BCPSS from 1998 through 2003, this report compares achievement scores of students in schools with NIFDI support with those in other DI schools and in the rest of the Baltimore system. This report examines the extent to which using the Direct Instruction curriculum and consistent technical support in implementation of this curriculum can promote higher levels of achievement. Additional analyses further articulate the extent to which differences in achievement remain when changes over time and the socio-economic and demographic characteristics of the schools are controlled. These analyses disclose the extent to which the Direct Instruction curriculum, with and without NIFDI technical support, can help counter the detrimental impact of a context of poverty on student achievement. In this report results are summarized in graphs and are also expressed as effect sizes. Two appendices provide full statistical details. The methodology is briefly described below, followed by a summary of the results obtained with each research question and a discussion of the implications.

Methodology

The analysis uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Measures of reading achievement were obtained each spring using the Comprehensive Test of Basic Skills (CTBS), a widely used, nationally normed standardized test. Subtest measures of Vocabulary and Comprehension, as well as a Composite measure were available. Results in this part of the report are presented as percentiles. Percentiles can be simply interpreted as the percentage of

[1] MacIver and Kemper (2002) examined data for a shorter time span from 6 of the BCPSS schools included in this study, but did not differentiate between the NIFDI-supported and other DI schools.

students that would have scores lower than a given student. For instance, if a student has a score at the 60th percentile, 60 percent of all students had scores that were lower.

The average school in the BCPSS had large proportions of poor and minority students. On average, schools were quite poor, with 75 percent of the student bodies on free or reduced lunch. This varied however, from a low of 22 percent to a high of 93 percent. The district also had a large proportion of minority students. On average, the student body of elementary schools in the district was 84 percent African American, although the percentage varied from a minimum of three percent to almost 100 percent. Reflecting a high level of segregation, schools varied in their representation of non-Hispanic white students, from having virtually no white students to being 94 percent white, with an average value of 14 percent. Asian, Hispanic and Native American students were quite rare in the district, with an average representation of one percent in the student bodies. There was also substantial variation in the poverty status of the schools.

The NIFDI-supported schools were similar to the Other DI schools and the other BCPSS schools in both racial-ethnic composition and rates of poverty. There were no significant differences between the three groups in the average proportion of students receiving free and reduced lunch or the representation of African American or non-Hispanic white students. The NIFDI schools, however, had significantly higher proportions of Hispanic and Native American children.

One of the schools in the Other DI group had a much lower rate of poverty. On average, over the years in the study, only about one-third of the students in this school received free or reduced lunch. In all of the other schools that received Direct Instruction, seventy-five to ninety percent of the students had free or reduced lunch. Appendix I provides additional details on the sample and measures used in the analysis.

The analysis focused on 1) comparing the achievement of students who received Direct Instruction with those who did not and 2) comparing the achievement of students in the NIFDI-supported schools with those in the two other groups. For the three way comparisons, schools that implemented Direct Instruction, but without NIFDI support throughout the entire study period, are termed "Other DI schools." The remaining schools in the system are termed "Control schools."

Five research questions were addressed:

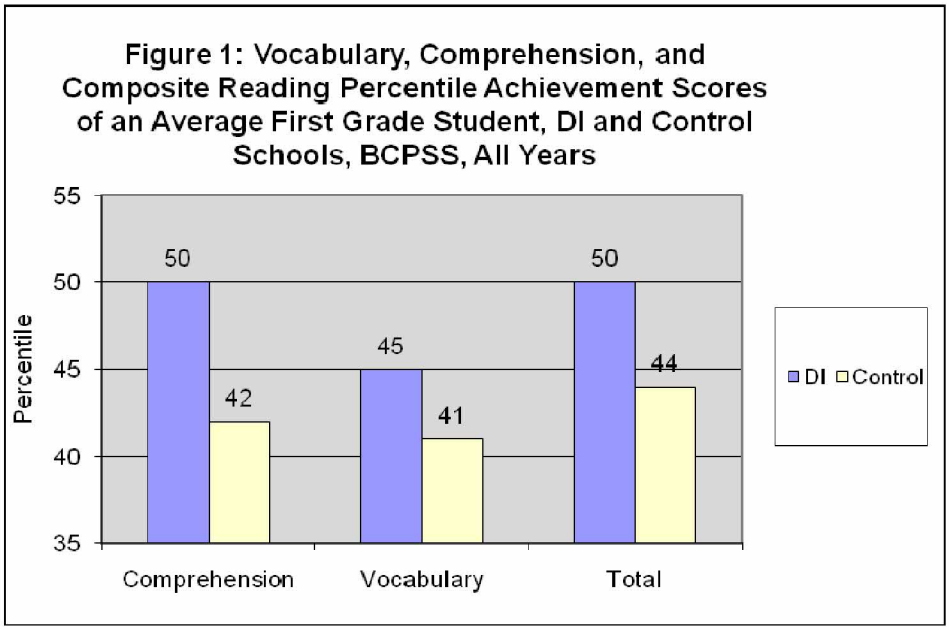
- Did students who received Direct Instruction have higher achievement scores than students in the other schools?
- Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in other groups?
- Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools were equalized?
- Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?
- What is the magnitude of the effect of NIFDI-supported Direct Instruction?

Results

A general summary of the results related to each of the research questions is given below. Full statistical details are provided in Appendix II.

1. Did students who received Direct Instruction have higher achievement scores than students in the other schools?

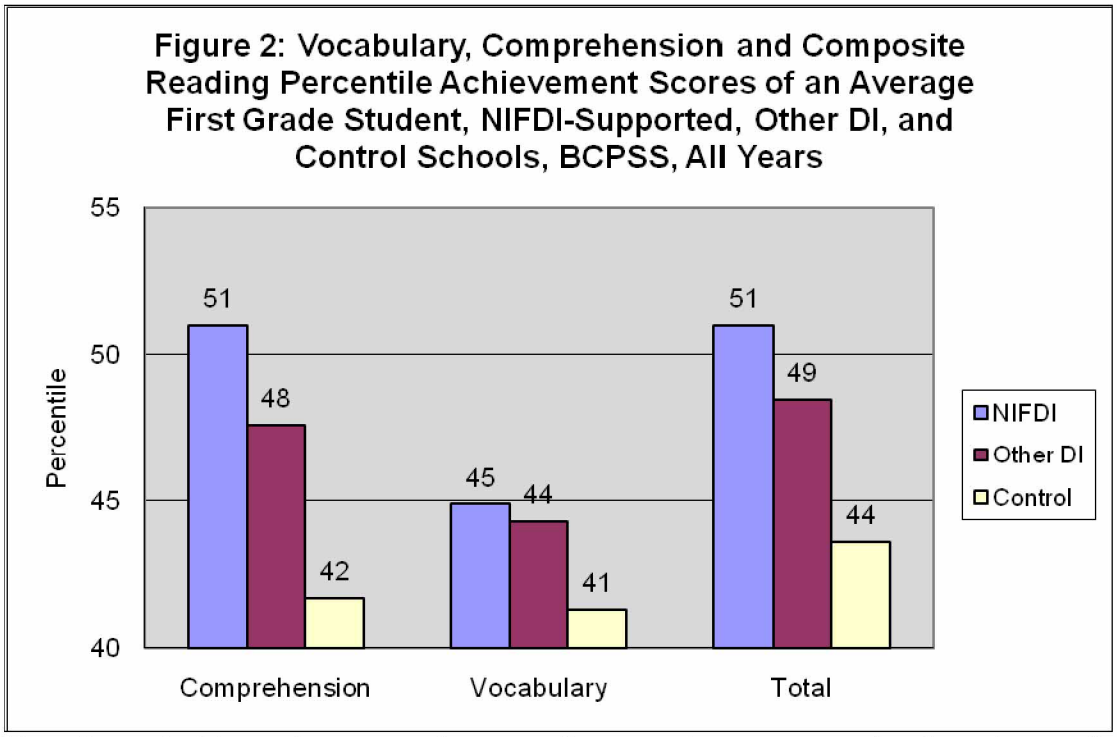
Figure 1 summarizes achievement data for students who received Direct Instruction and those who did not receive DI on each of three measures of reading achievement. These results combine the data from all the available years (1997-98 through 2002-03) and represent the percentile score of the average student in a school. Combining data from all the years in the study, the average first grader that received Direct Instruction scored at the 50th percentile, the national norm, in both the measure of comprehension and the composite score, but somewhat below the national average, at the 45th percentile, on vocabulary. The average student in the Control schools had scores that were substantially lower: at the 42nd percentile for comprehension, the 41st percentile for vocabulary, and the 44th percentile for the composite measure. These differences were statistically significant at well beyond the .001 level of significance.



Note: Scores in this graph represent the percentile score of an average student in each type of school scored, averaging data over all years in the analysis: 1998-2003.

2. Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in the other groups?

As illustrated in Figure 2, students in the NIFDI-supported schools had higher average levels of achievement than students in both the Other DI schools and the Control schools. Combining data over all years in the study, the average student in a NIFDI-supported school scored at the 51st percentile on the comprehension and composite measures and at the 45th percentile on vocabulary. The average student in the Other DI schools scored at the 48th percentile on comprehension, at the 44th percentile on vocabulary, and the 49th percentile on the composite measure. Scores were lowest for students in the Control schools. The average student in a Control school had comprehension at the 42nd percentile, vocabulary scores at the 41st percentile and composite scores at the 44th percentile. Differences between the scores in NIFDI-supported schools and those in Control schools were statistically significant on all three measures. However, differences between the scores for students in NIFDI-supported schools and those in the Other DI schools were statistically significant only for the measure of comprehension.

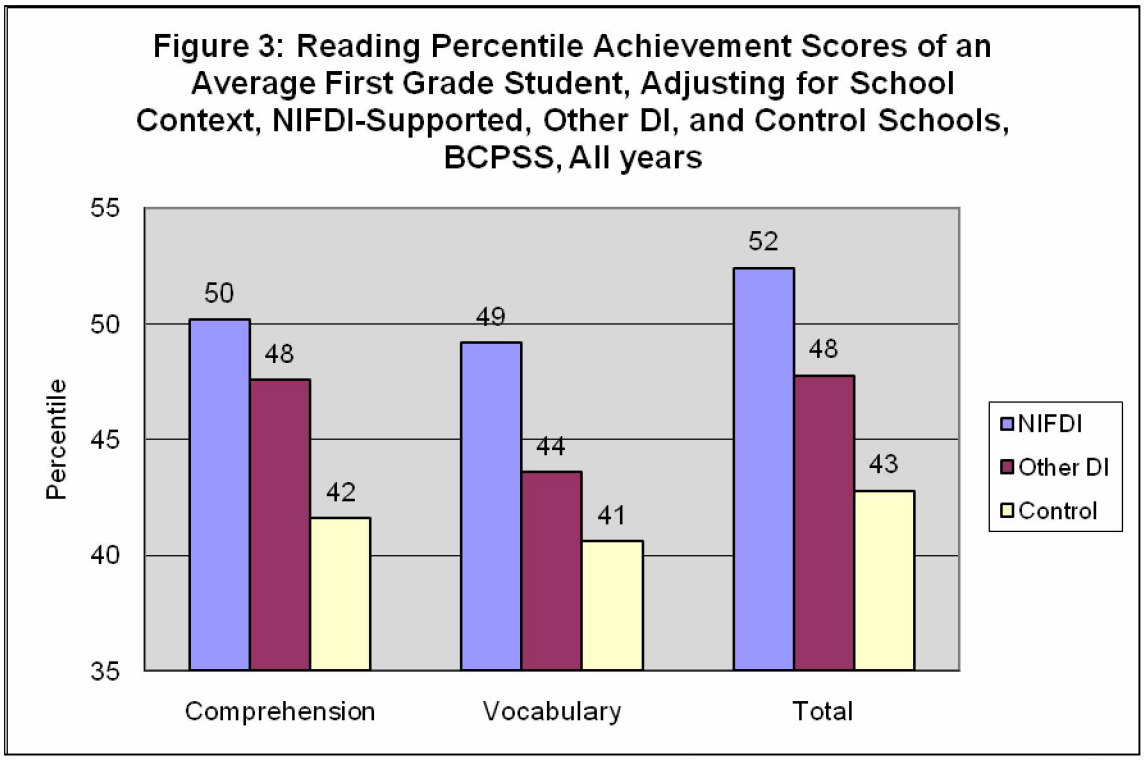


Note: Scores in this graph represent the percentile score of an average student in each type of school averaging over all the years in the analysis: 1998-2003.

3. Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools were equalized?

Many studies have demonstrated that the socio-economic characteristics of a school have a strong influence on achievement, and thus it is important to control for this factor.[2] As noted above, elementary schools in the Baltimore City Public School System varied substantially in their racial-ethnic composition and poverty rate. As in other studies, students in BCPSS schools with fewer poor children had higher average levels of achievement.

When the socio-economic characteristics of the schools were statistically controlled the differences between NIFDI-supported schools and other schools became even more marked. Figure 3 summarizes these results. The numbers in this graph represent the percentile score for an average student in each group if schools were equal in their racial-ethnic and poverty status. (Details regarding these calculations are in Appendix II.) The major difference between these results and those in Figure 2 is that the differences between scores of students in the NIFDI-supported schools and those in the Other DI schools are now statistically significant on all three measures. The higher levels of achievement for students in the Other DI schools relative to the Control schools shown in Figure 2 reflect the high achievement of students in the low poverty school in the Other DI group. When statistical controls adjust for this disparity – making the three groups of schools more similar – any advantage for students in the Other DI group disappears.

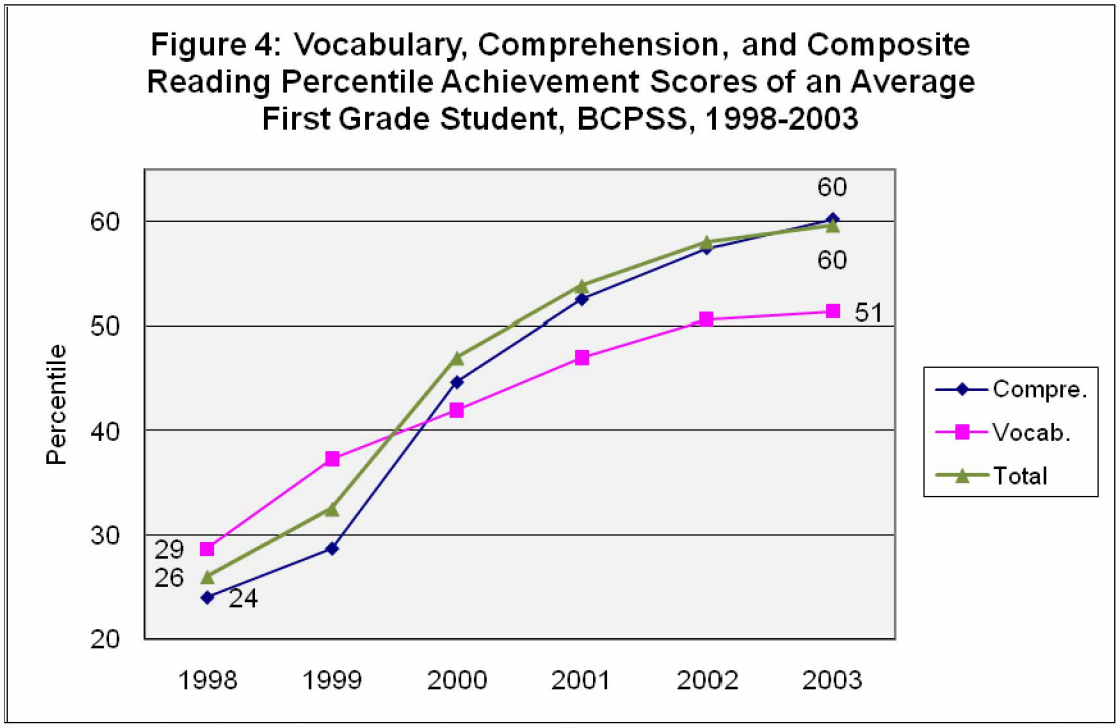


Note: Scores in this graph represent the percentile score of an average student in each type of school if the schools were equal in their socio-economic and demographic characteristics, averaging over all the years in the analysis: 1998-2003.

[2] See Stockard and Mayberry, 1992, for a complete review of this literature.

4. Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?

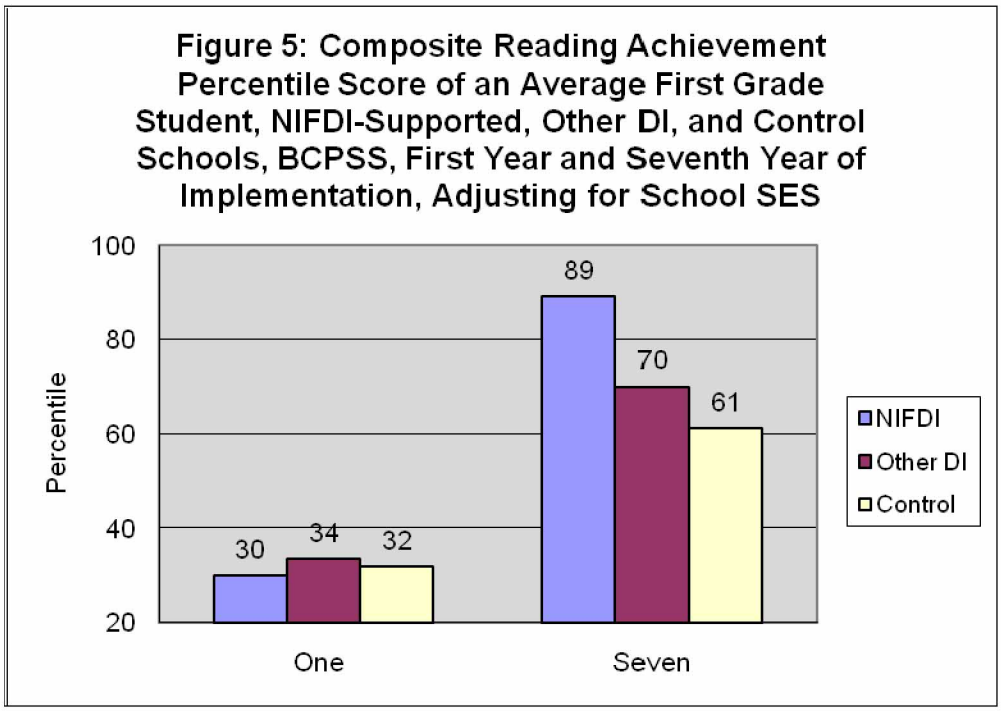
Changing school practices takes time and effort. It is reasonable to expect that the advantage of an effective instructional program would become more apparent as schools have more experience with a new curriculum, teachers gain more practice, and the procedures become part of the institutionalized and accepted practices within a school. As noted above, Direct Instruction was not the only new reading curriculum introduced to the BCPSS during the years included in the analysis. In 1998 the system adopted Open Court Reading for all K-2 schools not in the DI program, and concerted attention was paid throughout the district to enhancing student achievement. The data indicate that these efforts were fruitful. Data in Figure 4 represent the percentile at which an average student achieved at each year for each measure. It can be seen that the reading achievement scores of the average student in the BCPSS increased markedly from 1998 to 2003 on all three of the achievement measures. For instance, in 1998, the average student in BCPSS had a composite achievement score at the 26th percentile. By 2003, the average student scored at the 60th percentile.



Note: Scores in this graph represent the percentile score of an average student in the BCPSS for each year of the analysis and for each measure of achievement.

However, the change in average achievement levels was significantly greater for students in the NIFDI-supported schools than in either the Other DI or Control schools. Figure 5 summarizes the results of this analysis. The bars in the graph give the percentile score on the composite reading measure for an average student in each type of school in the first and last year of implementation of Direct Instruction. All of the calculations equalized school socio-economic status. (Details regarding these calculations are included in Appendix II.) The results show that after the first year of implementation, an average student in a NIFDI-supported school had achievement scores that were equal to or even slightly less than

an average student in the other schools. However, after seven years of implementation, an average student in a NIFDI-supported school had scores that were substantially higher than an average student in other schools.[3] The scores of students in all three groups of schools increased significantly over time, but the increase in scores of students in the NIFDI-supported schools was much greater. In other words, on average all first graders in the BCPSS had significantly higher achievement scores at the end of the study period than at the beginning, but this increase was significantly higher in the schools that were served by NIFDI.



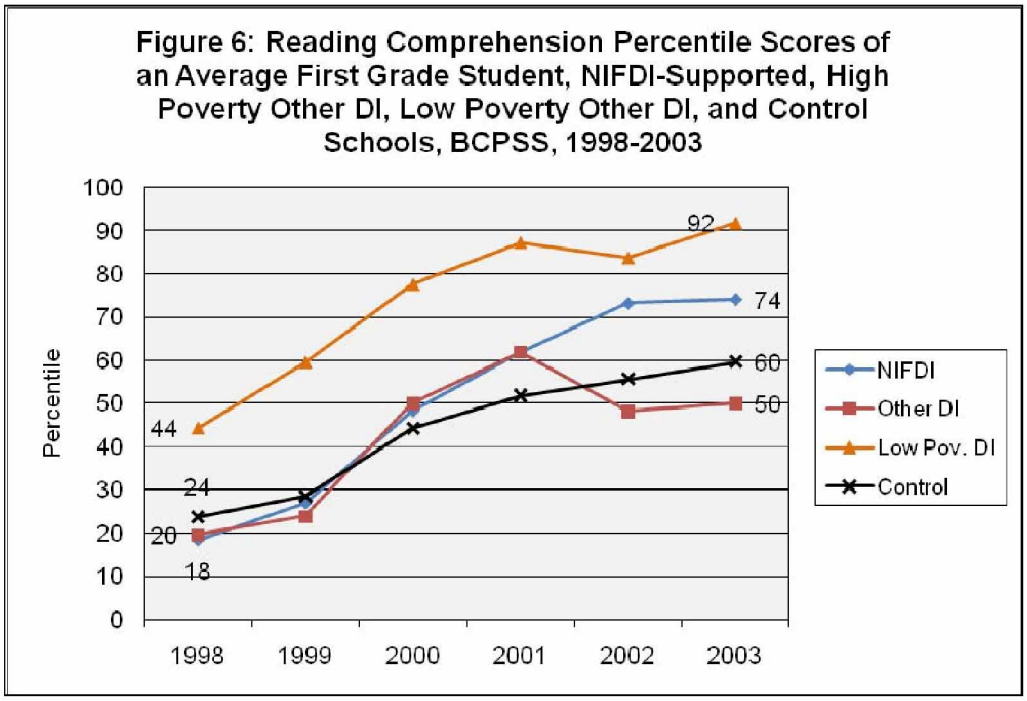
Note: Scores in this graph represent the percentile score on the composite measure of reading achievement for an average first grade student in each type of school in the first year of implementation of DI and the 7th year.

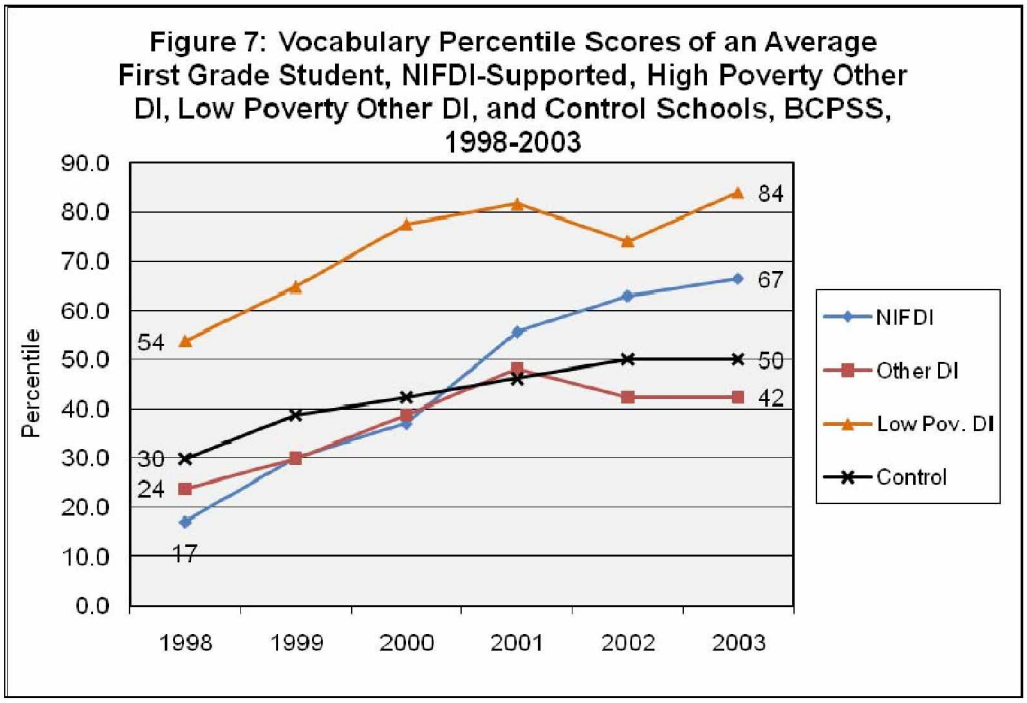
The data in Figure 5 were calculated from the results of statistical analyses that adjusted for variations in school advantage and other differences in the schools. However, the same pattern occurs when the simple raw achievement scores for schools in each group are examined over the time period in the study: 1998-2003. Figures 6 through 8 summarize these data. They show the percentile at which an average student in each group scored in each of the 6 years. In these figures the data for the low-poverty school in the Other DI group is separated from the others.

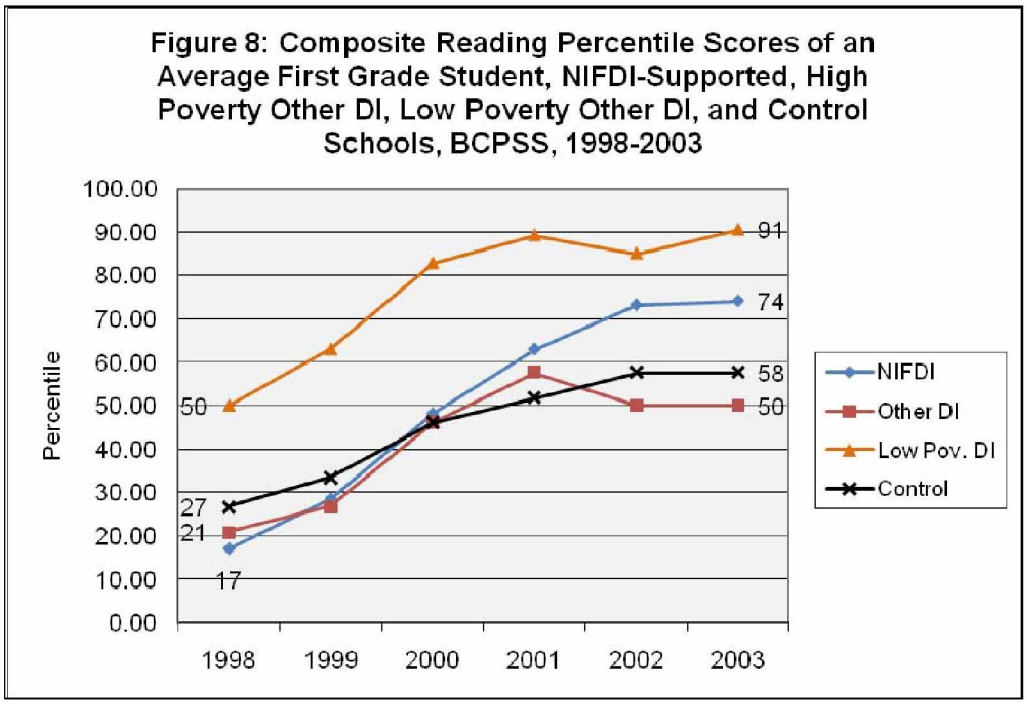
At the beginning of implementation in 1998, an average student in NIFDI-supported schools had lower scores than average students in the other schools, scoring at the 17th percentile in vocabulary and the composite score and at the 18th percentile in comprehension. An average student in the high poverty Other DI schools (those that were demographically similar to the NIFDI-supported schools) had scores only slightly higher than the average student in a NIFDI-supported school, ranging from the 20th to 24th percentile, while an average student in the control schools scored from the 24th to the 30th percentile. In contrast,

[3] All but four of the NIFDI-supported schools and all but one of the Other DI schools began implementation in 1996-97, one year before the start of the data set, thus having 7 years of implementation by the spring of 2003.

in 1998 an average student in the low-poverty school that implemented DI scored from the 44th percentile (on comprehension) to the 54th percentile (on vocabulary). In general, in 1998, an average BCPSS student had very low achievement scores.





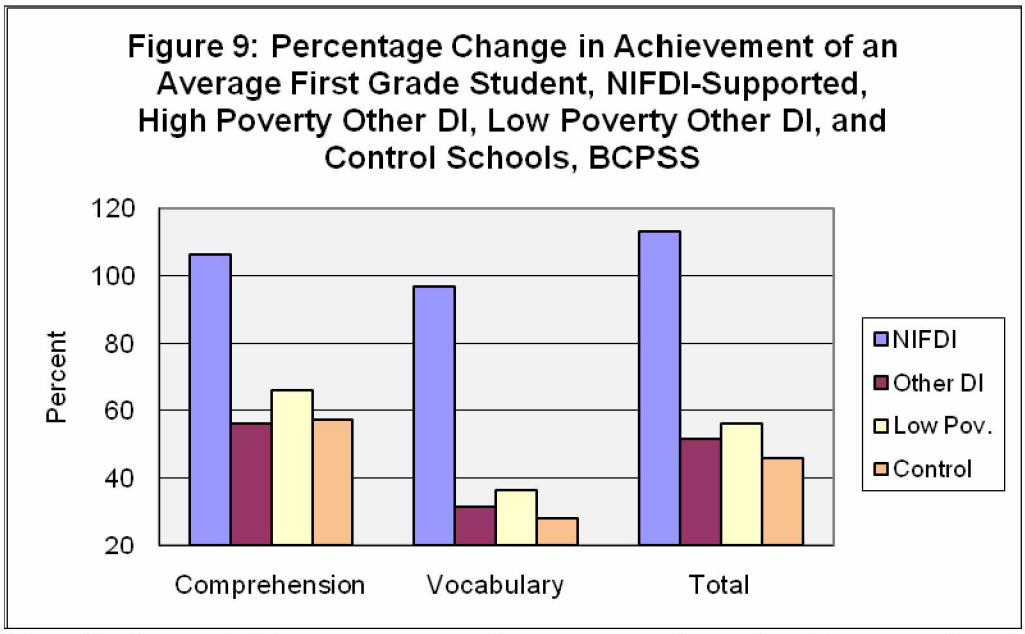


Note: Scores in Figures 6, 7 and 8 represent the percentile score of an average first grade student in each type of school for each year in the analysis. Data for students in the low poverty school in the Other DI group are separated from other schools in that group.

By 2003 the situation had changed markedly. The average student had substantially higher scores than in 1998. For instance, scores for an average student in the Other DI, high poverty schools, ranged from the 42nd (vocabulary) to the 50th percentile (comprehension and composite), while scores for an average student in the Control schools ranged from the 50th (vocabulary) to the 60th percentile (comprehension). An average student in the low-poverty Other DI school had very high scores, ranging from the 84th percentile in vocabulary to the 92nd percentile in comprehension. The changes were even more marked, however, for students in the NIFDI-supported schools. By 2003, the average student in a NIFDI-supported school had percentile scores that ranged from 66 (for vocabulary) to 74 (for comprehension and the composite score). These scores were significantly higher than those of students in both the Other DI schools (excluding the low poverty school) and the control schools.

Figure 9 illustrates the magnitude of these changes. Values in Figure 9 represent the difference in the 1998 and 2003 scores as a percentage of the 1998 score, thus showing the percentage increase over time. The differences between the groups are very large. For instance, for the composite scores, the percentage increase in achievement for first grade students in the NIFDI-supported schools was over twice that for students in the other three groups: 113 percent versus 46 percent (the control schools) to 56 percent (the low poverty school in the Other DI group).[4]

[4] These change figures were calculated using normal curve equivalent (NCE) scores rather than percentiles. As explained more fully in Appendix I, NCE scores are preferable for statistical procedures and were used for all calculations in the analysis.



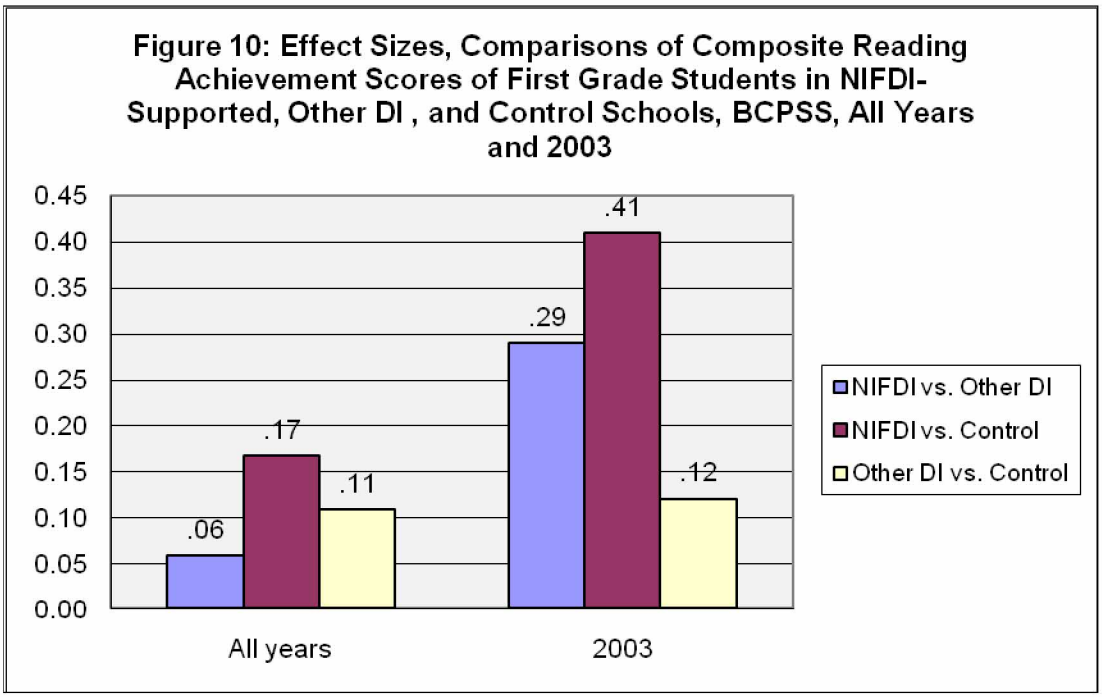
Note: The figures in this graph represent the percentage change in achievement from 1998 to 2003 for each of the groups of schools, separating the low poverty school in the Other DI group from others.

5. What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?

Unless specifically noted, all of the results presented earlier in this report are statistically significant. That is, various statistical tests indicate that the results do not appear by chance. However, it is possible that statistical significance can simply reflect sample size. When data from large samples are subjected to statistical tests relatively small substantive differences can produce significant results. To combat this problem, researchers use measures of effect size. These measures provide descriptions of the magnitude of the results. One of the most commonly used measures is Cohen's *d*. It is simply calculated as the difference between two average values divided by the common standard deviation. Thus, it reports the magnitude of a difference between two groups as a proportion of the standard deviation. A d value, or effect size, of .25 or larger has traditionally been considered educationally significant, indicating that an intervention has an important impact on students' achievement.[5]

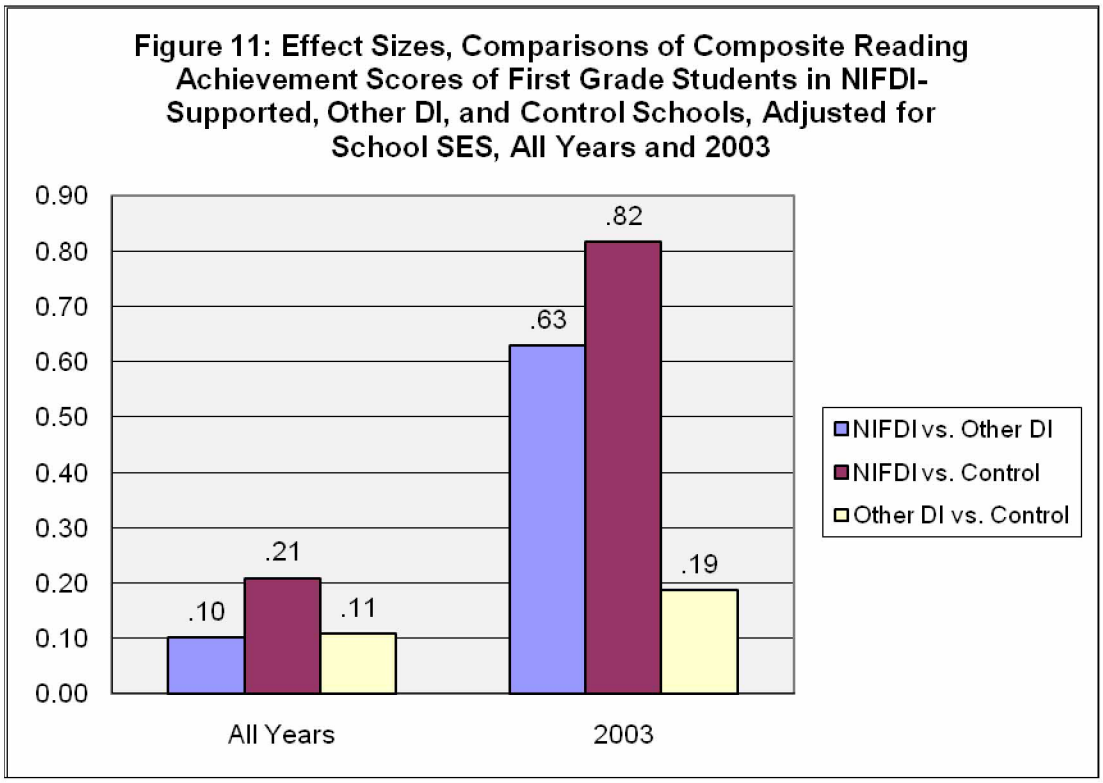
Figures 10 and 11 summarize the magnitude of the various effects reported earlier in this report. They compare the three groups: NIFDI-supported schools, Other DI schools, and the Control Schools. Figure 10 summarizes the effects without any adjustment for school characteristics. When data from all the years are examined together, the effect sizes comparing achievement of students in the NIFDI-supported schools and those in the two other groups of schools are lower than the .25 criterion (.06 for the comparison of NIFDI and the other DI schools and .17 for the comparison of the NIFDI and Control schools). However, by 2003 the effect sizes were substantially larger: .29 for the comparison of students in the NIFDI schools and in the other DI schools and .41 for the comparison with students in the Control schools. The comparisons of students in the Other DI group with those in the Control schools, while positive, were substantially lower in both comparisons (.11 for all years and .12 for 2003).

[5] This criterion is taken from Fashola and Slavin (1997).



Note: Values in the table are effect sizes, a measure of the magnitude of the difference between two groups. Effect sizes of .25 or larger are traditionally considered educationally significant.

Figure 11 gives the effect sizes for the measure of composite reading achievement for all years and 2003 after the impact of school socio-economic status has been controlled. Averaged across all years, the effect sizes in comparison to other DI schools was .10 and in comparison to the Control schools was .19. However, at the end of the study period, the effect of being in a NIFDI-supported school relative to other DI schools was .63 and relative to the Control schools was .82. The effect size of Other DI relative to the Control schools remained small and below the level of educational significance (values of 0.11 for all years and .19 in 2003).



Note: Values in the table are effect sizes, a measure of the magnitude of the difference between two groups. Effect sizes of .25 or larger are traditionally considered educationally significant.

Summary

This report analyzed differences in first grade reading achievement of students in the Baltimore City Public School System from 1997-98 through 2002-03. The BCPSS is typical of large urban school districts in serving large numbers of poor and minority students with low levels of academic achievement. Eleven schools in the BCPSS implemented Direct Instruction programs under the guidance of NIFDI during this time period. Five other schools used DI material but had NIFDI support for only the initial years. Beginning in 1998 the remaining schools used the Open Court curriculum. In the spring of each year all first graders were administered the Comprehensive Test of Basic Skills. Data from these tests were analyzed for this report.

Despite an extremely large body of empirical evidence, some continue to suggest that Direct Instruction is not effective in enhancing student achievement. The results from this study, with over 40,000 students and 6 years of data, would appear to contradict that statement. The results confirm previous research that has found the Direct Instruction curriculum to be effective in promoting student achievement. Averaging across all the years in the study, students exposed to Direct Instruction had higher achievement scores than students in the control schools.

The results also, however, provide an important caveat regarding the role of strong implementation and technical support. Once the socio-economic characteristics of schools were controlled (a necessary step given the low poverty rate of one of the Other DI schools), students in the NIFDI-supported schools had significantly higher achievement than both the students in the Control schools and those in the Other DI schools. By the end of the study

period the achievement scores of students in the NIFDI-supported schools were well above national averages and close to those of students in the low poverty Other DI group. Thus, while these results confirm the role of Direct Instruction in promoting high achievement, they also suggest that consistent, well-designed technical support may be important to ensure that these superior results occur.

The analysis may address some common objections to Direct Instruction. For instance, it has been claimed that Direct Instruction involves a relatively mindless application of a script with little intellectual engagement or skill required by the teacher. The results presented in this report showed that strong gains in achievement only appeared several years after schools began to implement DI and then only with consistent technical support. One could ask, "If DI simply employs mindless, rote tasks, why does it take so much time for the effects to appear?" A logical reply could be that it is not a mindless, rote task at all. Instead, learning to teach Direct Instruction involves a great deal of skill and practice. Those that are most successful have more practice and more skilled, technical guidance.

Second, some have suggested that Direct Instruction is useful for teaching rote, elementary skills such as decoding, but has more limited utility in teaching higher level skills such as comprehension. The results in this paper would contradict that view as well. Increases in achievement for students in the NIFDI-supported schools on comprehension were larger than those in vocabulary, the area more likely to be "rote" in nature.

Third, some have suggested that DI works only with low performers, but holds back high performing students. The results with the low poverty school that implemented DI, albeit without NIFDI support after the first few years, belie this suggestion. Even though students in this school had scores that averaged at the national mean at the beginning of the study, their scores increased markedly during the 6 years for which data were gathered. In addition, by the later years of the study, first grade students in the NIFDI schools had joined the ranks of high performers, with achievement scores well above the national norms. The scores of first graders in these schools continued to be maintained at these high levels and to even improve, providing no indication that their progress was impeded.

Appendix I
Sample and Measures

This appendix provides details on the sample and the measures used in the analysis.

Sample

The analysis uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Eleven schools implemented Direct Instruction with technical assistance provided by NIFDI. An additional 5 schools started implementation with NIFDI but then ceased involvement or began to work with another provider. The former group is referred to below as the NIFDI-supported schools, and the latter group as Other DI schools.[6] The remaining 103 schools are termed the Control schools. Schools in the two intervention groups are listed in Table A-1[7]

Table A-1: Schools in Treatment Groups and Start Dates of Treatment

NIFDI Schools	Starting Year
Arundel	96 - 97
CC Barrister	97 - 98
City Springs	96 - 97
Collington Square	98 - 99
Dickey Hill	98 - 99
Federal Hill	97 - 98
General Wolfe	96 - 97
Hampstead Hill	96 - 97
Langston Hughes	98 - 99
Margaret Brent	98 - 99
Dr. Rayner Browne	98 - 99
Other DI Schools	
William Pinderhughes	96 - 97
George Kelson	96 - 97
Robert Coleman	96 - 97
Roland Park	96 - 97
Westport	97 - 98

[6] One additional school, Charles Carroll, was part of the original intervention group but was closed shortly after the start of the study period. Because data are not available throughout the time span of the study, data from that school are not included.

[7] There were two additional schools in the set of control schools, but they, unfortunately, had the same name: Highland Town. The number assigned to the schools was not available for all years, and alternative spellings of the schools' name across years made it impossible to clearly differentiate them. Thus these two schools were eliminated from the analysis. Data were available for one home-schooled student and that was also omitted.

Table A-1 also lists the date at which programs were implemented in each school. In the analysis schools were designated as belonging to a given condition only during the years in which their school was receiving an implementation. For instance, Collington Square had no intervention in the first year of data collection (1997-98), but began implementing Direct Instruction reading programs in 98-99. Thus, students in Collington Square were determined to be in the NIFDI reading intervention group in 98-99 and subsequent years, but in the control condition in 1997-98. The analysis also considers the amount of time that a school had implemented the program. For instance, Collington Square was considered to be in its first year of implementation of reading in 1998-99, its second in 1999-2000, etc.

Measuring Achievement

The Comprehensive Test of Basic Skills (CTBS), a widely used standardized achievement test, was administered to all first graders in the spring of each year, from 1997-1998 through 2002-2003. The 4th edition was administered in the spring of 1998 and 1999 and the 5th edition was administered in the spring of 2000 through 2003. Two subtest scores, Reading Comprehension and Reading Vocabulary, and a Composite Reading Achievement score were analyzed.

Normal Curve Equivalent (NCE) scores were used for all statistical analyses. NCE scores are calculated by translating the raw or percentile scores to a distribution that is normally distributed. Like percentile scores, NCE scores range from 1 to 99 with a mean of 50. However, while percentile scores are evenly distributed (the graph of percentile scores would look like a rectangle), NCE scores comprise a normal distribution. The transformation results in scores that can be meaningfully added and subtracted, so that the difference of an NCE score of 50 and 55 (=5) is the same as the difference between 30 and 35. This interval scale allows computations of virtually all of the common statistics used in analysis. Percentiles and raw scores do not meet the technical requirements of interval scales, and thus statisticians urge researchers to analyze data based on NCE scores rather than percentiles or raw scores. All statistical calculations for this report were done using NCE scores.

The meaning of NCE scores is, however, not intuitively obvious. For this reason, results in the body of the text and the executive summary have been translated into percentile scores. After completing statistical calculations with NCE scores the resulting descriptive information was converted into percentiles for display in the graphs in the body of the report. This conversion was done using a standard conversion table.[8]

Measuring School Context

Given the strong influence of school context on student achievement, it was important to develop an efficient, yet strong, measure of the demographic context of the schools in the sample. Preliminary analysis indicated that the demographic characteristics of schools were very highly correlated from one year to another. Thus, one summary measure was developed for each school that would be valid for all the years included in the data set.

There were a few cases in which data were not available for a school for all years. A regression-based method was used to predict values of missing cases from other years (e.g.

[8] The conversion can also be accomplished manually using a normal curve table. The NCE scores can be converted to z-scores (z = ((nce-50)/21.06))). The percentile that corresponds to the z score can then be found in the normal curve table.

predicting 1998 levels of proportion African-American from levels in 1999 through 2003).[9] Values for each variable were averaged across the years to produce aggregate measures of the demographic characteristics of the school.

Table A-2: Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels, Total and By Treatment Group

A: Descriptive Statistics

	NIFDI		Other DI		Control		Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Asian	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01
African-American	0.75	0.34	0.93	0.13	0.85	0.27	0.84	0.27
Hispanic	0.05	0.12	0.002	0.003	0.01	0.01	0.01	0.04
Native American	0.01	0.02	0.001	0.001	0.00	0.00	0.003	0.01
Non-Hispanic White	0.17	0.25	0.06	0.12	0.14	0.25	0.14	0.25
Free and Reduced Lunch	0.83	0.06	0.72	0.21	0.74	0.15	0.75	0.15
Factor 1	-0.12	0.96	-0.07	1.06	0.02	1.01	0.01	1.00
Factor 2	1.18	2.82	-0.34	0.25	-0.11	0.47	0.00	1.00
N	11		5		103		119	

B: Analysis of Variance and Tukey Post-Hoc Tests

	Analysis of Variance		Post-Hoc Comparisons (significance)		
	F	p	NIFDI v. Other DI	NIFDI v. Control	Other DI v. Control
Asian	2.01	0.14	0.61	0.12	0.96
African-American	0.85	0.43	0.46	0.54	0.78
Hispanic	5.51	0.01	0.08	0.004	0.95
Native American	6.99	0.001	0.04	0.001	0.92
Non-Hispanic White	0.35	0.70	0.68	0.90	0.78
Free and Reduced Lunch	2.06	0.13	0.35	0.12	0.97
Factor 1	0.12	0.89	1.00	0.89	0.98
Factor 2	9.76	<.001	0.009	<.001	0.85

[9] There were two cases where a predicted value fell outside of the theoretical range. Both of these involved the predictions for proportion white, where the predicted values were less than zero. For these cases (Malcolm X and Mildred Monroe schools), an average of the other years in the data file was used as the predictor. The average values for both cases were .01 or less.

Panel A of Table A-2 gives means and standard deviations on each of the demographic characteristics for each group of schools. Analysis of variance indicated no significant differences between the schools in their proportion of African American, non-Hispanic White, or Asian students, as well as no difference in the proportion receiving free or reduced lunch. The NIFDI schools, however, had significantly more Hispanic and Native American students than the students in the other two groups of schools.

A factor analysis was conducted to develop summary measures of school context. Six variables were included in this analysis: the average proportion of Asian, Hispanic, African American, Native American, and white students as well as the average proportion of students receiving free and reduced lunch. A principle components extraction method was used, with varimax rotation. Two significant factors (with eigenvalues greater than 1.0) were found.[10] Factor scores were computed for each school and saved.

Results for the factor analysis are given in Table A-3. The commonalities indicate that, except for the measure of proportion of Asian students, there is a great deal of shared variance among the indicators, ranging from .66 for the proportion of students receiving free and reduced lunch to .94 for the proportion of African American students. Eigenvalues are also relatively high, with 73.5 percent of the total variance between schools explained by the two factors.[11]

Table A-3: Factor Analysis of School Characteristics, Elementary Schools, BCPSS,1998-2003

	Rotated Factor Loadings		Communalities
	Factor 1	Factor 2	
Proportion Asian	0.64	0.09	0.42
Proportion African-American	-0.88	-0.40	0.94
Proportion Hispanic	0.13	0.85	0.74
Proportion Native American	0.08	0.87	0.77
Proportion White	0.90	0.27	0.89
Proportion Free/Reduced Lunch Eligible	-0.80	0.14	0.66
Eigenvalues	2.66	1.75	
% of total variance	44.4	29.1	

N = 120 schools

The first factor, which accounts for 44 percent of the total variance, has a strong positive loading of the proportion of white students, a slightly smaller positive loading for the proportion of Asian students, and strong negative loadings for the proportion of African American students and the proportion receiving free and reduced lunch. Thus, schools with positive scores on this factor would have proportionately more white and Asian students and

[10] The factor analyses were also conducted with data for each year separately (i.e. the non-aggregated data), and the results were virtually identical to those obtained with the aggregated data.
[11] N = 120 for these analyses, for Carroll School was included. Results are identical when Carroll is omitted.

many fewer African-American and poor students. As can be seen in Table A-2, there are no significant differences between the three treatment groups on this factor.

The second factor, which accounts for 29 percent of the total variance, has positive loadings on the proportion of Hispanics and the proportion of Native Americans in a school, and a negative loading, of a somewhat smaller magnitude, of the proportion of African Americans. Thus, schools with a positive score on this factor would have proportionately more Hispanic and Native American students and somewhat fewer African American students. The results in Table A-2 indicate that the NIFDI schools had significantly higher scores on this factor.

In addition, one of the schools in the Other DI group, Roland Park, was strikingly different from the other schools in its rate of children receiving free and reduced lunch. On average only .34 of the students in Roland Park had free or reduced lunch, the third lowest rate in the BCPSS. The other schools in the Other DI group had free lunch proportions ranging from .76 to .84, while schools in the NIFDI group had proportions ranging from .74 to .90. While the presence of Roland Park in one of the treatment groups allows the examination of the effects of DI in more advantaged schools, it was also important to control for the presence of this school in the various analyses.

Appendix II: Detailed Statistical Results

This appendix expands upon the material presented in the text including an extended discussion of the analysis techniques employed and the results that address each research question.

Methodology

As noted in the text, the analysis of data focused on comparing the achievement of students who received Direct Instruction, in the NIFDI schools and in the Other DI schools, with those in the Control Schools. Six research questions were addressed, each providing a slightly more complex method of control to help ensure the accuracy of the results.

The first research question was, "Did students who received Direct Instruction have higher achievement scores than students in the other schools?" Simple t-tests were used to answer this question, comparing the average achievement scores of students in the NIFDI and other DI schools with scores of students in the Control schools, combining data from all years of implementation.

The second research question asked, "Did students who received Direct Instruction in schools that received consistent NIFDI technical support have higher achievement scores than students in the other groups?" The analysis for this question compared the average scores of students in the three groups of schools (NIFDI-supported, Other DI, and Control Schools) across all the years of data using simple one-way analysis of variance and post-hoc tests to compare subgroup means.

The third research question asked, "Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized?" This analysis is especially important given the unique low poverty status of one school in the Other DI group. Two approaches were used. The first was a simple one-way analysis of variance, separating data from the low poverty Other DI school from other schools within that group. The second method was mixed model regression, including measures of school advantage and disadvantage from the factor analysis as predictors. To the extent that Direct Instruction, as implemented and supported by NIFDI, is more effective at promoting student achievement, it would be expected that students in the NIFDI group would have higher achievement scores than those in the other two groups, even when school characteristics were controlled.

Mixed models are particularly appropriate for analyzing multi-level data, such as data regarding students and the schools that they attend. In these models a "random variable" is used to control for differences between schools (often termed the level 2 entity) while calculating regression coefficients regarding the impact of variables from both students and schools on achievement. The random variable is equivalent to having a separate intercept in the regression equation for each school. The coefficients associated with the various individual and school related variables are then calculated while this between school variance is controlled. The analysis also allows one to calculate the amount of variance in the dependent variable that occurs between schools and the extent to which various independent variables can account for this between school variance.

The fourth research question asked, "Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?" It was expected that the impact of the NIFDI implementation would become stronger over time. This question was analyzed both through examination of mean scores over time, with analysis of variance and post-hoc comparison of means, as well as through adding variables regarding length of implementation to the mixed models.

The final research question asked, "What is the magnitude of the effect of NIFDI-supported Direct Instruction." To address this question, the results were translated into effect sizes. Effect sizes are commonly used to summarize the magnitude of differences between two groups. They are calculated by simply dividing the difference in the means of two groups by their common standard deviation. The resulting figure thus describes the magnitude of a difference between two groups as a percentage of the standard deviation. Unlike tests of significance, which can be heavily influenced by sample size, effect sizes are unaffected by the size of a sample. Effect sizes greater than .25 have traditionally been considered educationally significant (Fashola and Slavin 1997).

Research Question 1: Did students who received Direct Instruction have higher achievement scores than students in the other schools?

The first research question addressed was, **"**Do students who received Direct Instruction have higher achievement scores than students in the other schools?" This was answered through comparing the average achievement scores of students in the combined NIFDI and Other DI groups of schools with students in the Control schools. In this analysis data for all years in the sample (1997-98 through 2002-03) were combined.

Table A-4 summarizes the information used for these comparisons. Panel A reports the means and standard deviations of the normal curve equivalent scores for students who received Direct Instruction and students in the Control schools on each of the measures of achievement as well as the results of t-tests addressing the hypothesis that any differences occur by chance. Panel B translates the average Normal Curve Equivalent scores in Panel A to the corresponding percentiles. Data in Panel B were used to create Figure 1 in the text. The values in Panel B may be interpreted as the score of an average student in each type of school over all the years in the study.

The results in Table A-4 show that students in schools with Direct Instruction had significantly higher achievement scores than those in the Control Schools on all three measures: Comprehension, Vocabulary, and the Composite score. The average student in the Control schools, across all the years in the study, had scores ranging from the 41st to 43rd percentile. In contrast, the average student in a DI school had scores ranging from the 45th to the 50th percentile. On both comprehension and the composite score the average DI student had a score at the national average. The lowest scores for both groups occurred with the measure of Vocabulary.

Table A-4: Average Reading Achievement Scores of Students With and Without Direct Instruction, All Years, First Graders, BCPSS

Panel A: Descriptive Statistics, Normal Curve Equivalent Scores, and t-test Results

Comprehension	Control Schools	Had Direct Instruction
Mean	45.6	49.8
S.D.	23.7	24.7
N	41043	4820
t-value	11.43	5905.60
d.f.	5906	
Prob.	<.001	
Vocabulary		
Mean	45.4	47.2
S.D.	23.1	24.7
N	40825	4754
t-value	4.78	5764.08
d.f.	5764	
Prob.	<.001	
Composite		
Mean	46.6	50.1
S.D.	23.0	24.5
N	40425	4705
t-value	9.23	5711.99
d.f.	5711	
Prob.	<.001	

Panel B: Percentile Scores Corresponding to Mean Values

	Control School Students	Direct Instruction Students
Comprehension	42	50
Vocabulary	41	45
Composite	44	50

Research Question 2:"Did students who received Direct Instruction in schools that received consistent NIFDI technical support have higher achievement scores than students in the other groups?"

Simple one-way analysis of variance was used to analyze differences in average achievement between the three groups: NIFDI-supported schools, Other DI schools, and the Control schools. Table A-5 summarizes the results of this analysis. Panel A gives the means and standard deviations on the three measures of achievement. Panel B includes results of one-way analyses of variance and Tukey post-hoc tests comparing each pair of schools. Finally, Panel C translates the average normal curve equivalent scores in Panel A into the corresponding percentiles. The results in Panel C were used to create Figure 2 in the text.

Table A-5: Average Reading Achievement Scores by Group, All Years, First Graders, BCPSS

Panel A: Descriptive Statistics (NCE Scores) and Analysis of Variance Results

Comprehension	NIFDI	Other DI	Control	Total
Mean	50.5	48.7	45.6	46.0
S.D.	24.4	25.2	23.7	23.8
N	2991	1829	41043	45863
Vocabulary				
Mean	47.3	47	45.4	45.6
S.D.	25.2	24	23.1	23.3
N	2948	1806	40825	45579
Composite				
Mean	50.6	49.2	46.6	47.0
S.D.	24.5	24.6	23	23.2
N	2911	1794	40425	45130

Panel B: Analysis of Variance and Post-Hoc Pair-Wise Comparisons

	Analysis of Variance		Tukey Post-Hoc Tests (p)		
	F Value	prob.	NIFDI v. Other DI	NIFDI v. Control	Other DI v.Control
Comprehension	73.17	<.001	0.03	<.001	<.001
Vocabulary	12.85	<.001	0.86	<.001	0.01
Composite	49.19	<.001	0.09	<.001	<.001

Panel C: Percentile Scores Corresponding to Means in Panel A

	NIFDI	Other DI	Control	Total
Comprehension	51	48	42	42
Vocabulary	45	44	41	42
Composite	51	48	44	44

Inspection of the means shows that students in the NIFDI-supported schools had higher achievement than students in the Other DI schools or the Control schools on all three

measures. The analysis of variance results indicate that there are significant differences between these three groups on all three measures. All of the pair-wise comparisons with the control schools, for both the NIFDI-supported and the Other DI schools, are statistically significant. However, the differences between the NIFDI schools and the Other DI schools are only significant for the measure of comprehension.

Panel C, which translates the Normal Curve Equivalent scores to percentiles, indicates that the average student in a NIFDI-supported school, across all the years in the study, scored slightly above the national average on Comprehension and the Composite score, but somewhat below the average in Vocabulary. Percentiles for the average student in the Other DI group and the Control schools were below the national average for all measures.

Research Question 3: Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized?

Research Question 3 deals with the extent to which students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized. This question is answered in two ways. The first is with descriptive data and analyses of variance and the second uses mixed model regressions.

Descriptive Data and Analyses of Variance

The descriptive statistics, analysis of variance, and post-hoc comparisons of means are in Table A-6. The analysis parallels that of Table A-5 except that, within the Other DI group, the results for the low poverty school are separated from those for the other schools.

Table A-6: Average Reading Achievement Scores by Group, All Years, First Graders, BCPSS, Four Groups
Panel A: Descriptive Statistics (NCE Scores) and Analysis of Variance Results

Comprehension	NIFDI	HPODI	LPODI	Control	Total
Mean	50.5	44.2	65.3	45.6	46.0
S.D.	24.4	24.1	22.0	23.7	23.8
N	2991	1434	395	41043	45863
Vocabulary					
Mean	47.3	42.2	63.8	45.4	45.6
S.D.	25.2	22.8	20.4	23.1	23.3
N	2948	1411	395	40825	45579
Composite					
Mean	50.6	44.2	66.8	46.6	47.0
S.D.	24.5	23.1	21.7	23	23.2
N	2911	1399	395	40425	45130

Panel B: Analysis of Variance and Post-Hoc Pair-Wise Comparisons (probabilities)

	F Value	prob.	NIFDI v. HPODI	NIFDI v. LPODI	NIFDI v. Control	HPODI v. LPODI	HPODI v. Control	LPODI v. Control
Comprehension	130.85	<.001	<.001	<.001	<.001	<.001	0.131	<.001
Vocabulary	96.95	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Composite	130.65	<.001	<.001	<.001	<.001	<.001	0.001	<.001

Panel C: Percentile Scores Corresponding to Means in Panel A

	NIFDI	HPODI	LPODI	Control	Total
Comprehension	51	39	76	42	42
Vocabulary	45	36	74	41	42
Composite	51	39	79	44	44

Note: HPODI signifies the high poverty Other DI schools, which are demographically similar to the NIFDI-supported schools, LPODI refers to the low poverty Other DI school.

The results in Table A-6 show that when the low poverty other DI (LPODI) school is removed from the Other DI group, the differences between the achievement scores of students in that group and students in the NIFDI-supported schools become even larger. All of the differences between the students in the NIFDI-supported schools and those in the high poverty Other DI (HPODI) schools are highly statistically significant, as are all the comparisons with the control schools. Comparisons of students in NIFDI-supported schools with those in the LPODI school are also highly significant, but the result is in favor of the low poverty school. In fact, averaging over all years in the study, first grade students in this low poverty, DI school had the second highest average score in the BCPSS on both the vocabulary and the composite measures and the highest measure of all schools in comprehension.

Mixed-Model Results

The second method of addressing this research question was through the use of mixed model regressions. Table A-7 summarizes the models that were tested and gives summary information on the results with each model. Panel A gives the variables that are included and Panel B gives the model fit statistics.

Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average reading achievement. The correlation ratio attached to Model 1 is the proportion of

variance in the dependent variable that is between schools as opposed to between students. It can be seen that from nine to ten percent of the variance in achievement is between schools rather than simply between students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model and school differences are controlled. These null hypotheses can be easily rejected with Model 1 and with all subsequent models. There is significant variation between schools and also between students in all models that we examine. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those available to in this analysis.

Table A-7: Variables in Models and Measures of Fit, Analysis of First Grade Reading Achievement, BCPSS, 1998-2003

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3	Model 4
Random intercept for Schools	x	x	x	x
Year		x	x	x
Factors 1 and 2		x	x	x
Group (NIFDI, Other DI)			x	x
Years of Treatment				x
Treatment * NIFDI				x
Factor 1 * NIFDI				
Factor 1 * Other DI				
Panel B: Model Fit Statistics				
Comprehension				
Random Effects Estimate	49.2	38.2	36.3	35.8
s.e.	6.7	5.2	4.9	4.9
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	522.9	467.63	467.52	466.09
s.e.	3.5	3.1	3.1	3.1
sig.	<.0001	<.0001	<.0001	<.0001
BIC	417661	412535	412528	412396
- 2 Log Likelihood	417647	412506	412490	412348
Change in LL		5140.2	16.4	141.9
df		3	2	2
P		<.001	<.001	<.001
PRE measure		0.11	0.0002	0.003
Correlation ratio	0.09			
Vocabulary				
Random Effects Estimate	52.3	40.2	39.8	40.1
s.e.	7.1	5.4	5.4	5.5
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	497.7	480.3	480.2	477.9
s.e.	3.3	3.2	3.2	3.2
sig.	<.0001	<.0001	<.0001	<.0001

BIC	412833	411204	411198	410994
-2 Log Likelihood	412818.7	411175	411160	410946
Change in LL		1643.6	15.1	214
df		3	2	2
P		<.001	<.001	<.001
PRE measure		0.03	0.0003	0.005
Correlation ratio	0.10			
Composite				
Random Effects Estimate	54.7	41.8	40.4	40.2
s.e.	7.4	5.6	5.4	5.4
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	490.7	444.95	444.8	442.9
s.e.	3.3	3.0	3.0	3.0
sig.	<.0001	<.0001	<.0001	<.0001
BIC	408143	403719	403710	403524
-2 Log Likelihood	408128	403690	403672	403476
Change in LL		4437.8	18.7	195.9
df		3	2	2
sig.		<.0001	<.001	<.0001
PRE measure		0.09	0.0003	0.004
Correlation ratio	0.10			

Note: The correlation ratio is computed by dividing the random effects estimate for schools by the sum of the estimate for schools and the residual. Thus, the ratio represents the proportion of total variation that is between schools. The PRE measure is the ratio of the difference of residual estimates of two models divided by the estimate from the less complex model. Thus it tells the proportionate change in the variance that occurs by adding more variables to a model.

The models become incrementally more complex, with each subsequent model adding one or more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the year in which data were collected and the two factor scores to the baseline model. Year was included to test the hypothesis that test scores vary over the years in the study (1997-98 through 2002-03). This is important to control for any general changes within the district. Including the two factor scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African American students and more non-Hispanic white students, would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient). Thus, the results with Models 2 and 3 provide a second way of testing Research Question three.

The -2 log likelihood measures and the BIC values in Panel B of Table A-7 can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the

comparisons between these values are in the bottom part of each section of Panel B of Table A-7. For example, Model 2 provides a significantly better fit to the data than Model 1 for all measures. For comprehension, the change in the -2 Log Likelihood = 5174 (417,647 - 412,506). With three degrees of freedom (because three new variables were added to Model 2 compared to Model 1), this result is highly significant. The comparisons of Model 3 with Model 2 and of Model 4 with Model 3 also indicate that adding the variables in the more complex model significantly improves the fit. (See the significance associated with the change in the -2 log likelihood.)

The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. Looking at all the models in Table A-4 it may be seen that the lowest BIC values appear for Model 4. The Proportionate Reduction of Error (PRE) measures are another descriptive measure of the incremental changes in fit of the models and simply reflect the proportionate changes in the residual variance from one model to the next. The greatest changes occur from Model 1 to Model 2 and the proportionate reduction of error in prediction is substantially less for each of the more complex models. Again, this could be expected given the large number of factors that affect student achievement.[12]

Table A-8 gives the coefficients associated with each of the models in Table A-7, beginning with Model 2. As expected, the coefficient associated with Factor 1 (where higher scores indicate schools with fewer students on free and reduced lunch and more non-Hispanic whites) is positive and highly significant. Also as expected, the coefficient associated with Factor 2 (where higher scores indicate more Hispanic and Native American children) is negative, but is not significant in Model 2. The coefficient for year is positive and significant in all models, indicating that, over time, students' achievement scores increased.

The coefficients in Model 3 directly test Research Question 3 by including both the dummy variables for treatment group and the controls for the measures of school context. Two dummy variables were used to designate students in schools receiving the two treatments: NIFDI directed interventions and DI interventions not administered by NIFDI. The Control schools are the omitted category for both variables. If NIFDI support has a unique impact we would expect stronger associations with that variable. It should be recalled that students were only coded as having the treatment if they were in a school in a year when the school was in a treatment condition.

The coefficient associated with being in a NIFDI-supported school is positive and highly significant ($p<.0001$) for all three measures of achievement. In contrast, the coefficient associated with being in an Other DI school, while positive, is not significant in any of the models. In other words, once the demographic context of the school is controlled, students in the NIFDI-supported schools had significantly higher achievement than both students in the Control schools and students in the Other DI schools.

Table A-9 shows the predicted achievement scores for students in each of the three groups of schools based on the results in Model 3. These values were calculated by substituting average values for the entire sample for year and the two factor scores into Model 3 and calculating predicted values for students in each of the three groups. Thus, the

[12] It should be noted that the values of BIC, the changes in the Log-Likelihood Ratio, and the PRE measures would alter if the order in which variables were introduced were changed. For example, the PRE measure associated with adding the treatment groups to the model is 10 times as large when the groups are entered before the measures of year and factor, but still not as large as when year and factor are entered.

resulting values give the achievement score that would be predicted if students were equalized on the factor scores and time. Panel A gives the Normal Curve Equivalent scores and Panel B gives the percentile equivalent of these scores. Values in Panel B may be interpreted as the percentile at which an average student in each of the three groups would be expected to achieve if schools were equal in their demographic characteristics. The values in Panel B were used in Figure 3 in the body of the text.

Table A-8: Coefficients Associated with Mixed Model Regressions of First Grade Reading Achievement, BCPSS, 1998-2003

	Model 2			Model 3			Model 4		
Comprehension	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	-8827.2	120.6	<.0001	-8776.0	121.2	<.0001	-8363.1	127.6	<.0001
Year	4.4	0.1	<.0001	4.4	0.1	<.0001	4.2	0.1	<.0001
Factor 1	2.3	0.6	<.0001	2.3	0.6	<.0001	2.4	0.6	<.0001
Factor 2	-0.5	0.6	0.39	-1.0	0.6	0.10	-1.5	0.6	0.01
NIFDI	-----	-----	-----	4.6	1.2	<.0001	-3.8	1.4	0.01
Other DI	-----	-----	-----	3.2	2.8	0.26	0.4	3.0	0.91
Years of Treatment	-----	-----	-----	-----	-----	-----	0.7	0.3	0.020
Treatment * NIFDI	-----	-----	-----	-----	-----	-----	2.3	0.4	<.0001
Vocabulary	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	-4920.2	122.3	<.0001	-4868.9	123.0	<.0001	-4372.0	129.4	<.0001
Year	2.5	0.1	<.0001	2.5	0.1	<.0001	2.2	0.1	<.0001
Factor 1	2.8	0.6	<.0001	2.9	0.6	<.0001	2.9	0.6	<.0001
Factor 2	-0.8	0.6	0.16	-1.3	0.6	0.03	-2.0	0.6	0.001
NIFDI	-----	-----	-----	4.6	1.2	0.0001	-5.9	1.4	<.0001
Other DI	-----	-----	-----	1.6	2.9	0.59	-0.9	3.2	0.78
Years of Treatment	-----	-----	-----	-----	-----	-----	0.6	0.3	0.04
Treatment * NIFDI	-----	-----	-----	-----	-----	-----	3.1	0.4	<.0001
Composite	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	-8011.6	118.3	<.0001	-7956.1	119.0	<.0001	-7490.8	125.2	<.0001
Year	4.0	0.1	<.0001	4.0	0.1	<.0001	3.8	0.1	<.0001
Factor 1	2.7	0.6	<.0001	2.7	0.6	<.0001	2.8	0.6	<.0001
Factor 2	-0.7	0.6	0.24	-1.2	0.6	0.04	-1.8	0.6	0.003
NIFDI	-----	-----	-----	5.0	1.2	<.0001	-4.7	1.4	0.001
Other DI	-----	-----	-----	2.6	3.0	0.38	-0.1	3.2	0.98
Years of Treatment	-----	-----	-----	-----	-----	-----	0.7	0.3	0.03
Treatment * NIFDI	-----	-----	-----	-----	-----	-----	2.8	0.4	<.0001

Table A-9: Predicted Achievement Scores Assuming Equal School Contexts

	NIFDI	Other DI	Control
Panel A: Normal Curve Equivalent Scores			
Comprehension	50.1	48.7	45.5
Vocabulary	49.6	46.6	45.0
Composite	51.3	48.8	46.2
Panel B: Percentile Scores			
Comprehension	50.2	47.6	41.6
Vocabulary	49.2	43.6	40.6
Composite	52.4	47.8	42.8

Note: Scores were computed by substituting average values for entire sample for Year and Factors 1 and 2 into the equations associated with Model 3 in Table A-8. The average value for year was 2000.3. The average value for Factor 1 was -.03203, the average value for Factor 2 was -.04252. The value for year did not equal 2000.5 (the average of 1998-2003) because there were fewer students in the BCPSS in later years than in earlier years. The average values for Factors 1 and 2 do not equal zero because the unit of analysis for computing the average values was students rather than schools. If the alternative values were used in computations, the pattern of substantive results would not differ.

Research Question 4: Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?

Research question four tests the hypothesis that the advantage to students in the NIFDI schools would become stronger as practices and procedures were institutionalized within a school. As noted in the text, it takes time for teachers and other school staff to fully adjust to and incorporate the nuances of a new curriculum. Thus, it would be expected that the advantages accruing to attending a NIFDI school would become greater over time.

Model 4 in Tables A-7 and A-8 addresses this hypothesis by adding the interaction of membership in the NIFDI group and the years of treatment. Years of treatment measures how long DI had been implemented in a school.[13] It was expected that the coefficient associated with the interaction of years of implementation and NIFDI would be positive, indicating stronger effects of NIFDI as years of treatment increased.

The results show very strong support for this hypothesis. The positive and significant coefficients associated with year indicate that achievement increased for all students in the BCPSS over the years of the study, and the positive and significant coefficients associated with the variable "years of treatment" indicate that, besides this general increase, students in

[13] The interaction of year of treatment and membership in the Other DI schools is not needed in the equation because this is captured by the years of treatment variable.

the schools that implemented DI had additional increases. Yet, students in a NIFDI-supported school had even stronger increases in achievement from year to year. With all three measures, the results with Model 4 indicate a strong and significant impact of the interaction of being in a NIFDI school and years of implementation.

The coefficients associated with Model 4 in Table A-7 can be used to calculate the average change expected in the three groups of schools from one year to the next. For instance, first grade students in the control schools had an average gain from one year to the next of 3.8 points in their normal curve equivalent scores on the composite measure of reading achievement (the coefficient associated with year), controlling for differences in the demographic context of the schools. First grade students in the Other DI schools had an average gain of 4.5 normal curve equivalent points with each additional year of implementation of Direct Instruction (summing the impact of year and the impact of years of implementation: 3.8 + 0.7). In contrast, the average change in the NIFDI schools, net of school context, was 7.3 NCE points from one year of implementation to the next (3.8 + 0.7 + 2.8), a yearly expected increase in the average scores of first graders that was almost twice the magnitude of that of the Control schools.

Table A-10 gives the scores that would be predicted for students in the three groups of schools if they attended schools with similar demographic characteristics for both one year of implementation of Direct Instruction and 7 years of implementation. These scores were obtained by using the regression coefficients in Table A-8 and substituting the average factor scores and year for all cases in the sample, as was done for the calculations reported in Table A-9.[14] Panel A gives the resulting normal curve equivalent scores and Panel B translates these scores into the corresponding percentiles. The scores in Panel B can be interpreted as the percentile that an average student in a group of schools would be expected to achieve at a given year and were used to create Figure 5 in the text.

Results show that with only one year of implementation the average scores of students in the three groups are very similar. However, after seven years of implementation, differences are dramatic. Assuming that they attended schools with similar socio-economic contexts, the average student in a NIFDI-supported school would be expected to score between the 84th and 89th percentile, the average student at an Other DI school would score from the 58th to the 72nd percentile, and the average student at a control school would score from the 52nd to the 63rd percentile, depending on the measure of achievement.

[14] The average values for Factor 1 and Factor 2 that were used for calculations in Table A-9 were also used for the calculations for Table A-10. For year, the average value used in the calculations was 1998.7 for year 1 of implementation and 2003 for year 7. The value of 1998.7 was determined by calculating the average year in which schools were in their first year of implementation of Direct Instruction. The value 2003 was used for year 7 because that was the only year in which DI schools had 7 years of implementation.

Table A-10: First Grade Reading Achievement Scores by Group and Years of Implementation, Adjusting for School Advantage, BCPSS

Panel A: Normal Curve Equivalent Scores

	Years of Implementation	NIFDI	Other DI	Control
Comprehension	1	38	40	39
	7	74	62	57
Vocabulary	1	39	41	41
	7	71	54	51
Composite	1	39	41	40
	7	76	61	56

Panel B: Expected Percentile Score of Average Student

	Years of Implementation	NIFDI	Other DI	Control
Comprehension	1	29	32	30
	7	87	72	63
Vocabulary	1	30	34	34
	7	84	58	52
Composite	1	30	34	32
	7	89	70	61

Table A-11 presents the data that were used to construct Figures 6, 7, and 8 in the text. Unlike data in Tables A-9 and A-10, these data are simply the average achievement scores in each year on each measure for each of the three groups. Data for the low poverty school in the Other DI group are presented separately. No other adjustments were included for school advantage/disadvantage, nor are there any adjustments for how long the schools had been implementing DI. They simply compare the average scores in the schools in the four different groups from 1998 to 2003. The F-tests summarize the analysis of variance results testing the null hypothesis that the mean scores of students in the 4 groups are equal. The pair-wise tests, using Tukey post-hoc comparisons, test the hypothesis that each pair of means is equal.[15]

The results show that in the first year for which data are available (1998) students in the NIFDI schools had the lowest average scores of students in any of the four groups. Pair-wise tests show that these differences with students in the control schools and in the low poverty Other DI school were statistically significant on all three measures and differences with students in the high poverty Other DI schools were significant on the measure of

[15] Recall that a number of the schools (all but one of the Other DI schools and four of the NIFDI-supported schools) began using DI in 1996-97, the year before the first year of data in the data set (see Table A-1).

vocabulary. By the second year students in the NIFDI schools had begun to catch up, and by the last two years they had scores that were significantly higher than students in both the Control schools and the high poverty Other DI schools. Students in the low poverty Other DI school had significantly higher scores than students in the other schools on all three measures throughout the scope of the study.

Table A-11: Reading Achievement Scores by Year and Group, First Graders, BCPSS, 1998-200

Panel A: Means(NCE Scores) and Analysis of Variance Results

	Group				4-Way Comparisons	
	NIFDI	HPODI	LPODI	Control	F	prob.
Comprehension						
1998	31	32	47	35	14.59	<.001
1999	37	35	55	38	14.91	<.001
2000	49	50	66	47	16.92	<.001
2001	56	56	73	51	32.854	<.001
2002	63	49	70	53	42.108	<.001
2003	64	50	78	55	49.59	<.001
Percent Increase	106	56	66	57		
Vocabulary						
1998	30	35	52	39	26.78	<.001
1999	39	39	58	44	19.45	<.001
2000	43	44	66	46	19.99	<.001
2001	53	49	69	48	28.457	<.001
2002	57	46	64	50	26.46	<.001
2003	59	46	71	50	45.39	<.001
Percent Increase	97	31	37	28		
Composite						
1998	30	33	50	37	21.286	<.001
1999	38	37	57	41	16.98	<.001
2000	49	48	70	48	19.20	<.001
2001	57	54	75	51	35.044	<.001
2002	63	50	71	54	39.9	<.001
2003	64	50	78	54	55.29	<.001
Percent Increase	113	52	56	46		

Panel B: Pair-Wise Comparisons (sig.)

	NIFDI v. HPODI	NIFDI v. LPODI	NIFDI v. Control	HPODI v. LPODI	HPODI v. Control	LPODI v.Control

Comprehension						
1998	1.00	<.001	0.003	<.001	0.01	<.001
1999	0.43	<.001	0.77	<.001	0.07	<.001
2000	0.93	<.001	0.14	<.001	0.13	<.001
2001	1.00	<.001	<.001	<.001	0.003	<.001
2002	<.001	0.08	<.001	<.001	0.12	<.001
2003	<.001	<.001	<.001	<.001	0.05	<.001
Vocabulary						
1998	0.04	<.001	<.001	<.001	0.02	<.001
1999	1.00	<.001	<.001	<.001	0.004	<.001
2000	1.00	<.001	0.07	<.001	0.50	<.001
2001	0.14	<.001	<.001	<.001	0.91	<.001
2002	<.001	0.066	<.001	<.001	0.18	<.001
2003	<.001	<.001	<.001	<.001	0.07	<.001
Composite						
1998	0.48	<.001	<.001	<.001	0.01	<.001
1999	0.95	<.001	0.011	<.001	0.04	<.001
2000	1.00	<.001	0.97	<.001	1.00	<.001
2001	0.48	<.001	<.001	<.001	0.22	<.001
2002	<.001	0.05	<.001	<.001	0.13	<.001
2003	<.001	<.001	<.001	<.001	0.03	<.001

Panel C: Percentile Equivalents of Normal Curve Equivalent Scores

	NIFDI	HPODI	LPODI	Control
Comprehension				
1998	18.3	19.6	44.3	23.8
1999	26.8	23.8	59.6	28.6
2000	48.2	50.0	77.6	44.3
2001	61.2	61.2	86.2	51.8
2002	73.2	48.2	82.9	55.7
2003	74.1	50.0	90.8	59.6
Vocabulary				
1998	17.1	23.8	53.8	30.0
1999	30.0	30.0	64.8	38.8
2000	37.0	38.8	77.6	42.4
2001	55.7	48.2	81.7	46.2
2002	63.0	42.4	74.1	50.0
2003	66.5	42.4	84.1	50.0
Composite				
1998	17.1	21.0	50.0	26.8
1999	28.6	26.8	63.0	33.5
2000	48.2	46.2	82.9	46.2
2001	63.0	57.6	88.3	51.8

2002	73.2	50.0	84.1	57.6
2003	74.1	50.0	90.8	57.6

Another way to examine the change in scores is to calculate the percentage change in average NCE scores from 1998 to 2003. The last row of each section of Panel A gives these figures for each group and each measure. The percentage change in average scores was much stronger, often more than twice as great, for students in the NIFDI-supported schools than for students in the other schools. For instance, with the composite achievement measure, average normal curve equivalent scores for students in the NIFDI-supported schools increased by 113 percent. The comparable figures for students in the other groups ranged from 46 percent to 56 percent. (These data were used to create Figure 9 in the text.)

Finally, Panel C of Table A-11 translates the Normal Curve Equivalent scores in Panel A into the corresponding percentiles. (These percentiles were used to create Figures 6 through 8 in the text.) The percentile scores confirm the pattern that appears in the other analyses. In each of the four groups the average student had substantially higher scores in 2003 than in 1998, but the increases were strongest for students in the NIFDI-supported schools. For instance, on the composite measure of reading achievement the average student in the NIFDI-supported school scored at the 17th percentile, but by 2003 the average student in these schools scored at the 74th percentile. The average student in the high poverty Other DI schools scored at the 21st percentile in 1998 but at the 50th percentile by 2003. The average student in the Control schools had similar gains, from the 27th percentile in 1998 to the 58th percentile in 2003. An average student at the low poverty Other DI school was at the national mean in 1998, but by 2003 was at the 91st percentile. In general, the raw, unadjusted scores confirm the more efficient and parsimonious results obtained with the mixed models, which incorporate the statistical controls for school context.

Research Question 5: What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?

The fifth and final question asked, "What was the magnitude of the effects that were found?" Cohen's d, a standard measure of effect size, was calculated to provide a descriptive measure of the extent to which students in the NIFDI-supported schools had achievement scores that were superior to students in the other schools. Cohen's d is simply the difference between the average score of two groups divided by the common standard deviation. Traditionally, d scores of .25 or larger have been deemed educationally significant.

Table A-12 gives the d scores comparing average achievement of students in NIFDI-supported schools with those in Other DI and Control schools as well as the comparison of students in Other DI schools with students in Control Schools**.** Effect sizes are given in Panel A for the unadjusted scores and in Panel B for scores calculated with school context controlled. With the unadjusted scores combined over all years, the effect sizes comparing students in NIFDI-supported schools with other students are positive, but below the .25 threshold. The effect sizes for 1998 are negative, indicating that the students in the NIFDI schools had substantially lower achievement than students in the other schools at the beginning of the implementation. However, by 2003, the d scores comparing students in the NIFDI schools with the other two groups surpass the .25 threshold for all measures of achievement. In other words, while the d score at the beginning of the study period indicates an educationally significant disadvantage for students in the NIFDI group, this had totally

reversed by the end of the study period and the students in the NIFDI-supported schools had an educationally significant advantage.

Similar, but even stronger, results occur with the measures derived from the results that are independent of school context (see Panel B). The effect sizes for students in schools that had experienced NIFDI-supported intervention far surpass the threshold of educational significance by the 7th year of implementation, ranging from .50 to .71 for comparisons to the Other DI schools and .72 to .85 for comparisons with the Control Schools. In contrast, the effect sizes involving comparisons of the Other DI schools and the Control Schools never reach the .25 level of educationally meaningful.

Thus, the results with the effect sizes parallel the results with the other statistical analyses. Students in NIFDI-supported schools had higher achievement than students in both the Other DI schools and in the Control Schools, and the magnitude of this advantage increased substantially over the years of implementation.

Table A-12: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Reading Achievement, BCPSS, 1998-2003

Panel A: Gross Effect Sizes

		Gross		
		All years	1998	2003
Vocabulary	NIFDI vs. Other DI	0.01	-0.33	0.28
	NIFDI vs. Control	0.08	-0.36	0.39
	Other DI vs. Control	0.07	-0.03	0.11
Comprehension	NIFDI vs. Other DI	0.07	-0.13	0.27
	NIFDI vs. Control	0.20	-0.16	0.39
	Other DI vs. Control	0.13	-0.04	0.12
Composite	NIFDI vs. Other DI	0.06	-0.22	0.29
	NIFDI vs. Control	0.17	-0.26	0.41
	Other DI vs. Control	0.11	-0.03	0.12

Panel B: Effect Sizes Independent of School Context

		All Years	1 year	7 years
Vocabulary	NIFDI vs. Other DI	0.13	-0.08	0.71
	NIFDI vs. Control	0.19	-0.09	0.85
	Other DI vs. Control	0.07	-0.01	0.14
Comprehension	NIFDI vs. Other DI	0.06	-0.08	0.50
	NIFDI vs. Control	0.19	-0.03	0.72
	Other DI vs. Control	0.13	0.04	0.22

Composite	NIFDI vs. Other DI	0.06	-0.08	0.63
	NIFDI vs. Control	0.17	-0.05	0.82
	Other DI vs. Control	0.11	0.02	0.19

Note: Effect sizes for the total "gross" figures were calculated from the raw achievement scores in Table A-5. Those for the "gross figures for the control schools and the NIFDI-supported schools for 1998 and 2003 are in Table A- . Figures for Other DI schools include Roland Park and the other schools. For the composite scores the means were 35.8 for 1998 and 57.1 for 2003, for vocabulary 37.8 for 1998 and 52.4 for 2003, and for comprehension 34.3 for 198 and 57.4 for 2003. Those for the effects "independent of school context" were calculated from the scores in Tables A-9 and A-10. The common standard deviation of 24.0 was used for all calculations.

References

MacIver and Kempner. 2002. The impact of Direct Instruction on elementary students' reading achievement in an urban school district. *Journal of Education for Students Placed at Risk* 7: 197-220.

Fashola, Olatokunbo S. and Robert E. Slavin. 1997. Promising programs for elementary and middle schools: Evidence of effectiveness and replicability. *Journal of Education for Students Placed at Risk* 2: 251-307.

Stockard, Jean and Maralee Mayberry. 1992. *Effective Educational Environments.* Newbury Park, California: Corwin/Sage, 1992.

From: WhatWorks
Sent: Friday, February 08, 2008 5:45 PM
To: 'jeans@uoregon.edu'
Subject: re: Questions on What Works Clearing House Procedures
Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated December 2, 2007 below.

The first question you raised concerned the cut-off year for studies that are reviewed by the WWC. To maintain a focus on current research, our reviews examine research conducted in the past 20 years. This cut-off date of 20 years is the standard for every topic reviewed by the WWC. The review for Early Childhood Education (ECE), under which Direct Instruction falls, marked its cut-off date at 1985 (the review began in 2005).

The ECE protocol states that "if sufficient time and resources remain after we have completed our review of research on interventions implemented post 1985, the ECE team will consider reviewing older research on curricula that are still in widespread use." This determination has not been made to date.

In terms of the age/grade range for the review of Direct Instruction research, the ECE team determined the parameters for the review. For further details on inclusion criteria please see the ECE protocol at http://ies.ed.gov/ncee/wwc/reports/early_ed/index.asp

In response to your third question about the qualifications of WWC topic review team members, each team is comprised of several trained professionals, each playing an integral part in the review process. The Principal Investigator (PI) for each topic is a well-known expert in his/her field and is responsible for leadership in conceptualizing the specific topic area, identifying and addressing issues during the review, and developing and reviewing topic and intervention reports developed for the topic. Leadership includes overseeing the quality in the production of the reports and making decisions, based on methodological and substantive expertise, that are not otherwise covered in the WWC protocols and procedures for report production.

The Project Coordinator (PC) is an established education researcher with relevant methodological and substantive expertise. The coordinator oversees the work of the WWC Review Team; manages that specific review; reviews research ratings; and writes and revises the work plan, protocol, and draft and final reports in collaboration with the PI.

Individual reviewers, who prepare initial summaries of studies for the WWC, are professional researchers with experience in research design and methodology. These reviewers undergo a rigorous training and certification process before conducting WWC reviews. .For more information about the staff that make up the WWC team, visit http://ies.ed.gov/ncee/wwc/overview/index.asp.

The WWC recently created a Quality Review Team to respond to concerns raised by study authors, curriculum developers or other relevant parties about WWC reviews published on our website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. If a quality review concludes that the original review was flawed, a revision will be published. These quality reviews are one of tools used to ensure that the standards established by the Institute of Educational Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.

If you have concerns about a published WWC review that you think warrant a quality review, please send those concerns to our Help Desk at info@whatworks.ed.gov. Please identify the study in question, the specific issue(s) that you think were handled incorrectly, and where relevant, explain what you think is the correct interpretation of the study.

I hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: Jean Stockard [mailto:jeans@uoregon.edu]

Sent: Sunday, December 02, 2007 12:30 PM

To: Mark Dynarski

Subject: Questions on What Works Clearing House Procedures

Dear Dr. Dynarski:

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction (NIFDI). One of my first tasks was to read as much of the background research in the area as I could. Along the way I ran into the "What Works Clearinghouse." In my review I developed several questions that I am hoping you can answer.

First, I noticed that the WWC has listed relatively few of the studies that deal with Direct Instruction. While a few were rejected because they didn't meet some specific standards of design, it appears that many others were rejected because they were published before

1985. I have not been able to find a scientific justification for selecting this date. No studies were cited that suggested children's learning styles had altered after that time or that schools' organizational structures or teachers' actions had changed. In fact, the only research article that I found to justify the methods was actually published in 1974 – 10 years prior to the beginning of your targeted dates. This article dealt with phonological awareness, and had nothing to do with systematic literature reviews and, of course, included nothing to suggest that processes related to phonological awareness have changed over time.

In all the other areas with which I am familiar we use the full range of research articles and, unless there are clear reasons established through the research literature, do not set an arbitrary date for inclusion or exclusion of results. The choice of this date seems to have particularly affected the DI literature because a great deal of the pioneering work was published before the cut-off point. Because I'm new to this particular work, I wondered if you could provide me with any explanation based in the research literature for your choice of this cut-off year. I did find a suggestion that, "if sufficient time and resources remain," studies published before 1985 might be reviewed. Do you know if this step will be taken? From my reading of the literature I believe that your limit on dates of studies may have seriously limited the selection of effective resources that educators will find on your website; and I know that this is a primary aim of your project.

Second, I noticed that a number of studies were rejected because they included children outside the K-3 grade range. I found this somewhat unusual, for the children that are most in need of effective instruction are those who have not learned to read by older ages. Because the effectiveness of DI has been so well demonstrated with regular students at the K-3 grade range, the more recent focus on other ages seems only greater evidence of its worth. This is especially so in the case of special education students and those with behavioral and other issues. Thus, I couldn't understand the rationale of omitting all the studies that targeted these most needy students. Any explanation you might have of this decision would also be greatly appreciated.

Third, and perhaps most important, as I read studies more closely I found that the Clearinghouse's interpretation of some of the studies didn't seem to accurately reflect the actual content of the reports. I know that these reviews are often done by staff members or graduate students who might not be familiar with all of the intricacies of social research. Thus, I assume that we can send corrections and they will be added.

I look forward to hearing from you, either through e-mail (jeans@uoregon.edu) or by phone (541-346-5005). I'm sure that we share the ultimate goal of helping all students – no matter what their background – develop academic skills and appreciate your appreciate your help.

Sincerely,
Jean Stockard, Professor Emerita

MATHEMATICA
Policy Research, Inc.

Paul T. Decker
President and Chief Executive Officer

P.O. Box 2393
Princeton, NJ 08543-2393
Telephone (609) 799-3535
Fax (609) 799-0005
www.mathematica-mpr.com

September 8, 2008

Dr. Jean Stockard, Ph.D.
Director of Research
National Institute of Direct Instruction
P.O. Box 11248
Eugene, OR 97440

Dear Dr. Stockard:

Thank you for your June 25 letter concerning the What Works Clearinghouse (WWC). Mathematica Policy Research wants to ensure the quality and accuracy of all information contained in the WWC, and it works with study authors, curriculum developers, and consumers of the WWC to correct inaccurate information. We therefore appreciate your interest in and feedback on the WWC.

The WWC was designed to establish and apply a set of standards to identify rigorous research related to the impacts of education interventions. Studies that meet these standards are considered to have designs with causal validity, which gives practitioners increased confidence that the results of the study actually reflect the true impact of the intervention being examined. These standards were developed by leading experts in education research methodology, and they are applied through a systematic review process that includes repeated checks for quality and accuracy. They represent research criteria that we believe to be vital to the purpose of the WWC.

I have closely reviewed the concerns in your letter. I have also asked for input from Mark Dynarski, the director of the WWC effort, and representatives from the Institute for Education Sciences at the U.S. Department of Education. Based on my review and the input I received, I am convinced that the standards being used by the WWC adequately reflect the original mission of the WWC, as outlined above, and that those standards are being applied appropriately and consistently. In the attached document, Mark Dynarski provides more detailed responses to the issues you raise concerning the WWC's review of Direct Instruction.

Your letter makes several suggestions for the WWC moving forward. Some suggestions are counter to the systematic review process that is the foundation of the WWC. For example, as explained in the attachment, the criteria for date and age ranges being reviewed are established with specific rationales, and the systematic review process gives equal weight to peer-reviewed and non-

peer-reviewed sources. As we expand the WWC's activities into new topic areas, we will be mindful of the suggestions you raised in your letter.

I hope this information clarifies the WWC mission and the way in which the WWC operations support that mission.

Sincerely,

(b)(6)

Paul T. Decker

RESPONSES TO CONCERNS IN 6/25/08 LETTER FROM JEAN STOCKARD

Contrasting conclusions between WWC and extant literature

The letter states that WWC conclusions differ from extant literature including meta-analyses and literature reviews. It is important to note that the WWC is designed to produce a systematic review of literature. A sound definition of a "systematic review" is in the recent publication *Knowing What Works in Health* Care by the Institute of Medicine (IOM):[1] *A systematic review is a scientific investigation that focuses on a specific question and uses explicit, preplanned scientific methods to identify, select, assess, and summarize similar but separate studies.* (pg. 82)

Consistent with the recommendations in the IOM report, the WWC applies evidence-based methodological standards consistently to each study it reviews. These standards, which are available on the WWC website, were developed by leading education research methodologists.

The systematic review conducted by the WWC goes beyond the procedures performed in meta-analyses. Again using the IOM definition, a meta-analysis "quantitatively combines the results of similar studies in an attempt to allow inference from the sample of studies included to the population of interest" (pg. 82). However, as Robert Slavin has noted, meta-analyses rarely describe even one study in any detail.[2] The WWC uses meta-analytic techniques to summarize the results of studies that meet WWC standards. However, the important distinction is that the WWC uses a rigorous set of standards, applied consistently, to determine *which* studies are included in the meta-analytic computations. This ensures that WWC summary measures are based only on studies with causal validity.

It is not surprising that WWC findings may differ from those of other analyses. Some meta-analyses may not have such rigorous standards for including studies; others may have standards that differ from those used by the WWC. Because of these differences, the WWC cannot judge the results of its systematic reviews by how they compare to other analyses. Rather, it works to ensure that its standards appropriately identify studies with strong causal validity and applies those standards consistently to each study reviewed.

---

[1] Institute of Medicine (IOM). *Knowing What Works in Health Care: A Roadmap for the Nation.* Washington, DC: The National Academies Press. 2008.

[2] Slavin, R.E. (1995). Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, Vol.48, No.1 pp.9-18.

Inclusion and exclusion procedures

Limiting studies to 1985 or later

The WWC's default time period for reviews is a study publication date of 1985 or later. This timeframe, which was established in 2005, is used for two reasons. First, by limiting reviews to research to this time period, WWC reviews reflect reasonably current research. In particular, the time period range ensures that effect sizes and improvement indices are based on a counterfactual condition that reflects classrooms as they operate within a recent time period. Second, the timeframe ensures that the research reviewed is examining versions of interventions that are most likely to be available to practitioners today.

WWC principal investigators have the option to expand the period for which studies can be reviewed, if they believe that important research will be excluded. The principal investigators for the Beginning Reading area chose to maintain the default period in large part to maintain currency with the classroom context for beginning readers. For instance, the fact that preschool enrollment has increased,[1] combined with the fact that more preschool and kindergarten programs run full-day,[2] means that students in the early grades may be better prepared to receive reading instruction today than students 25 years ago. Moreover, it is possible that any changes in reading readiness over this period have not been evenly distributed, since differences in reading ability by socioeconomic status and race are apparent at the kindergarten level.[3] Other contextual factors have changed over the past 20 years, including advances in teacher training, increases in home literacy activities, and changes in the content of and variety of curricula used in classrooms.

Any of these changes could have implications for the effectiveness of an intervention. If school readiness has increased, than an intervention that was effective 25 years ago may not be effective in more recent years. If teachers are receiving stronger training and using newer curricula, the counterfactual condition against which interventions were measured 25 years ago have changed, and possibly with it the magnitude of its effects. The Beginning Reading principal investigators judged that they had an inadequate basis for

---

[1] The proportion of 3 and 4 year olds enrolled in school increased from 37 percent in 1980 to 56 percent in 2006. See Snyder, T.D., Dillow, S.A., and Hoffman, C.M. *Digest of Education Statistics 2007*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2008.

[2] The proportion of pre-kindergartners and kindergartners attending full-day pre-kindergartners or kindergarten increased from 32 percent in 1980 to 59 percent in 2006. See Snyder, Thomas D. *Mini-Digest of Education Statistics, 2007*. Washington, DC: National Center for Education Statistics, Institute of Educational Sciences, U.S. Department of Education, 2008.

[3] For example, the Early Childhood Longitudinal Study of Kindergarteners (ECLS-K) found gaps in the reading knowledge and skills of kindergarteners by race: black and Hispanic children scored just under half of a standard deviation below whites on a test of reading knowledge and skills. Analysis of the ECLS-K and other surveys (i.e., Children of the National Longitudinal Study of Youth and the Infant Health and Development Program) show that socioeconomic status accounts for about half of the standard deviation of racial differences in reading test scores. See Duncan, G. and K. Magnuson. "Can Family Socioeconomic Resources Account for Racial and Ethnic Test Score Gaps?" *Future of Children*, Vol. 15, No. 1, Spring 2005.

assuming that effects of interventions measured more than 20 years ago would be experienced if schools adopted those interventions today.

Restrictive grade range for Beginning Reading studies

The letter raised concerns that the WWC's exclusion of research conducted on children outside of the kindergarten to 3rd grade age range was too restrictive. Because the reviews are focused on assessing interventions for beginning reading, the principal investigators, in coordination with the Institute of Education Sciences (IES), concluded that the review should focus on intervention effects for children in kindergarten through third grade (roughly ages 5 to 8). Some studies examine effects of interventions for students within the specified grade range and also students in higher grades. These studies are included in the review when the WWC is able to isolate the effects for the students who fall within the Beginning Reading grade range. The studies are excluded from the review if the results cannot be disaggregated to isolate effects for relevant grade range.

The grade range criterion is important for the integrity of the review process. The WWC-computed improvement indices and effect sizes are intended to reflect the effect of the intervention on the population in question. Including students above the topic area age range could lead to misstatements about intervention effects on children within the grade range.

It should be noted that the WWC attempts to determine a study's effects for the relevant grade range. When authors present findings aggregated across a broader age range and indicate that findings were analyzed for the relevant age range, it is standard WWC procedure to contact authors to request findings disaggregated for the grade range. Sometimes design limitations or other factors preclude authors from providing disaggregated results. In such cases, the WWC excludes the study.

Fidelity of treatment implementation

The letter notes that the WWC review process may downplay implementation fidelity. Definitions of implementation fidelity vary and many studies include little information to gauge fidelity, especially information about whether an intervention has been implemented within normal operating regimes of districts, schools, and teachers, not under specialized laboratory conditions. Moreover, there is no standard metric with which to rate and assess fidelity across studies that assures comparability.

The WWC's approach emphasizes the importance of replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies. Intervention reports include an "extent of evidence" classification that allows practitioners to place more weight if they choose on interventions for which the extent of evidence is large, meaning the results are drawn from multiple studies and a large number of classrooms and students.

Concerns about interventions and studies reviewed by the WWC

Reading Recovery

The letter expressed concern that Reading Recovery is an intervention outside of the Beginning Reading protocol. The Beginning Reading protocol states that interventions that target specific populations (for example, readers below grade level, and at-risk students) are eligible for the review. Reading Recovery is a short-term tutoring intervention program intended to serve the lowest achieving first-grade students (i.e., those in the bottom 20 percent). As such, it falls within the Beginning Reading protocol.

The letter expressed concerns that the reviews of studies of Reading Recovery mischaracterized the findings from those studies. For both the Baenen et al. (1997) and Iversen and Tunmer (1993) studies, the results presented by the WWC review represent the findings when the WWC standards and procedures are applied to these studies.

With respect to the Baenen et al. study, it is important to note that the beginning reading protocol prioritized one-year results. In effect, the Beginning Reading review is only intended to examine whether beginning reading interventions have an effect within one year. This one-year period is applied consistently to each study reviewed to ensure the results can be compared across studies and interventions. The Baenen et al. study's two- and three-year general reading achievement measures are not ignored, however; they are presented in Appendix A4.4 of the Reading Recovery Technical Appendices (http://ies.ed.gov/ncee/wwc/pdf/techappendix01_209.pdf).

With respect to the Iversen and Tunmer study, and consistent with the protocol, the WWC examined the results most relevant to the question of whether Reading Recovery improves reading proficiency compared to a reasonable counterfactual. That the study examined other comparisons is not ignored however. Appendices A4.1, A4.2 and A4.3 present results from other comparison groups. As with any study it reviews, the WWC does not base the findings of its review on the conclusions drawn by the authors.

We disagree that studies that compare an intervention to a no-treatment condition (as was done in Baenen et al.) provides a "built-in advantage," as the letter suggests. Many practitioners are interested in knowing whether an intervention is effective relative to customary classroom practices. The WWC reviews studies comparing treatments to no-treatment as well as studies comparing one treatment to another. In each case, the counterfactual is clearly documented in the review.

Exclusion of Reading Mastery Program

The letter expresses concern that the Reading Mastery Program is excluded from review by the WWC. The WWC has reviewed studies of Reading Mastery. A report summarizing the results of those reviews was published on the WWC website on August 12, 2008.

Inclusion of unpublished manuscripts for review

The letter suggests that by reviewing unpublished manuscripts, the WWC reviews studies of lower quality. Reviewing only studies that have passed a peer reviewing process could neglect unpublished information that may contain important and different findings. Mark Lipsey and David Wilson caution against this type of publication bias in *Practical Meta-Analysis* (2001). The damage such publication bias can cause in a systematic review is substantial. In *Knowing What Works in Healthcare*, the IOM notes that publication bias is well-established and that systematic reviews need to take steps to counter it, because otherwise "harmful interventions may appear to be worthwhile and beneficial interventions may appear to be useless" (page 97). For this reason WWC procedures are explicitly designed to review evidence from sources other than those published in journals. However, regardless of publication status, studies must meet the same evidence standards.

The letter asserts that the WWC should make available to the public all the research it reviewed. The more important issue is that the WWC provides an explicit citation to the study and thereby enables readers to obtain studies they wish to review. Unpublished information obtained from authors in response to queries that arise in reviews also are made available to requestors, and authors sign a form indicating that the information they provide to the WWC is available to the public.

Exclusion of Carlson and Francis (2002) study

The letter expressed concern over the WWC review's conclusion that there was a confound with the Direct Instruction intervention in Carlson and Francis (2002).[1] The WWC standard regarding intervention confounds was established to ensure that the results in a WWC review reflect what educators can expect if they implement the intervention being reviewed. A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program.

Exclusion of Waldron-Soler et al. (2002) study

As described in the protocols for the WWC Early Childhood Education and Beginning Reading reviews, to establish baseline equivalence, treatment and comparison groups must differ on the pretest measures by less than half a standard deviation, or the differences must be insignificant in an adequately powered statistical test. The Waldron-Soler (2002) study reported pretest differences exceeding half a standard deviation on at least two measures.

---

[1] The letter references Carlson and Freeman (2002). We have no record of a Carlson and Freeman (2002) study, and have assumed the letter refers to Carlson, C.D., & Francis, D.J. (2002). "Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE)." *Journal of Direct Instruction*, 3(1), 29-50.

The same standards were applied to the three Success for All studies you cite (Dianda and Flaherty, 1995; Ross, Alberg and McNelis, 1997; and Ross, McNelis, Lewis and Loomis, 1998). Unlike the Waldron-Soler study, pretest differences between the treatment and comparison groups were less than one half of a standard deviation.

Exclusion of Tobin (2003, 2004) studies

The Tobin (2003) study indicated that treatment and comparison groups were substantively different on pretest measures, with the differences exceeding half a standard deviation. Because the Tobin (2004) study follows the same students for additional time, it did not meet standards for the same reason.

The same standards were applied to the Success for All study cited in the letter (Smith, Ross, Faulks et al., 1993). Unlike the Tobin studies, pretest differences between the treatment and comparison groups were not statistically significantly different, and they were less than one half of a standard deviation. In keeping with WWC procedures, the WWC obtained information on pretest differences through communication with the study authors.

The "Create your own summary" feature

The letter expresses concern about the "create your own summary" feature of the WWC website because the results sort interventions by the magnitude of the improvement index. Because the improvement index calculations accurately reflect the application of the WWC evidence standards and effect size computations to the studies of those interventions, the WWC believes it is informative and useful to practitioners to sort results by improvement index.

The improvement index is one summary measure of the effect of beginning reading interventions. Users can also sort interventions alphabetically (by intervention name), by evidence rating, and by extent of evidence indices.

The What Works Clearinghouse Beginning Reading Reports and Rating of Reading Mastery: An Evaluation and Comment

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

September 2008

Technical Report 2008-04



© 2008 National Institute for Direct Instruction
PO Box 11248 • Eugene, OR 97440
1-877-485-1973

The What Works Clearinghouse Beginning Reading Reports and Rating of Reading Mastery: An Evaluation and Comment

NIFDI Technical Report 2008-4
September, 2008
Executive Summary

The What Works Clearinghouse is a federally funded program established in 2002 that evaluates educational interventions on the basis of the "rigor of research evidence" and provides summary ratings on its website. Ratings have appeared only since 2007, and some of the work of the Clearinghouse has already been subjected to harsh criticism. This paper adds to the body of critiques.

Direct Instruction's Reading Mastery curriculum is a comprehensive core reading program that has been found, in decades of research, to be highly effective. Yet, the WWC's recent review of Reading Mastery concluded that no studies met their evidence standards and they were "unable to draw any conclusions based on research" regarding its effectiveness.

A careful examination of the research cited by the WWC found that this conclusion was not well founded. The WWC failed to examine close to 100 research studies that were cited in other well-known reviews of the literature. In addition, reviews of studies that were included were flawed. For instance, one large, well-regarded study was rejected because training of teachers to use the curriculum was considered a "confound" to the use of the curriculum itself. Other studies were dismissed because the treatment (Reading Mastery) and control groups were unequal at baseline. Yet, these differences were statistically controlled and, in two cases, the treatment groups had lower pretest scores but equal and higher posttest scores.

Examination of reviews of other curricula indicates that at least some of the material posted on the website is misleading. Summaries do not reflect the conclusions of the articles and, in at least one case, are exactly opposite to the published conclusions.

Attempts have been made to communicate these concerns and recommendations for change to the WWC and Mathematica Policy Research, Inc., which administers the Clearinghouse. These concerns and suggestions, which parallel those of other researchers, have been rejected.

Given the very serious problems in the review procedures and policies, the NIFDI research staff suggests extreme caution in the use of any ratings or conclusions from the WWC.

Table of Contents

	Page
Executive Summary	2
The What Works Clearinghouse	4
Direct Instruction, Reading Mastery, and the WWC	5
The WWC Ignored a Significant Body of Research on Reading Mastery	6
The WWC Failed to Review Many Existing Studies of Reading Mastery	8
Inaccuracies in the Reviews of Reading Mastery Studies	10
Summary	12
Inaccuracies in Other WWC Reports	12
Recommendations	14
References	16
Appendix A: Communications with the What Works Clearinghouse and Mathematica	19
Appendix B: A Partial List of Studies of Reading Mastery and its Precursors Completed Before 1985	
Appendix C: Studies of Reading Mastery published 1985 or Later But Not Included In the WWC Review	
Appendix D: The Impact on Summary Ratings of Ignoring Fidelity of Implementation	

The What Works Clearinghouse Beginning Reading Reports and Rating of Reading Mastery: An Evaluation and Comment

This document discusses serious concerns with the policies, procedures and judgments of the federally funded What Works Clearinghouse, focusing both on a recent Clearinghouse review of the Direct Instruction program Reading Mastery as well as more general issues regarding their procedures and conclusions. The report begins with a brief description of the stated mission of the WWC and criticisms of their procedures. The second section provides a detailed analysis of the WWC's recent report on Reading Mastery, describing the extensive extant body of scientific research on the program and serious errors within the WWC report. The third section describes more general concerns with the work of the WWC and misrepresentations of research on one other curricular program, and the final section provides recommendations for changes in WWC procedures. Appendices provide supporting material.

The What Works Clearinghouse

The What Works Clearinghouse (WWC) was established in 2002 through a grant from the U.S. Department of Education's Institute of Education Sciences. Billing itself as "a central and trusted source of scientific evidence for what works in education," the Clearinghouse reviews studies of educational interventions, assessing the "rigor of research evidence" and "giving educators the tools to make informed decisions" (http://ies.ed.gov/ncee/wwc/aboutus/). The WWC has focused its work on analyzing individual curricula, giving its highest rankings to generally small, but tightly controlled, experimental studies on the impact of different curricula.[1]

After a very slow start, the Clearinghouse began issuing reviews in 2007, periodically posting summary judgments of individual curricular programs on its website. The assessments appear to be internally developed by WWC staff with no external oversight of the process or results. Unlike most of the scientific literature, the assessments are not exposed to peer review or other independent judgment by experts in the field. Not surprisingly, the work of the WWC has faced extensive criticism for its procedures and methodology including the ways in which studies are identified, the manner in which summary rankings are constructed, the accuracy of the reviews that have been produced, and the absence of any outside or independent oversight (e.g. McArthur 2008, Slavin, 2008).

The author of this document queried the WWC regarding its practices in e-mail communications from December, 2007 through April, 2008. Based on the responses received to those queries and further examination of WWC materials, a detailed letter was sent to Mathematica, the organization responsible for administering the Clearinginghouse on June 25, 2008. This letter summarized concerns with the WWC's procedures and policies. A reply was received from Mathematica on September 8, with a cover letter from the CEO of Mathematica and a more detailed report, apparently prepared by Mark Dynarski, the project director of the

[1] This focus is somewhat surprising given the emphasis within the educational community and research literature on the importance of comprehensive changes within schools to produce long-lasting changes in achievement. In addition, the WWC's "Statement of work" claims that the goal of the clearinghouse was to develop reviews of "educational interventions," parenthetically defined as "programs, products, practices, and policies" (WWC Statement of Work, Section I.B., paragraph 1).

WWC. All of these communications are included in Appendix A and are occasionally referenced in this document.

Unfortunately, none of the information that has been provided by the WWC or Mathematica has eased the concerns with the WWC's procedures that were outlined in the original queries. The material below summarizes some of these concerns and indicates why the conclusions of the WWC, both with respect to Reading Mastery and to other instructional programs, should be seriously questioned.

Direct Instruction, Reading Mastery and the WWC

Numerous studies have demonstrated the effectiveness of Direct Instruction with a wide variety of students in very different settings. Meta-analyses that examine large bodies of research literature also consistently point to the superiority of Direct Instruction. For instance, a recent meta-analysis of the achievement effects of different models (Borman et al 2003) found that Direct Instruction was the most researched of any of those studied and had the most research conducted by those not affiliated with the developers. As they put it, "the research base for Direct Instruction is very extensive and of very good quality" (p. 187). Among the many different models used in their analysis, Direct Instruction had the strongest evidence of effects on student achievement. Numerous other analyses have reached similar conclusions (e.g. Adams and Engelmann 1996, AFT 1998, Beck and McCaslin 1978, Herman, et al 1999, Kennedy 1978, Schieffer et al 2002). Additional studies expound upon the importance of the principles that underlie Direct Instruction, especially systematic and explicit instruction (see Adams 1990, Anderson et al 1985, Baker 1994, Bond and Dykstra 1967, Chall 1967, Foorman 1995, Fukkink and deGlopper 1998, Grossen 1997, ICHHD 2000, Juel and Minden-Cupp 2000, National Reading Panel 2000, NICHHD 1996, Pflaum, et al 1980, Smith et al 2001, Snider 1990, Snow et al 1998, Stanovich 1994). Finally, Direct Instruction has been lauded as not just the purveyor of effective curricular materials, but also as an important model of comprehensive school reform, guiding the transformation of schools into effective organizations in which all students can have high achievement (e.g. Borman 2002, Buechler 2002, CSRQ 2006).

Direct Instruction's Reading Mastery curriculum is a comprehensive core reading program with multiple levels and is the most widely used program within the DI curricular library. It has long been acclaimed as highly effective, with strong research accumulated over decades of work, supporting its development and documenting its success.

Despite the large body of research on Direct Instruction and the various associated curricular programs, Reading Mastery was not included in the initial WWC review process for Beginning Reading. Through mid-summer, 2008, the WWC website simply stated, without further explanation, that they chose not to include Reading Mastery in their review of Beginning Reading material.[2] In mid-August, 2008, however, the WWC issued a report on Reading Mastery with the following conclusion:

[2] In July, 2007, the WWC posted a review of Corrective Reading on its Beginning Reading page. Corrective Reading is a Direct Instruction program designed for children in the upper elementary grades who are reading below grade level. No explanation was given for why this program was included in the Beginning Reading category when its target audience did not match either the intended grade level or the general classroom focus of the review. While the review accepted one study of Corrective Reading as meeting its criteria and determined that it showed

> No studies of Reading Mastery that fell within the scope of the Beginning Reading review meet WWC evidence standards. The lack of studies meeting WWC evidence standards means that, at this time, the WWC is unable to draw any conclusions based on research about the effectiveness or ineffectiveness of Reading Mastery.[3]

To support their conclusion the WWC report lists 61 citations on Reading Mastery that were examined and a summary of the reasons they did not meet the WWC's standards of evidence.

Because the WWC conclusion regarding Reading Mastery contrasts so starkly with previous findings of the scientific community, the research staff of the National Institute for Direct Instruction (NIFDI) carefully examined the listed works and compared them to the extant literature and other reviews. As detailed below, our examination uncovered numerous examples of faulty judgments by the WWC. These involve three major points, all of which are elaborated below: 1) the WWC report ignored a large body of the extant research through seemingly arbitrary review decisions, 2) the WWC report did not find or review a significant proportion of the research on Reading Mastery that should have met their review criteria, and 3) the few reviews that were conducted often misconstrued and misinterpreted the research evidence.

Based on our review, others' critiques of the work of the WWC, and the very extensive research base that has documented the effectiveness of Direct Instruction and Reading Mastery, we conclude that the WWC statement regarding Reading Mastery is inaccurate and misleading. A large body of literature points to the very strong effect of Direct Instruction and Reading Mastery on children's reading achievement.

The WWC Ignored a Significant Body of Research on Reading Mastery

The WWC limited its review of studies for the Beginning Reading category to those that involve students in general education in grades K-3 and that were published between 1985 and 2007. When NIFDI's Director of Research questioned the decision to ignore studies published before 1985, the WWC responded that the "cut-off date of 20 years is a parameter set to ensure the research reviewed is most relevant to classrooms as they operate today" (e-mail correspondence, April 9, 2008, see Appendix A). The report included with the September 8 letter from Mathematica expands upon this reasoning by suggesting that numerous other changes over time such as increased levels of reading readiness (supposedly as a result of increased preschool enrollment), "stronger training" of teachers and the use of "newer curricula" "could have implications for the effectiveness of an intervention" (Dynarski, page 2). The exposition in the

---

"potentially positive effects," the issues raised in this document regarding WWC's review of Reading Mastery also apply to its review of Corrective Reading. One study of Reading Mastery was also included in the ELL portion of the WWC webpage, but again this placement does not reflect the target audience of Reading Mastery. It should be noted, however, that the high ratings these studies of both CR and RM received, even with these alternative audiences, should be seen as further evidence of the quality of the programs.

[3] The WWC Reading Mastery report, including the studies they reviewed, can be found at http://ies.ed.gov/ncee/wwc/pdf/WWC_ReadingMastery_081208.pdf. Interestingly, the September, 2008 letter from Mathematica (Dynarski report, page 4) notes the addition of the RM review to the website on August 12, 2008. It is unclear to what extent the review was prompted by our letter of June 25, 2008.

correspondence pointedly excludes any mention of empirical data to support the causal linkage that is suggested.[4]

We have not been able to find any studies cited on the WWC website, or elsewhere, that document that children's learning styles, classrooms, teachers' actions, or schools' organizational structures have drastically changed since 1985. As noted above, an extensive body of research cites the importance of systemic and explicit instruction in promoting achievement. This research spans the last 4 decades and has produced consistent results, no matter when the research was conducted (see Adams 1990, Anderson et al 1985, Baker 1994, Bond and Dykstra 1967, Chall 1967, Foorman 1995, Fukkink and deGlopper 1998, Grossen 1997, ICHHD 2000, Juel and Minden-Cupp 2000, National Reading Panel 2000, NICHHD 1996, Pflaum, et al 1980, Smith et al 2001, Snider 1990, Snow et al 1998, Stanovich 1994). In all the other areas of scientific study with which we are familiar, such as medicine, psychology, or the other social sciences, the full range of research articles are used to formulate policy, and there is no arbitrary date for inclusion or exclusion of results. If changes over time are hypothesized, the possibility of such changes is subjected to empirical test. Many analyses of the research literature systematically examine variations in results by the time when the study occurred.

This limitation of the time period for review had a disproportionately large impact on the number of studies of Direct Instruction included in the WWC's scope of study compared to other interventions precisely because DI has such a long and strong research base. A strong body of seminal research on DI was conducted before 1985, including Project Follow Through, the largest educational experiment in the history of the United States.[5] In general, research on interventions tends to be most active soon after the creation of the program. Because Reading Mastery was developed in the 1960s and high quality research conducted at that time demonstrated its effectiveness with a wide range of students, researchers have had little incentive to conduct additional studies. Given the consistently positive results from earlier decades, the current recent research focus has moved to populations of students with more extensive needs, such as older children who had not yet learned to read.[6]

A significant amount of research on Reading Mastery was ignored by the WWC because of the seemingly arbitrary choice of 1985 for a cut-off. Appendix B of this report includes a partial list of studies of Reading Mastery and its precursors that were completed before 1985.[7] The list was compiled by reviewing published literature reviews

---

[4] The citations provided are to macro-level data regarding increased preschool and kindergarten enrollment as well as to a study regarding the relationship of race and socio-economic status to achievement. None relate, either directly or indirectly, to variations over time in the impact of curricula or to changes over time in children's learning styles or achievement.

[5] Project Follow Through employed a strong experimental design and an extremely large sample to compare a wide variety of elementary curricula. The results provided very strong evidence of the superiority of Direct Instruction in raising academic achievement as well as students' self image.

[6] Studies of children in special education should have been within the purview of the review. In fact, the description of "populations to be included" in the WWC's Beginning Reading protocol states that "because students with learning disabilities…lag behind the population as a whole in reading achievement, studies involving these groups are of particular interest." No mention is made of excluding studies that involve students in special education.

[7] The cut-off date of 1985 used for the Reading Mastery report differs from the general WWC protocol for the Beginning Reading studies, which uses 1983 as the cut-off date. We have found no explanation for this discrepancy

and meta-analyses, all of which should have been available to the WWC staff. The list includes 38 items.[8] It should be stressed that this list is not exhaustive, but simply relied on easily available meta-analyses and reviews that should have been accessible to the WWC staff.

In short, unless the WWC can provide scientific evidence to the contrary, it would appear reasonable to suggest that high quality studies conducted before 1985 are still relevant today and should be included in any evaluation of Reading Mastery. Any conclusion regarding the effectiveness of Reading Mastery that does not consider this part of the literature should be challenged as incomplete and inaccurate.

The WWC Failed to Review Many Existing Studies of Reading Mastery

In addition to excluding studies published prior to 1985, the WWC review omitted a large number of studies published after 1985 that could have been included. Appendix C provides a partial list of these studies and includes 56 different citations. Like Appendix B, this list was largely generated by examining earlier reviews that should have been available to the WWC staff. Although the WWC lists a methodology for retrieving studies, the fact that they could miss so many citations that were in already published reviews suggests serious difficulties with their selection criteria.

As with the vast majority of work in this area, the studies listed in Appendix C generally found that Reading Mastery was superior to other approaches. For instance, an article by Kamps and associates (2003) uses multi-level analytic techniques to examine data from schools that implemented different reading curricula and found that students in Reading Mastery had significantly stronger achievement gains than in any of the other programs, both for those with and without social behavioral risks. Similarly, Gunn and associates (2000) report a study in which students were randomly assigned to interventions and found that both Hispanic and non-Hispanic students who received instruction in Reading Mastery had significantly more improvement over time than other students.

It should also be noted that the WWC list of writings on Reading Mastery includes material that would generally not be included in summaries of the scientific literature. Many items on the list are not research reports and were never intended as such. A large proportion is reports by school principals, teachers, and/or superintendents of the success their schools have experienced with Direct Instruction. These testimonials appeared in publications of the Association for Supervision and Curriculum Development and Council of Chief State School Officers (accounting for 8 of the items on the list of 61 studies) and on the website of SRA, the publisher of DI materials (26 of the items).[9] Two other studies on the list clearly targeted students outside the stated age range, with the titles indicating that they looked at "ninth and tenth graders" (Airhart 2005) or at "middle

[8] A large number of these studies examined DISTAR, the name by which the content now in Reading Mastery, was once known.

[9] Most of the testimonials include results from norm referenced achievement tests, often involving data from several years. These could have been used, if the WWC desired, to compare the achievement in these schools with national norms.

school students" (Shippen et al 2006). An additional study (Flores and Ganz 2007) did not look at Reading Mastery, but at Corrective Reading, a different Direct Instruction curricular program. Three studies in the list did not have reading achievement as a dependent measure, but instead were content analyses of textbook stories (Jordan 2005), a qualitative study of three students' understandings of what they read (Wilson 2005), and an examination of students' social behaviors (Smolkowski et al 2005). Finally, a few others appeared to have never been subjected to the usual scientific standards of peer review, such as articles posted on websites, papers given at professional meetings, and unpublished dissertations (6 studies in all).[10, 11]

When these 46 studies are omitted from the list, the WWC actually examined only 15 studies of Reading Mastery that would generally be considered part of the scientific research base. This number is less than one-fourth of the number given in the report. The listing of materials in the WWC's report on Reading Mastery implies that the research base that was reviewed was much larger than it, in fact, was.

In general, it appears that the WWC list of materials that were reviewed suffers from errors of both exclusion and inclusion. A large number of studies appear to have been ignored. Our review of already existing reviews of the literature revealed almost one hundred articles that could have been consulted if a full analysis were to occur. These citations were easily found through simply searching the reference list of already published literature reviews and thus are no doubt an undercount of the available resources. The WWC review missed a large proportion of the research base, not just through the time parameters that were used but also, apparently, through a search of the literature that was far from thorough.[12] At the same time, much of the material listed in their report should have fallen outside the domain of a comprehensive review of the

[10]Meta-analyses often include such unpublished works in their reviews, and that has been the case with several meta-analyses of beginning reading. Given that the WWC review is not subject to peer review, omitting these studies from their work could be one way to add additional controls. If the WWC's work were subject to peer review, including unpublished work would be less problematic. Dynarski's report included with the September, 2008, letter addresses the issue of peer review, citing a study of publication bias in medical research. Whether or not the result with medical research applies to educational research is an empirical question and a citation to the education or more general social science literature would be more appropriate. We know of no study that has found that high quality social science work remains unpublished. In any case, the standard practice in meta-analyses and other reviews is to treat the issue of the relationship of publication status to results as an empirical question and compare the unpublished and published articles.

[11] The WWC categorized one of the studies that had not been subjected to peer review (CSRQ 2006) as "ineligible…because it does not examine the effectiveness of an intervention." This study, prepared by the highly regarded Comprehensive School Reform Quality Center, reviewed and summarized the evidence on the effectiveness of 22 different school reform models, including Direct Instruction. As with other such reviews, there were more studies available regarding the effectiveness of Direct Instruction than for any other model and the results indicated that Direct Instruction was consistently one of the most effective, if not the most effective, in each of the categories examined. Thus, while the CSRQ document did not address the effectiveness of a single intervention, to dismiss it because it examined numerous models could be seen as misleading.

[12] Ironically, one of the literature reviews used to generate the listings in Appendices B and C was the 2006 report of the Comprehensive School Reform Quality Center (source e in the listings) and discussed in footnote 10 above. The WWC lists this report as one that they consulted and found "ineligible for review," but apparently did not use its references for further examination.

scientific literature, providing a misleading representation of the amount of literature that was examined.

Inaccuracies in the Reviews of Reading Mastery Studies

The WWC report provides one sentence summaries of the reason that a study was judged not to meet their established "evidence standards." These reasons involved either questions about the comparability of the control and experimental groups or the existence of a possible "confound" to the treatment. Examples are given below of the studies rejected for these reasons and explanations of the faulty logic involved in the conclusions. The errors in these examples are so serious that they should raise serious concerns regarding the WWC's evaluation procedures.

Equivalence of Experimental and Control Groups – Three of the published studies examined by the WWC that addressed student achievement were rejected because, although they used a "quasi-experimental design" with both an intervention and a control group, they did "not establish that the comparison group was comparable to the treatment group prior to the start of the intervention."[13] The NIFDI research staff carefully examined these three articles and found that, in each case, the conclusion of the WWC could be questioned. In one of the studies (Brent et al, 1986) the presentation of the results clearly noted that adjustments had been made, using standard multivariate procedures, for any possible pretest differences between the treatment and control conditions. In the two other studies (O'Brien and Ware 2002, Thomson 1991), the pretest scores of the treatment group (those receiving Reading Mastery) were lower than the scores of students in the control group. Such a situation is routinely considered as an indication of a conservative test, for it biases results against the treatment, and is not used as a reason to automatically discount the results. Again, supporting the large body of extant evidence, these three studies all showed positive effects of Reading Mastery on students' achievement.

Confounding Factors – Nine studies were rejected from the WWC's review because, although they used a quasi-experimental design, the reviewers determined that there was "a confounding factor, such as combining with other interventions, which makes it impossible to attribute the observed effect solely to Reading Mastery." Unfortunately, the WWC provided no specific details in the report published on the web regarding how they reached this conclusion for each study. Again, however, the review by the NIFDI Research staff suggests that this judgment may not be appropriate for at least some of the works.

For instance, one of the studies rejected in the Reading Mastery review for having a "confounding factor" was Carlson and Francis's (2002) evaluation of the RITE program. For some reason this study had mistakenly been included in WWC's earlier review of the Direct Instruction Corrective Reading program and had also been rejected in that review for a supposed confound. This rejection was questioned in the letter written to Mathematica on June 25, 2008, and the response to this query in the September 8 correspondence sheds greater light on the WWC's definition of a "confound." This explanation also, however, raises serious questions about the validity of the WWC review and its implications for educators.

[13] The WWC report lists 6 studies that were rejected for this reason, but three of these were unpublished works such as those described above.

The Carlson and Francis study involved thousands of students, hundreds of teachers, and dozens of schools and used sophisticated multivariate techniques to analyze the data. The only intervention in the experimental schools was Direct Instruction's Reading Mastery, and as with the other studies, the results strongly favored this curriculum. After scouring the article, the NIFDI research staff found one possible explanation for the WWC's determination. On page 143, the authors described the intervention: "In addition to the teaching of skills directly related to the RM curricula, the … program also strives to provide teachers with strong classroom management techniques" (p. 143).[14] The September 8, 2008, communication from Mathematica confirms this interpretation, with the statement that, "A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program" (Dynarski, p. 5).

In reality, strong classroom management is part and parcel of the Direct Instruction approach. Training teachers in such management is part of the in-service training that teachers receive in learning how to implement the curriculum as well as prominently included within the teacher's guide to the program. Part of the reason that DI is so successful is that it provides not just well designed curricular materials but well developed, research-based guidance on how the curriculum should be administered. Because these guidelines are well documented in the various guides to the programs, they should have been part of the knowledge base of a competent reviewer. The classroom management was not a confounding element of the intervention, but was an integral part of the curriculum and its appropriate delivery.

More importantly, virtually all curricular programs include elements of teacher training and discussions of classroom management. It is reasonable to argue that if studies of Reading Mastery that include training for teachers are to be excluded, all other studies that include training for teachers should also be excluded. We know of virtually no legitimate curriculum that does not include some type of instructional overview for teachers.

In addition, complete training in a program is vital to ensuring that it is implemented with fidelity. Disallowing such elements from a design could produce a very serious threat to internal validity of a study. Surprisingly, the WWC's approach seems to discount the importance of the fidelity of treatment implementation, suggesting that "there is no standard metric" with which to rate and assess fidelity" and that a better approach is their reliance on "replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies" (Dynarski 2008, p. 3).

At least two very serious problems are immediately apparent with these statements. First, those familiar with any number of structured learning and behavioral approaches know that there is in fact a well developed literature of fidelity measures and that such measurements are very important both in helping to train practitioners and to assess the extent to which programs are being adequately implemented. To suggest that such measures do not exist is to demonstrate an unfortunate unfamiliarity with the literature. Second, to rely on replications as a substitute for

[14] The WWC Reading Mastery bibliography lists the Carlson and Francis article as in the Journal of Direct Instruction. The article also appeared in the Journal of Education for Students Placed at Risk in 2002, volume 2, issue 2, pp. 141-166. The page number in the text refers to this citation.

measures of fidelity could well produce very misleading results. Numerous replications of a poorly implemented program can, in no stretch of the imagination, provide an adequate test of a program. To assume that poor implementations are so rare that their impact would "average" out is, at best, a contention that should be subject to empirical test. More realistically, relying on such an assumption would likely produce very inaccurate and misleading results, for the errors that result from ignoring fidelity are systematically biased. As detailed in Appendix D, including studies with poor fidelity within summary analyses produces higher ratings of poor programs and lower ratings of good programs, thus providing very misleading conclusions for consumers.

Finally, two other studies that were rejected for an alleged confound were those by Gunn and associates (2002, 2005). These works report the results of a study that included both a behavioral intervention and Reading Mastery. While it is understandable that a naïve reviewer might quickly reject these articles, a careful reading shows that the authors directly address the issue of any possible confound and, based on both their results and other works in the field, note that this was unlikely. Again, these articles, plus a third by the authors that was not included in the WWC listing, find results strongly in favor of Reading Mastery.[15]

Summary

To summarize, examination of the What Works Clearinghouse's review of Reading Mastery indicates a number of aspects that should lead consumers to question the accuracy of the report. Part of the problem involves the ways in which studies were selected for inclusion and exclusion from review. An arbitrary time frame resulted in a large number of studies being excluded. In addition, however, the Clearinghouse ignored an even larger number of studies published within the chosen time frame. It appears that the WWC made no attempt to examine any of the large and easily available meta-analyses and literature reviews and use the studies listed in those documents. In addition, the accuracy of the WWC report should be questioned because of the quality of the reviews. The reviews appear to be both inaccurate and misleading.

Inaccuracies in Other WWC Reports

As noted above, the letter of June 25, 2008 to Mathematica was written before the Reading Mastery review was posted. One issue discussed in the letter is of such concern that it deserves additional discussion in this comment: the misrepresentation of results for at least one program that was given high ratings.

The WWC has given high ratings to the Reading Recovery program, a short-term tutoring intervention. The WWC website concludes that "Reading Recovery® was found to have positive

[15] Of the remaining three peer reviewed studies examined by WWC, two were excluded from review because they did not include a comparison group, but relied on comparison of pretest and posttest scores for a single group of subjects (Humphries et al 2005 and Marchand-Martella et al 2006). A third (Kamps and Greenwood 2005) was excluded because it "does not examine the effectiveness of an intervention." Examination of this study indicates that, in fact, just the opposite is true – the study reports the results of the first few months of a randomized trial of interventions comparing several more highly structured curricula, including Reading Mastery, with less structured approaches. Although descriptive statistics indicate that the students exposed to Reading Mastery have had substantially higher rates of progress than students in the traditional settings, inferential statistics that examine differences between the settings were not included. Thus, while definitive judgments could not be made regarding the statistical significance of the results, to indicate that this study does not involve an intervention is, at best, misleading.

effects on students' alphabetics skills and general reading achievement outcomes and potentially positive effects on comprehension and fluency" (http://ies.ed.gov/ncee/wwc/reports/beginning_reading/reading_recovery/). A careful reading of the articles cited by the website, however, indicated that, in at least two cases, the WWC review did not accurately reflect the content of the research studies.

For instance, one article (Baenen, et al 1997) was cited as showing positive impacts of RR, yet even a cursory reading of the article indicates that the authors found remarkably little success, especially in the long-term, for students in the program. The article concludes (p. 176) that although about one-half of the students in the tutoring program had successfully reached first grade reading levels at the end of the year, success rates declined in subsequent years of implementation. In addition, by third grade, there was no difference in achievement scores, needs for retention, special education, or Chapter 1 assignments of students who had participated in Reading Recovery and other students. The authors concluded that Reading Recovery was very expensive to implement in relation to the benefits that it provided.

The seemingly erroneous conclusion of the WWC was brought to the attention of Mathematica in the June 25 letter. The response (Dynarski 2008, p. 4) acknowledged the results outlined above, but defended the WWC's conclusions by stating that their reports "prioritized one-year results" and that the findings regarding the results in later grades were included in a technical appendix. Yet, as teachers, parents and students would attest, how well one reads at the end of third grade is much more crucial in determining eventual success than how well one reads at the end of first grade. Learning to read is not a one-year process, but is a multi-year endeavor. If gains in first grade do not persist through third grade, the first grade gains have very little worth. Ignoring the very important conclusions of Baenen and associates would seem to do a disservice to parents and teachers.

An even more disturbing example involves an article by Iversen and Tunmer (1993), which was also cited by the WWC as supporting the conclusion that RR is effective. The major purpose of this study was to compare the standard RR program to a "modified" program that included explicit instruction in phonological skills. The major variable of interest to Iversen and Tunmer was how long children took to reach a level of competency where they could discontinue special tutoring, the major goal of a tutoring program such as RR. Students in both the unmodified Reading Recovery program and the modified program (including instruction in phonologically based elements) eventually caught up with the other children, but the students in the modified program were able to discontinue tutoring much earlier. The standard Reading Recovery program was found to be 37 percent less efficient than the modified program. In addition students in the modified program continued to have higher levels of achievement and higher rates of learning at the end of the school year. The authors provide an extensive discussion and additional analyses that demonstrate the fallacy involved in Reading Recovery about the ways in which word recognition skills develop. They clearly conclude that Reading Recovery is not an efficient method for teaching children to read and that phonological training is superior.

The WWC chose to ignore any results regarding this comparison group, which received phonological training and had superior achievement, "because it was a modified version of the

standard program." The September response to our query regarding their judgment stated, "the WWC examined the results most relevant to the question of whether Reading Recovery improves reading proficiency compared to a reasonable counterfactual." That counterfactual was, apparently, having no tutoring at all, not tutoring with a more effective program. The statement goes on to note that the results with the other comparison groups were mentioned in an appendix, implying that such information could be available for those who were interested.

Even though the WWC did acknowledge the negative results regarding Reading Recovery in a technical Appendix, the overall conclusion given on its website is certainly misleading and does not reflect the conclusions of the articles. The chance of a parent or school official accessing a technical appendix to find information that contradicts that given in the major pages of the web site is extremely remote. The actions of the WWC regarding the material on Reading Recovery not only resulted in misrepresenting the effectiveness of Reading Recovery but also that of phonologically based programs and provides further evidence that its conclusions should be seriously questioned.

Recommendations

Based on a careful review of the evidence, we suggest that consumers approach the ratings of the WWC with extreme skepticism. With respect to Reading Mastery the list of sources that were reviewed was extremely selective and tapped only a very small proportion of the extant literature. Perhaps even more disturbing, the reviews of the works that were selected were often inaccurate and misleading. We see no reason to accept their conclusion of "no evidence" as valid.

From our correspondence with the WWC there also appear to be a number of very serious policy and procedural issues that raise questions regarding the accuracy of any ratings that they have developed. As detailed above, their policies in a number of areas appear to differ substantially from traditional scientific practices. These include such areas as decisions about how to select studies to examine, classifying teacher training procedures as a confounding element, relying on replications to "average out" problems with fidelity of implementations, and focusing on one-year studies rather than multi-year results, even when such multi-year results are available and more indicative of student success.

Part of the reason that the WWC has reached erroneous conclusions may reflect the way in which its review process departs from well established traditions of scientific research. In our June, 2008, letter to Mathematica, we made several suggestions for changes to their procedures. These suggestions in many cases overlapped with those made by other critics of WWC (see Slavin 2008 and McArthur 2008). For instance, we suggested that using standard quality control mechanisms, such as two reviewers for every article with a third if results differ and a peer review process before publication of ratings, could go far to help ensure more accurate results. We suggested that the cumulative nature of the scientific enterprise should be acknowledged, and the WWC's results should be compared with the already well developed body of meta-analyses and literature reviews. When WWC's results differ from these reviews, they should investigate why and adjust the results as needed. In addition, metrics that are commonly used within the social sciences, such as measures of effect size would greatly enhance faith in the WWC's conclusions. Issues of both internal and external validity in decisions regarding acceptable

designs should be weighed. Much more accurate information could be provided if the WWC understood the wide variety of research approaches that can and must be used within real-life educational settings. Instead of simply rating studies' quality by the nature of the research design, elements related to sample size, statistical significance, substantive significance, and length and fidelity of intervention should be included, with global ratings reflecting the preponderance of evidence regarding interventions from all available data. Finally, it is crucial that the WWC ensure that reviewers are knowledgeable in the substantive areas that they are reviewing as well as in the methodological details.

Unfortunately, the concerns expressed in the correspondence with Mathematica and the WWC, as well as the concerns of earlier critics Slavin (2008) and McArthur (2008) appear to have been unheeded and, in fact, rejected. As stated in that letter, the recommendations were made "in the spirit of the intent behind the What Works Clearinghouse – a desire to provide schools and families with the most accurate information." The letter also noted that "the problems with the material posted on the website appear to be much more widespread than can be handled in a piece-meal fashion. To allow inaccurate material to remain does a disservice to schools, families and students."

The communications with Mathematica appeared before the review of Reading Mastery was posted. Unfortunately, the quality of that review has made the earlier judgments of problems with the WWC even more apparent and worthy of grave concern. If Mathematica refuses to remove the faulty information from the WWC website, consumers should be extremely wary of the ratings posted there and consult the standard social science literature for more accurate information.

References

Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.

Adams, Marilyn J. 1990. Beginning to read: Thinking and learning about print. Cambridge, Mass.: The MIT Press.

American Federation of Teachers. 1998. Building on the best, learning from what works: Seven promising reading and language arts programs. Washington, D.C.: AFT.

Anderson, R. C., E. H. Hiebert, J. A. Scott, and I. A. G. Wilkinson. 1985. Becoming a nation of readers: The report of the Commission on Reading. Washington, D.C.: National Institute of Education.

Baenen, Nancy, Alissa Bernholc, Chuck Dulaney, and Karen Banks. 1997. Reading Recovery: Long-Term Progress After Three Cohorts. Journal of Education for Students Placed at Risk 2: 161-181.

Baker, Scott K. Edward J. Kameenui, Deborah C. Simmons, and S. A. Stahl. 1994. Beginning reading: Educational tools for diverse learners. School Psychology Review 23: 372-391.

Beck, I.L. and E. S. McCaslin. 1978. An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. LRDC Report No. 1978/6 Pittsburgh.

Bond, G. and R. Dykstra. 1967. The cooperative research program in first-grade reading instruction. Reading Research Quarterly 2: 5-142.

Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.

Buechler, Mark. 2002. Catalog of School Reform Models: Program Report. Portland, Oregon: Northwest Educational Research Lab.

Chall, J.S. 1967. Learning to read: The great debate. New York: McGraw Hill.

Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.

Dynarski, Mark. 2008. "Responses to Concerns in 6/26/08 Letter from Jean Stockard," sent on September 8, 2008, from Mathematica, included in Appendix A of this report.

Foorman, B. R. 1995. Research on "the great debate": Code-oriented versus whole language approaches to reading instruction. School Psychology Review 24: 376-392.

Fukkink, R.G. and K. deGlopper. 1998. Effects of instruction in deriving word meaning from context: A meta-analysis. Review of Educational Research 68: 450-469.

Grossen, B. 1997. A synthesis of research on reading from the National Institute of Child Health and Human Development. Eugene, Oregon: University of Oregon.

Herman, R., D. Aladjam, P. McMahon, E. Masem, I. Mulligan, O. Smith, A. O'Malley, S. Quinones, A. Reeve, and D. Woodruff. 1999. *An educator's guide to schoolwide reform*. Washington, D.C.: American Institutes for Research.

Institute of Child Health and Human Development (ICHHD). 2000. Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). Washington, D. C.: U. S. Government Printing Office.

Iversen, Sandra and William E. Tunmer. 1993. Phonological processing skills and the reading recovery program. Journal of Educational Psychology 85: 112-126.

Juel, C. and C. Minden-Cupp. 2000. Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly 35: 458-492.

Kennedy, M.M. 1978. Findings from the Follow Through Planned Variation Study. Washington, D.C.: U.S. Office of Education.

McArthur, Genevieve. 2008. Does what works clearinghouse work? A brief review of FastForWord®. Australasian Journal of Special Education 32: 101-107.

National Institute of Child Health and Human Development (NICHHD). 1996. Thirty years of NICHD research: What we now know about how children learn to read. Effective School Practices 15: 33-46.

National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, D.C.: National Institute of Child Health and Human Development.

Pflaum, S., H. J. Walberg, M. L. Karigianes, and S.P. Rasher. 1980. Reading instruction: A quantitative analysis. Educational Researcher 9: 12-18.

Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.

Slavin, Robert E. 2008. Perspectives on evidence-based research in education – What works? Issues in synthesizing educational program evaluations. Educational Researcher 37: 5-14.

Smith, S., D. Simmons, M. Gleason, E. Kameenui, S. Baker, M. Sprick, B. Gunn, and C. Thomas. 2001. An analysis of phonological awareness instruction in four kindergarten basal reading programs. Reading and Writing Quarterly 17: 25-50.

Snider, V.E. 1990. Direct instruction reading with average first graders. Reading Improvement 27: 143-148.

Snow, C. E., M. S. Burns, and P. Griffin (eds.). 1998. Preventing reading difficulties in young children. Washington, D. C.: National Academy Press.

Stanovich, K.E. 1994. Romance and reality. The Reading Teacher 47: 280-291.

Appendix A
Communications with the What Works Clearinghouse and Mathematica

1) e-mail –December 2, 2007, from J. S. to WWC
2) Reply from WWC, February 8, 2008
3) JS to WWC, March 16, 2008
4) WWC to JS, April 9, 2008
5) JS to Mathematica, June 25, 2008 (letter sent by U.S. post)
6) Follow-up letter JS to Mathematica, July 31, 2008 (sent fed-ex with receipt notice)
7) Reply from PD to JS, August 6, 2008
8) JS reply to PD, August 7, 2008
9) A. S. to JS, September 6, 2008
10) Reply from Mathematica, September 8, 2008

1) e-mail –December 2, 2007, from J. S. to WWC

Sent: Sunday, December 02, 2007 12:30 PM

To: Mark Dynarski

Subject: Questions on What Works Clearing House Procedures

Dear Dr. Dynarski:

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction (NIFDI). One of my first tasks was to read as much of the background research in the area as I could. Along the way I ran into the "What Works Clearinghouse." In my review I developed several questions that I am hoping you can answer.

First, I noticed that the WWC has listed relatively few of the studies that deal with Direct Instruction. While a few were rejected because they didn't meet some specific standards of design, it appears that many others were rejected because they were published before 1985. I have not been able to find a scientific justification for selecting this date. No studies were cited that suggested children's learning styles had altered after that time or that schools' organizational structures or teachers' actions had changed. In fact, the only research article that I found to justify the methods was actually published in 1974 – 10 years prior to the beginning of your targeted dates. This article dealt with phonological awareness, and had nothing to do with systematic literature reviews and, of course, included nothing to suggest that processes related to phonological awareness have changed over time.

In all the other areas with which I am familiar we use the full range of research articles and, unless there are clear reasons established through the research literature, do not set an arbitrary date for inclusion or exclusion of results. The choice of this date seems to have particularly affected the DI literature because a great deal of the pioneering work was published before the cut-off point. Because I'm new to this particular work, I wondered if you could provide me with any explanation based in the research literature for your choice of this cut-off year. I did find a suggestion that, "if sufficient time and resources remain," studies published before 1985 might be reviewed. Do you know if this step will be taken? From my reading of the literature I believe that your limit on dates of studies may have seriously limited the selection of effective resources that educators will find on your website; and I know that this is a primary aim of your project.

Second, I noticed that a number of studies were rejected because they included children outside the K-3 grade range. I found this somewhat unusual, for the children that are most in need of effective instruction are those who have not learned to read by older ages. Because the effectiveness of DI has been so well demonstrated with regular students at the K-3 grade range, the more recent focus on other ages seems only greater evidence of its worth. This is especially

so in the case of special education students and those with behavioral and other issues. Thus, I couldn't understand the rationale of omitting all the studies that targeted these most needy students. Any explanation you might have of this decision would also be greatly appreciated.

Third, and perhaps most important, as I read studies more closely I found that the Clearinghouse's interpretation of some of the studies didn't seem to accurately reflect the actual content of the reports. I know that these reviews are often done by staff members or graduate students who might not be familiar with all of the intricacies of social research. Thus, I assume that we can send corrections and they will be added.

I look forward to hearing from you, either through e-mail (jeans@uoregon.edu) or by phone (541-346-5005). I'm sure that we share the ultimate goal of helping all students – no matter what their background – develop academic skills and appreciate your appreciate your help.

Sincerely,
Jean Stockard, Professor Emerita

2) Reply from WWC February 8, 2008

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated December 2, 2007 below.

The first question you raised concerned the cut-off year for studies that are reviewed by the WWC. To maintain a focus on current research, our reviews examine research conducted in the past 20 years. This cut-off date of 20 years is the standard for every topic reviewed by the WWC. The review for Early Childhood Education (ECE), under which Direct Instruction falls, marked its cut-off date at 1985 (the review began in 2005).

The ECE protocol states that "if sufficient time and resources remain after we have completed our review of research on interventions implemented post 1985, the ECE team will consider reviewing older research on curricula that are still in widespread use." This determination has not been made to date.

In terms of the age/grade range for the review of Direct Instruction research, the ECE team determined the parameters for the review. For further details on inclusion criteria please see the ECE protocol at http://ies.ed.gov/ncee/wwc/reports/early_ed/index.asp

In response to your third question about the qualifications of WWC topic review team members, each team is comprised of several trained professionals, each playing an integral part in the review process. The Principal Investigator (PI) for each topic is a well-known expert in his/her field and is responsible for leadership in conceptualizing the specific topic area, identifying and addressing issues during the review, and developing and reviewing topic and intervention reports developed for the topic. Leadership includes overseeing the quality in the production of the reports and making decisions, based on methodological and substantive expertise, that are not otherwise covered in the WWC protocols and procedures for report production.

The Project Coordinator (PC) is an established education researcher with relevant methodological and substantive expertise. The coordinator oversees the work of the WWC Review Team; manages that specific review; reviews research ratings; and writes and revises the work plan, protocol, and draft and final reports in collaboration with the PI.

Individual reviewers, who prepare initial summaries of studies for the WWC, are professional researchers with experience in research design and methodology. These reviewers undergo a rigorous training and certification process before conducting WWC reviews. .For more information about the staff that make up the WWC team, visit http://ies.ed.gov/ncee/wwc/overview/index.asp.

The WWC recently created a Quality Review Team to respond to concerns raised by study authors, curriculum developers or other relevant parties about WWC reviews published on our

website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. If a quality review concludes that the original review was flawed, a revision will be published. These quality reviews are one of tools used to ensure that the standards established by the Institute of Educational Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.

If you have concerns about a published WWC review that you think warrant a quality review, please send those concerns to our Help Desk at info@whatworks.ed.gov. Please identify the study in question, the specific issue(s) that you think were handled incorrectly, and where relevant, explain what you think is the correct interpretation of the study.

I hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

3) March 16, 2008 – JS to WWC

From: Jean Stockard [mailto:jeans@uoregon.edu]
Sent: Sunday, March 16, 2008 11:59 AM
To: WhatWorks
Subject: re: Questions on What Works Clearing House Procedures

Thank you for your reply of February 8 to my query of December 2. I am pleased to hear that you have established a Quality Review Team. Since my initial query I have continued to examine the reviews and have found quite a few that appear very problematic. I, and my colleagues, will be submitting our concerns to you within the coming weeks.

Unfortunately, I'm afraid that your reply failed to address two of my concerns. Your responses simply restated WWC's policies regarding the cut-off year for studies and the age-grade parameters without providing the requested justification. Again, I would be very interested in knowing the scientific and pedagogical justifications for not incorporating the entire corpus of work in your reviews and for limiting your reviews of studies to only those that included children in grades K-3.

My earlier, rather lengthy, explanation of my concerns is included in this e-mail. If you wish me to clarify any part of my query, please let me know. Thank you for your attention to this matter.

Sincerely,

Jean Stockard, Ph.D.
Emerita Professor

4) WWC to JS, April 9, 2008

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated March 16, 2008 below.

In response to your first concern, the cut-off date of 20 years is a parameter set to ensure the research reviewed is most-relevant to classrooms as they operate today, and to ensure that the scope of studies to be reviewed is manageable. That said, for every topic reviewed by the WWC, the Principal Investigator (PI) is given the flexibility to include studies published earlier, if they think the expansion is important for the review. In this case, the decision was made to keep the 20 year parameter in place.

In terms of the age-grade parameters established, this is determined by the PI and depends on the topic area and the studies under review. In this case, because the focus of the topic is on early childhood education, the age-grade parameter was set to K-3. Other topic areas have also focused on a subset of age-grade ranges. As the WWC expands, we anticipate expanding these topics to examine outcomes for other age-grade ranges.

The WWC solicited nominations from many sources for the topic areas and prioritized the topic areas based on the following criteria:

- potential to improve important student outcomes;
- applicability to a broad range of students or to particularly important subpopulations;
- policy relevance and perceived demand within the education community; and
- likely availability of scientific studies.

More information about this process and the specific topic areas can be found at http://ies.ed.gov/ncee/wwc/reports/.

Three Direct Instruction interventions were reviewed in two topic areas; the spread across topics was due to the samples included in the studies. Direct Instruction falls under the Early Childhood Education (ECE) area since it includes studies with preschool and kindergarten children where the majority (60% or more) of children in the sample are in preschool. This criterion for inclusion can be found at http://ies.ed.gov/ncee/wwc/pdf/ECE_protocol.pdf. One Direct Instruction intervention is also included in the English Language Learners topic area since it was used to supplement reading instruction for Spanish speaking students in grades K-3.

We hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

5) JS to Mathematica, June 25, 2008 (letter sent by U.S. Post)

June 25, 2008

Anita A. Summers, Chairperson and
Paul T. Decker, President and CEO
Mathematica, Policy Research Inc.
P.O. Box 2393
Princeton, NJ 08543-2393

Dear Drs. Summers and Decker:

I write to express my deep concerns with the material posted on the What Works Clearinghouse, which was established "to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education." I understand that Mathematica has now assumed responsibility for the Clearinghouse. I also know that Mathematica has a well established reputation for high quality research. I write to you in the hope that this correspondence can help you correct the errors in the current reports and establish future review procedures that are in accordance with standard scientific practices. The goals of the What Works Clearinghouse are too important to allow a flawed review process to continue or faulty reviews to be posted. None of us wants parents or schools to be given faulty information about the most effective curriculum for their students.

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction. In beginning to familiarize myself with the area I, of course, reviewed major analyses of Direct Instruction. As you no doubt know, in recent years, meta-analyses have become the most commonly accepted method for examining large bodies of research literature. The most recent meta-analysis of the achievement effects of comprehensive school reform models was conducted by Geoffrey Borman and associates (2002, Review of Educational Research). Their examination found that Direct Instruction was the most researched of any of the models and had the most studies conducted by third parties (those not affiliated with the developers). As they put it, "the research base for Direct Instruction is very extensive and of very good quality" (p. 187). Among the many different models used in their analysis, Direct Instruction had the strongest evidence of effects on student achievement.

In my reviews I found that this conclusion simply echoes that found in earlier meta-analyses and literature reviews. The overwhelming conclusion of the education community for many years, based on solid experimental evidence, is that Direct Instruction is one of, if not the, most effective curriculum currently available for teaching reading and mathematics.

In my examination of the literature I also reviewed the reports on Direct Instruction that had been prepared by the What Works Clearinghouse and was shocked and dismayed to find that the WWC conclusions were in marked contrast to those of the extant literature. My concerns with what I found prompted me to write to the Clearinghouse on two previous occasions (December 2, 2007 and March 16, 2008). The replies I received did not clearly address my

concerns. Even worse, my continuing examination of the work of the Clearinghouse has raised even more serious concerns about the quality of the reviews.

In recent months others have expressed similar concerns regarding the Clearinghouse's conclusions and procedures. Two examples of these critiques are Robert Slavin's article in the January/February, 2008 issue of Educational Researcher and Genevieve McArthur's article in the April, 2008 issue of the Australasian Journal of Special Education. While Slavin and McArthur focused their concerns on reviews related to Success for All (Slavin) and FastForWord® (McArthur), many of the concerns they raise parallel the issues I discovered in my review of the judgments of articles regarding Direct Instruction.

Below I outline in greater detail the extent of my concerns. I first discuss what I see as severe limitations of the process used to decide what studies to include in the review process and then give examples of very severe errors in the reviews themselves. I end with a few suggestions regarding what needs to be done to provide the most accurate information to the education community and to parents.

Problems with Exclusion/Inclusion Decisions

Some of the most disturbing elements of the WWC process involve the ways in which studies are selected for reviews. Some of these decisions arbitrarily limit the range of studies included and thus the information available to schools and parents. For instance, as I noted in my earlier communications, the decision to reject all studies for inclusion that were published before 1985 appears to have no scientific justification. The reply to my April 9 query claimed that the "cut-off date of 20 years is a parameter set to ensure the research reviewed is most relevant to classrooms as they operate today." I have not been able to find any studies cited on the website, or elsewhere, that document that children's learning styles, classrooms, teachers' actions, or schools' organizational structures have drastically changed over that time period. Most high quality studies conducted before 1985 are still relevant today. As I stated in earlier correspondence, in all the other areas with which I am familiar we use the full range of research articles and do not set an arbitrary date for inclusion or exclusion of results.

Similarly, some areas of the review have a very restrictive grade range for studies that will be included. For instance, for reviews of work on beginning reading, a number of studies, of both Direct Instruction and other approaches as well, were rejected because they included children outside the K-3 range, such as grades 3-5 or grades K-4. It is hard to understand why such studies would be summarily rejected.

The combination of the decision to limit reviews to work published after 1985 and to a narrow band of grades directly affected the numbers of studies of Direct Instruction that were reviewed. Even worse, this decision ignored the cumulative nature of science and research inquiry within education. Many studies of Direct Instruction with only K-3 students were completed before 1985. Given the consistently positive results from that work, the research focus then moved to populations of students with more extensive needs, such as older children who had not yet learned to read. Yet these studies were omitted because they were not focused on the general population. The real losers from these decisions are, of course, children and their

families who have been denied accurate and complete information, collected over a long span of time and from many different populations, about the most effective programs.

Some of the decisions regarding exclusion or inclusion of curricula and studies seem to have occurred with no stated justification. For instance, Direct Instruction's Reading Mastery program has long been acclaimed as a highly effective reading program, with strong research supporting its development and documenting its success. Yet, the WWC website states that they chose not to include Reading Mastery in their review of Beginning Reading studies (apparently not realizing that the RITE study noted below involved Reading Mastery). No explanation is given as to why this curriculum, which directly addresses beginning reading and which has a very large research base, was ignored. Again, the real losers with this decision are the students.

Finally, the range of work considered has been limited by downgrading findings from studies that do not incorporate strict random assignment. While we all know that random assignment is the "gold standard" for experimental work, the WWC's over-reliance on this criterion ignores the realities of how school organizations work. With this strict attention to random assignment, other aspects of research designs that are an even greater threat to internal validity can be ignored. The most important of these is no doubt ensuring the fidelity of treatment implementation, making sure that a program is implemented as the developers designed it. The complete rationale behind this concern is too lengthy to include in this letter. I worry that this criterion has involved a sterile and unthinking application of methodological rules that may work for growing corn in a field or worms in a lab. Yet they may be inappropriate, at best, and potentially harmful, at worst, for students in real-life schools and real-life neighborhoods where fidelity of treatment implementation is much more tenuous.

While the concerns noted above have resulted in fewer studies being included, another procedural decision of WWC has resulted in the inclusion of many studies that should probably not have been considered. Unlike most scholarly writing, the WWC material that I have reviewed has surprisingly little relationship to the extant published literature. As I have explored the studies that were reviewed for the website I found that many of them were very difficult to find and, even more shocking, did not appear in peer-reviewed journals, the standard mark of academic respectability. A surprisingly large number of the reports that were considered to have met the evidence standards, either fully or with reservation, were unpublished manuscripts. In some cases, I had to write to the original developers several times before receiving the manuscripts cited in the reviews. In contrast, numerous Direct Instruction articles that were rejected for review have been published in peer-reviewed journals and are available to the general public.

Finally, and perhaps most important, as I have read the studies and their evaluations more closely I have found that *the Clearinghouse's interpretations do not accurately reflect the actual* content of the reports. Below I give some examples of the problems that I have found to date. A very large proportion of the reviews I've read have had little correspondence to the articles themselves.

Quality Problems with the Reviews

One of the programs that received a high rating from WWC in Beginning Reading is Reading Recovery®. Most of the other programs reviewed by WWC, such as Direct Instruction, are designed to be used with entire classrooms of students as part of the regular instructional program. In contrast, Reading Recovery (RR) is a "pull-out" program, one that is applied only when students are having trouble in their regular classrooms. The inclusion of the tutorial based Reading Recovery in the Beginning Reading Category is, in fact, rather surprising, given the WWC's stated objective of reporting on "comprehensive … programs," those that enhance "whole school literacy," and "basals/textbooks intended for whole-school/whole-classroom use" (http://ies.ed.gov/ncee/wwc/PDF/BR_protocol.pdf, p. 5).

The WWC website concludes that "Reading Recovery® was found to have positive effects on students' alphabetics skills and general reading achievement outcomes and potentially positive effects on comprehension and fluency" (http://ies.ed.gov/ncee/wwc/reports/beginning_reading/reading_recovery/). A careful reading of the articles cited by the website, however, indicates that the reviewers apparently did not understand basic elements of the program or the reported research.

For instance, the article by Baenen and associates (1997) is one of the articles cited as supporting WWC's positive conclusions. In this study first graders were randomly assigned to participate in the pull-out tutoring program or to remain in their regular classroom. Note that the design does not compare Reading Recovery to an alternative curriculum, which would be a rigorous test of the treatment, but instead to a no-treatment situation. Even with this built-in advantage, the authors found remarkably little success, especially in the long-term, for students in the Reading Recovery program. The article concludes (p. 176) that although about one-half of the students in the tutoring program had successfully reached first grade reading levels at the end of the year, success rates declined in subsequent years of implementation. In addition, by third grade, there was no difference in achievement scores of students who had participated in Reading Recovery and other students. Two years after the treatment there were no differences in needs for retention, special education or Chapter 1 assignments. The authors concluded that Reading Recovery was very expensive to implement in relation to the benefits that it provided.

An even more disturbing example involves the article by Iversen and Tunmer (1993), which was also cited by the WWC as supporting the conclusion that RR is effective. The major purpose of this study was to compare the standard RR program to a "modified" program that included explicit instruction in phonological skills. The WWC chose to ignore any results regarding this comparison group "because it was a modified version of the standard program." Unfortunately, this decision resulted in completely ignoring the major conclusions of the article and thus misrepresenting the effectiveness of Reading Recovery and phonologically based programs as well.

The major variable of interest to Iversen and Tunmer was how long children took to reach a level of competency where they could discontinue special tutoring, the major goal of a tutoring program such as RR. Students in both the unmodified Reading Recovery program and the modified program (with instruction in phonologically based elements added) eventually

caught up with the other children. But the most important finding was that the students in the modified program were able to discontinue the tutoring much earlier. The standard Reading Recovery program was found to be 37 percent less efficient than the modified program. In addition students in the modified program continued to have higher levels of achievement and higher rates of learning at the end of the school year. The authors provide an extensive discussion and additional analysis that demonstrate the fallacy involved in Reading Recovery about the ways in which word recognition skills develop. They clearly conclude that Reading Recovery is not an efficient method for teaching children to read and that phonological training is superior.

Similar, very serious, problems occurred with the review of studies of Direct Instruction, but the mistakes with these articles led the reviewers to dismiss the studies out of hand. One of the most egregious cases involves the study by Carlson and Freeman (2002). The stated reason for dismissing this study was that "there was a confound, with the Direct Instruction intervention being modified or combined with other interventions." In fact, there was only one intervention, and it appears that the reviewer simply did not understand the nature of the intervention well enough to accurately read the article. As with virtually all other studies of Direct Instruction, the analysis, which used sophisticated multilevel techniques, had results strongly in favor of the DI curriculum.

It is possible that the reviewers responded to this section of the description of the intervention, which was termed "RITE" and used the Reading Mastery program of Direct Instruction: "In addition to the teaching of skills directly related to the RM curricula, the RITE program also strives to provide teachers with strong classroom management techniques" (p. 143). Strong classroom management is part and parcel of the Direct Instruction approach. Part of the reason that DI is so successful is that it provides not just well designed curricular materials but well developed, research-based guidance on how the curriculum should be administered. These guidelines are well documented in the various guides to the programs and should have been part of the knowledge base of a competent reviewer. The classroom management was not a confounding element of the intervention, but was an integral part of the curriculum.

In other cases, articles regarding a Direct Instruction curriculum were rejected for methodological issues, while those with a different curriculum, but the same methodological approach, were accepted. For instance, Waldron-Soler, et al (2002) found that students instructed in the DI program, Language for Learning, had significantly higher achievement than those in programs without explicit language instruction, using analysis of covariance to control for pretest scores. The study was rejected for consideration because "the intervention and comparison groups cannot be considered equivalent at baseline, even with the use of covariates in the analysis." However, the published article includes means and standard deviations for both the experimental and control groups. I calculated simple t-tests and found that in fact there were no significant differences between the groups at pre-test. Such a test was apparently not done by the reviewers, despite their stated conclusion.

To help understand why the Waldron-Soler article was dismissed, I examined studies of Success for All that WWC deemed had met the minimal requirements for inclusion. (Success for All is a reading program that has been found to be effective almost as often as Direct Instruction

and was modeled on many elements of DI.) Three of these papers, none of which had been published in a peer reviewed journal, also used analysis of covariance to adjust for pretest scores (Dianda and Flaherty 1995; Ross, Alberg and McNelis, 1997; and Ross, McNelis, Lewis and Loomis, 1998). Why this adjustment method would pass muster with the unpublished articles regarding Success for All, but be dismissed with the published article regarding Direct Instruction is very unclear and disturbing.

Similarly, two articles by Tobin (2003, 2004) on the Horizons DI reading program were rejected because the study groups were supposedly "incomparable." The studies, which were again published in peer reviewed journals, used a technique of matching students in the experimental and treatment groups on pretest measures. Tobin (2003) reported that there were no statistically significant pretest differences between the groups, and, as usual, the results supported the superiority of the DI program. Again I looked at the accepted studies of Success for All to try to understand why Tobin's articles had been rejected. I found that one of the accepted SFA studies (Smith, Ross, Faulks, et al, 1993) also individually matched students on pretest scores. Unfortunately, this unpublished paper only reports the results in graphical form and it was impossible to actually examine the means and standard deviations to test for any differences. Again, I am baffled as to why the unpublished Smith et al piece, which omitted crucial methodological details, would be accepted while the published Tobin pieces, with similar methodology and including the statistical details, were rejected.

Recent innovations to the WWC website have only compounded these errors. The "create your own summary" feature (http://ies.ed.gov/ncee/wwc/Reports/Topic.aspx?tid=01) invites viewers to create a summary of the reports by simply clicking a button. The resulting page lists interventions in the order of the "improvement index" score, implying that those at the top of the list are the most effective while those at the bottom of the list are least effective.

One of the programs often listed near the top of the Beginning Reading lists is Reading Recovery; but, as explained above, the reviews of the program were fatally flawed and inaccurate. Another program rated near the top is Early Intervention in Reading (EIR). Like Reading Recovery, EIR involves special tutoring for low performers, yet, as with the reviews of Reading Recovery, the judgment of EIR's effectiveness is questionable. The conclusion regarding EIR is based on one unpublished study, involving 59 students. Half of these students received tutoring in addition to their classroom instruction and their reading achievement was compared to those who received no additional systematic instruction. While the children who received tutoring did better than those who did not, there is no way to tell if EIR is superior to another type of intervention or if it is simply the additional practice time or attention that made the difference.

In short, given the flawed process of selecting studies for review and the extremely poor quality of the review process, any "effectiveness ratings" are undoubtedly inaccurate and misleading. To prominently display these ratings and encourage their use does a great disservice to schools and parents throughout the country and will eventually, I believe, harm Mathematica's reputation of scientific integrity.

What Should Be Done

I bring these very serious concerns to your attention in the hope that you will correct the mistakes that have been made and establish procedures that will prevent them from occurring in the future. The articles by Slavin and McArthur provide a number of very good suggestions. For instance, Slavin provides very cogent suggestions regarding the development of program ratings that more accurately reflect the strength of the evidence, altering the decision process regarding acceptable study designs, and giving greater weight to larger studies and those of longer duration. McArthur suggests that reviews be limited to studies that have passed peer review, that measures be adjusted to more accurately reflect the concepts that are supposedly being rated, that the method of rating programs more accurately reflect the results and, most important, that reviewers be familiar with the subject area and that all summaries be reviewed by independent experts in the area before publication on the website.

I fully support these suggestions and applaud their common notion that the WWC should use standard methods of scientific procedure and review and that those who do the reviews should be familiar with the areas under study. Acknowledging that there is overlap with the suggestions of Slavin and McArthur, I will list my own recommendations:

1) Make the criteria used in searching for studies explicit and replicable.

2) Concentrate on peer reviewed studies first.

3) Ensure that all studies accepted are available in full text for review by others.

4) Use all the available evidence regarding interventions, with no arbitrary cut-off dates or age or grade ranges that unfairly exclude a body of evidence.

5) Ensure that reviewers are knowledgeable in the substantive areas that they are reviewing as well as in the methodological details.

6) Develop standard quality control mechanisms, such as two reviewers for every article with a third if results differ and a peer review process before publication of ratings.

7) Develop procedures to ensure comparability from one area of the review to another.

8) Compare results that are obtained with the already well developed body of meta-analyses and literature reviews. When WWC's results differ from these reviews, investigate why and adjust results as needed.

9) Use measures of effect size that are more commonly used within the social sciences and comparable to those used in other review studies. Include elements related to sample size, statistical significance, substantive significance, and length of intervention.

10) Weigh issues of both internal and external validity in decisions regarding acceptable designs and recognize the wide variety of approaches that can and must be used within real-life educational settings.

11) Organize topics in ways that reflect how schools actually operate, such as core reading programs, supplemental programs, etc.

While all of these changes are vital to ensuring that future reviews are accurate, they do not deal with the faulty reviews that are already posted on the website. Given the very serious concerns that Slavin, McArthur, and I (and perhaps others of whom I'm not aware) have raised regarding the content of the WWC website, I believe that the reviews and ratings that are now posted must be removed until they can be reviewed for accuracy with well designed quality control measures. The problems with the material posted on the website appear to be much more widespread than can be handled in a piece-meal fashion. To allow inaccurate material to remain does a disservice to schools, families and students. I know that the vast majority of the work

posted on the website was not done under Mathematica's guidance. Continuing to post such faulty material can only, unfairly, sully Mathematica's reputation.

Thank you for attending to my concerns. Please know that they are presented in the spirit of the intent behind the What Works Clearinghouse – a desire to provide schools and families with the most accurate information. If you wish to discuss any of my concerns further, please do not hesitate to contact me. I look forward to hearing from you.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
jstockard@nifdi.org
Toll free phone # 877-485-1973

6) Follow-up letter, JS to Mathematica, July 31, 2008, sent with return receipt via Fed-ex.

July 31, 2008

Anita A. Summers, Chairperson and
Paul T. Decker, President and CEO
Mathematica, Policy Research Inc.
P.O. Box 2393
Princeton, NJ 08543-2393

Dear Drs. Summers and Decker:

It has now been more than a month since I wrote you regarding my concerns with the material posted on the What Works Clearinghouse website. However, I have received no response or even an acknowledgement of my letter. In case you did not receive the original communication, I have included a copy in this package. I have also sent this letter via Federal Express so that I may confirm that it has been received.

As you know, I believe that the analyses of curricular material posted on the WWC website are fatally flawed. They provide very erroneous information to educators and parents. Because Mathematica only recently assumed responsibility for the WWC and because I know that Mathematica has had, in the past, a reputation for high quality work, I felt it important to let you know of these problems privately so that you could address them.

The lack of response to my earlier letter suggests, however, that my concerns, as well as those of others, have not been given serious credence. Given the serious nature of the issues involved, unless I hear from you within the next 2 weeks, I will have to pursue alternative means of dealing with this matter. I am available to speak via phone (toll-free at 877-485-1973) or to communicate via e-mail (jstockard@nifdi.org) if you prefer those means of communication. I hope to hear from you soon.

Sincerely,

Jean Stockard, Ph.D.
Director of Research
jstockard@nifdi.org
Toll free phone # 877-485-1973

7) Decker to JS, August 6, 2008

Response to Your Letter

Dear Dr. Stockard:

Thank you for prompting me regarding a reponse to your concerns about the What Works Clearinghouse (WWC). I received your earlier letter and reviewed it carefully. As you know, your original letter contained a substantial amount of material as well as 11 specific recommendations to change the criteria and processes supported by the Clearinghouse. I therefore have asked for input from both our team working on the project as well as our client for this work, the Institute of Education Sciences (IES) at the U.S. Department of Education, to develop an appropriate response to your letter.

As you probably know already, Mathematica is administering the WWC under contract to IES, and our work is bound by the statement of work contained in the contract. Furthermore, we work closely with IES to determine how best to fulfill the statement of work as the details of the work unfold. Hence, any potential changes that Mathematica could make in response to your recommendations, including changes in the WWC web site, may be limited by the contract or by the guidance of our client.

Thank you for your attention to the What Works Clearinghouse initiative. I appreciate your thoughtful review, and I look forward to sending you a lengthier response.

Paul Decker

_____

Paul Decker
President and Chief Executive Officer
Mathematica, Inc.
Box 2393
Princeton, NJ 08543
609-275-2290

8) JS to Decker, August 7, 2008

Dear Dr. Decker:

Thank you for your response to my letters. I look forward to receiving your lengthier reply. Might you have any idea of the time frame for this response?

Again, thank you for attending to my concerns.

Sincerely,

Jean Stockard

9) A.S. to J.S., September 6, 2008

Ms. Stockard: Thank you for copying me on your Aug. 29 letter to Paul Decker. I am
following up on it with him, and am assured that he is addressing the issues meticulously. I will continue to connect with him on this matter.

Sincerely,

Anita A. Summers

10) Mathematica reply, September 8, 2008

Note that this has two part: The first is a cover letter from Paul Decker the President and CEO of Mathematica. The second is a more detailed response to the June 25 letter, apparently prepared by Mark Dynarski. This is referred to in the text as Dynarski, 2008.

MATHEMATICA
Policy Research, Inc.

Paul T. Decker
President and Chief Executive Officer

P.O. Box 2393
Princeton, NJ 08543-2393
Telephone (609) 799-3535
Fax (609) 799-0005
www.mathematica-mpr.com

September 8, 2008

Dr. Jean Stockard, Ph.D.
Director of Research
National Institute of Direct Instruction
P.O. Box 11248
Eugene, OR 97440

Dear Dr. Stockard:

Thank you for your June 25 letter concerning the What Works Clearinghouse (WWC). Mathematica Policy Research wants to ensure the quality and accuracy of all information contained in the WWC, and it works with study authors, curriculum developers, and consumers of the WWC to correct inaccurate information. We therefore appreciate your interest in and feedback on the WWC.

The WWC was designed to establish and apply a set of standards to identify rigorous research related to the impacts of education interventions. Studies that meet these standards are considered to have designs with causal validity, which gives practitioners increased confidence that the results of the study actually reflect the true impact of the intervention being examined. These standards were developed by leading experts in education research methodology, and they are applied through a systematic review process that includes repeated checks for quality and accuracy. They represent research criteria that we believe to be vital to the purpose of the WWC.

I have closely reviewed the concerns in your letter. I have also asked for input from Mark Dynarski, the director of the WWC effort, and representatives from the Institute for Education Sciences at the U.S. Department of Education. Based on my review and the input I received, I am convinced that the standards being used by the WWC adequately reflect the original mission of the WWC, as outlined above, and that those standards are being applied appropriately and consistently. In the attached document, Mark Dynarski provides more detailed responses to the issues you raise concerning the WWC's review of Direct Instruction.

Your letter makes several suggestions for the WWC moving forward. Some suggestions are counter to the systematic review process that is the foundation of the WWC. For example, as explained in the attachment, the criteria for date and age ranges being reviewed are established with specific rationales, and the systematic review process gives equal weight to peer-reviewed and non-

An Affirmative Action/Equal Opportunity Employer

peer-reviewed sources. As we expand the WWC's activities into new topic areas, we will be mindful of the suggestions you raised in your letter.

I hope this information clarifies the WWC mission and the way in which the WWC operations support that mission.

Sincerely,

(b)(6)

Paul T. Decker

RESPONSES TO CONCERNS IN 6/25/08 LETTER FROM JEAN STOCKARD

Contrasting conclusions between WWC and extant literature

The letter states that WWC conclusions differ from extant literature including meta-analyses and literature reviews. It is important to note that the WWC is designed to produce a systematic review of literature. A sound definition of a "systematic review" is in the recent publication *Knowing What Works in Health* Care by the Institute of Medicine (IOM):[1] *A systematic review is a scientific investigation that focuses on a specific question and uses explicit, preplanned scientific methods to identify, select, assess, and summarize similar but separate studies.* (pg. 82)

Consistent with the recommendations in the IOM report, the WWC applies evidence-based methodological standards consistently to each study it reviews. These standards, which are available on the WWC website, were developed by leading education research methodologists.

The systematic review conducted by the WWC goes beyond the procedures performed in meta-analyses. Again using the IOM definition, a meta-analysis "quantitatively combines the results of similar studies in an attempt to allow inference from the sample of studies included to the population of interest" (pg. 82). However, as Robert Slavin has noted, meta-analyses rarely describe even one study in any detail.[2] The WWC uses meta-analytic techniques to summarize the results of studies that meet WWC standards. However, the important distinction is that the WWC uses a rigorous set of standards, applied consistently, to determine *which* studies are included in the meta-analytic computations. This ensures that WWC summary measures are based only on studies with causal validity.

It is not surprising that WWC findings may differ from those of other analyses. Some meta-analyses may not have such rigorous standards for including studies; others may have standards that differ from those used by the WWC. Because of these differences, the WWC cannot judge the results of its systematic reviews by how they compare to other analyses. Rather, it works to ensure that its standards appropriately identify studies with strong causal validity and applies those standards consistently to each study reviewed.

---

[1] Institute of Medicine (IOM). *Knowing What Works in Health Care: A Roadmap for the Nation.* Washington, DC: The National Academies Press. 2008.

[2] Slavin, R.E. (1995). Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, Vol.48, No.1 pp.9-18.

Inclusion and exclusion procedures

Limiting studies to 1985 or later

The WWC's default time period for reviews is a study publication date of 1985 or later. This timeframe, which was established in 2005, is used for two reasons. First, by limiting reviews to research to this time period, WWC reviews reflect reasonably current research. In particular, the time period range ensures that effect sizes and improvement indices are based on a counterfactual condition that reflects classrooms as they operate within a recent time period. Second, the timeframe ensures that the research reviewed is examining versions of interventions that are most likely to be available to practitioners today.

WWC principal investigators have the option to expand the period for which studies can be reviewed, if they believe that important research will be excluded. The principal investigators for the Beginning Reading area chose to maintain the default period in large part to maintain currency with the classroom context for beginning readers. For instance, the fact that preschool enrollment has increased,[1] combined with the fact that more preschool and kindergarten programs run full-day,[2] means that students in the early grades may be better prepared to receive reading instruction today than students 25 years ago. Moreover, it is possible that any changes in reading readiness over this period have not been evenly distributed, since differences in reading ability by socioeconomic status and race are apparent at the kindergarten level.[3] Other contextual factors have changed over the past 20 years, including advances in teacher training, increases in home literacy activities, and changes in the content of and variety of curricula used in classrooms.

Any of these changes could have implications for the effectiveness of an intervention. If school readiness has increased, than an intervention that was effective 25 years ago may not be effective in more recent years. If teachers are receiving stronger training and using newer curricula, the counterfactual condition against which interventions were measured 25 years ago have changed, and possibly with it the magnitude of its effects. The Beginning Reading principal investigators judged that they had an inadequate basis for

---

[1] The proportion of 3 and 4 year olds enrolled in school increased from 37 percent in 1980 to 56 percent in 2006. See Snyder, T.D., Dillow, S.A., and Hoffman, C.M. *Digest of Education Statistics 2007*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2008.

[2] The proportion of pre-kindergartners and kindergartners attending full-day pre-kindergartners or kindergarten increased from 32 percent in 1980 to 59 percent in 2006. See Snyder, Thomas D. *Mini-Digest of Education Statistics, 2007*. Washington, DC: National Center for Education Statistics, Institute of Educational Sciences, U.S. Department of Education, 2008.

[3] For example, the Early Childhood Longitudinal Study of Kindergarteners (ECLS-K) found gaps in the reading knowledge and skills of kindergarteners by race: black and Hispanic children scored just under half of a standard deviation below whites on a test of reading knowledge and skills. Analysis of the ECLS-K and other surveys (i.e., Children of the National Longitudinal Study of Youth and the Infant Health and Development Program) show that socioeconomic status accounts for about half of the standard deviation of racial differences in reading test scores. See Duncan, G. and K. Magnuson. "Can Family Socioeconomic Resources Account for Racial and Ethnic Test Score Gaps?" *Future of Children*, Vol. 15, No. 1, Spring 2005.

assuming that effects of interventions measured more than 20 years ago would be experienced if schools adopted those interventions today.

Restrictive grade range for Beginning Reading studies

The letter raised concerns that the WWC's exclusion of research conducted on children outside of the kindergarten to 3rd grade age range was too restrictive. Because the reviews are focused on assessing interventions for beginning reading, the principal investigators, in coordination with the Institute of Education Sciences (IES), concluded that the review should focus on intervention effects for children in kindergarten through third grade (roughly ages 5 to 8). Some studies examine effects of interventions for students within the specified grade range and also students in higher grades. These studies are included in the review when the WWC is able to isolate the effects for the students who fall within the Beginning Reading grade range. The studies are excluded from the review if the results cannot be disaggregated to isolate effects for relevant grade range.

The grade range criterion is important for the integrity of the review process. The WWC-computed improvement indices and effect sizes are intended to reflect the effect of the intervention on the population in question. Including students above the topic area age range could lead to misstatements about intervention effects on children within the grade range.

It should be noted that the WWC attempts to determine a study's effects for the relevant grade range. When authors present findings aggregated across a broader age range and indicate that findings were analyzed for the relevant age range, it is standard WWC procedure to contact authors to request findings disaggregated for the grade range. Sometimes design limitations or other factors preclude authors from providing disaggregated results. In such cases, the WWC excludes the study.

Fidelity of treatment implementation

The letter notes that the WWC review process may downplay implementation fidelity. Definitions of implementation fidelity vary and many studies include little information to gauge fidelity, especially information about whether an intervention has been implemented within normal operating regimes of districts, schools, and teachers, not under specialized laboratory conditions. Moreover, there is no standard metric with which to rate and assess fidelity across studies that assures comparability.

The WWC's approach emphasizes the importance of replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies. Intervention reports include an "extent of evidence" classification that allows practitioners to place more weight if they choose on interventions for which the extent of evidence is large, meaning the results are drawn from multiple studies and a large number of classrooms and students.

Concerns about interventions and studies reviewed by the WWC

Reading Recovery

The letter expressed concern that Reading Recovery is an intervention outside of the Beginning Reading protocol. The Beginning Reading protocol states that interventions that target specific populations (for example, readers below grade level, and at-risk students) are eligible for the review. Reading Recovery is a short-term tutoring intervention program intended to serve the lowest achieving first-grade students (i.e., those in the bottom 20 percent). As such, it falls within the Beginning Reading protocol.

The letter expressed concerns that the reviews of studies of Reading Recovery mischaracterized the findings from those studies. For both the Baenen et al. (1997) and Iversen and Tunmer (1993) studies, the results presented by the WWC review represent the findings when the WWC standards and procedures are applied to these studies.

With respect to the Baenen et al. study, it is important to note that the beginning reading protocol prioritized one-year results. In effect, the Beginning Reading review is only intended to examine whether beginning reading interventions have an effect within one year. This one-year period is applied consistently to each study reviewed to ensure the results can be compared across studies and interventions. The Baenen et al. study's two- and three-year general reading achievement measures are not ignored, however; they are presented in Appendix A4.4 of the Reading Recovery Technical Appendices (http://ies.ed.gov/ncee/wwc/pdf/techappendix01_209.pdf).

With respect to the Iversen and Tunmer study, and consistent with the protocol, the WWC examined the results most relevant to the question of whether Reading Recovery improves reading proficiency compared to a reasonable counterfactual. That the study examined other comparisons is not ignored however. Appendices A4.1, A4.2 and A4.3 present results from other comparison groups. As with any study it reviews, the WWC does not base the findings of its review on the conclusions drawn by the authors.

We disagree that studies that compare an intervention to a no-treatment condition (as was done in Baenen et al.) provides a "built-in advantage," as the letter suggests. Many practitioners are interested in knowing whether an intervention is effective relative to customary classroom practices. The WWC reviews studies comparing treatments to no-treatment as well as studies comparing one treatment to another. In each case, the counterfactual is clearly documented in the review.

Exclusion of Reading Mastery Program

The letter expresses concern that the Reading Mastery Program is excluded from review by the WWC. The WWC has reviewed studies of Reading Mastery. A report summarizing the results of those reviews was published on the WWC website on August 12, 2008.

Inclusion of unpublished manuscripts for review

The letter suggests that by reviewing unpublished manuscripts, the WWC reviews studies of lower quality. Reviewing only studies that have passed a peer reviewing process could neglect unpublished information that may contain important and different findings. Mark Lipsey and David Wilson caution against this type of publication bias in *Practical Meta-Analysis* (2001). The damage such publication bias can cause in a systematic review is substantial. In *Knowing What Works in Healthcare*, the IOM notes that publication bias is well-established and that systematic reviews need to take steps to counter it, because otherwise "harmful interventions may appear to be worthwhile and beneficial interventions may appear to be useless" (page 97). For this reason WWC procedures are explicitly designed to review evidence from sources other than those published in journals. However, regardless of publication status, studies must meet the same evidence standards.

The letter asserts that the WWC should make available to the public all the research it reviewed. The more important issue is that the WWC provides an explicit citation to the study and thereby enables readers to obtain studies they wish to review. Unpublished information obtained from authors in response to queries that arise in reviews also are made available to requestors, and authors sign a form indicating that the information they provide to the WWC is available to the public.

Exclusion of Carlson and Francis (2002) study

The letter expressed concern over the WWC review's conclusion that there was a confound with the Direct Instruction intervention in Carlson and Francis (2002).[1] The WWC standard regarding intervention confounds was established to ensure that the results in a WWC review reflect what educators can expect if they implement the intervention being reviewed. A careful reading of Carlson and Francis indicates that findings cannot be separated into effects of Reading Mastery alone and effects of Reading Mastery supplemented by the support provided to teachers through the RITE program.

Exclusion of Waldron-Soler et al. (2002) study

As described in the protocols for the WWC Early Childhood Education and Beginning Reading reviews, to establish baseline equivalence, treatment and comparison groups must differ on the pretest measures by less than half a standard deviation, or the differences must be insignificant in an adequately powered statistical test. The Waldron-Soler (2002) study reported pretest differences exceeding half a standard deviation on at least two measures.

---

[1] The letter references Carlson and Freeman (2002). We have no record of a Carlson and Freeman (2002) study, and have assumed the letter refers to Carlson, C.D., & Francis, D.J. (2002). "Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE)." *Journal of Direct Instruction*, 3(1), 29-50.

The same standards were applied to the three Success for All studies you cite (Dianda and Flaherty, 1995; Ross, Alberg and McNelis, 1997; and Ross, McNelis, Lewis and Loomis, 1998). Unlike the Waldron-Soler study, pretest differences between the treatment and comparison groups were less than one half of a standard deviation.

Exclusion of Tobin (2003, 2004) studies

The Tobin (2003) study indicated that treatment and comparison groups were substantively different on pretest measures, with the differences exceeding half a standard deviation. Because the Tobin (2004) study follows the same students for additional time, it did not meet standards for the same reason.

The same standards were applied to the Success for All study cited in the letter (Smith, Ross, Faulks et al., 1993). Unlike the Tobin studies, pretest differences between the treatment and comparison groups were not statistically significantly different, and they were less than one half of a standard deviation. In keeping with WWC procedures, the WWC obtained information on pretest differences through communication with the study authors.

The "Create your own summary" feature

The letter expresses concern about the "create your own summary" feature of the WWC website because the results sort interventions by the magnitude of the improvement index. Because the improvement index calculations accurately reflect the application of the WWC evidence standards and effect size computations to the studies of those interventions, the WWC believes it is informative and useful to practitioners to sort results by improvement index.

The improvement index is one summary measure of the effect of beginning reading interventions. Users can also sort interventions alphabetically (by intervention name), by evidence rating, and by extent of evidence indices.

Appendix B
A Partial List of Studies of Reading Mastery and its Precursors Completed before 1985

Note: The letters in parentheses at the end of a citation refer to the source from which it was obtained. These sources are listed at the end of this appendix.

Apffel, J.A., J. Kelleher, M.S. Lilly, and R. Richardson. 1980. Developmental reading for moderately retarded children. Education and Training of the Mentally Retarded 10: 229-235. (b) (g)

Beck, I. L. and E. S. McCaslin. 1978. An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. LRDC Report No. 1978/6. Pittsburgh: University of Pittsburgh Learning Research and Development Center. (c)

Becker, W. C. and D. W. Carnine. 1980. Direct instruction: An effective approach to educational intervention with the disadvantaged and low performers. Pp. 429-473 in B.B. Lahey and A. E. Kazdin (eds.), Advances in Clinical Child Psychology, Volume 3, New York: Plenum. (a)

Becker, W. C. and S. Engelmann. 1976. Analysis of achievement data on six cohorts of low-income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model. Eric Document Reproduction Service No. ED 145922. (a)

Bock, G. and L. B. Stebbins. 1977. Education as experimentation: A planned variation model, Volume IV-B effects of Follow Through models. Cambridge, MA.: Abt Associates (ERIC Document Reproduction Service No. ED 148491). (a)

Booth, A., D. Hewitt, W. Jenkins, and A. Maggs. 1979. Making retarded children literate: A five-year study. The Australian Journal of Mental Retardation 5: 257-260. (d) (i)

Bowers, W. M. l972. An evaluation of a pilot program in reading for culturally disadvantaged first grade students. Doctoral dissertation, University of Tulsa, 1972. (Eric Document Reproduction Service No. ED 073439). (a) (b)

Bracey, S., A. Maggs, and P. Morath. 1975. Effects of a direct phonic apporach in teaching reading with six moderately retarded children: Acquisition and mastery learning stages. Slow Learning Child 22: 83-90. (b) (d) (i)

Branwhite, A. B. 1983. Boosting reading skills by Direct Instruction. British Journal of Educational Psychology 53: 291-298. (b) (c) (d) (g) (i)

Carnine, D. 1977. Phonics versus look-say: Transfer to new words. The Reading Teacher 30: 636-640. (h)

Carnine, D. 1980. Phonic versus whole-word correction procedures following phonic instruction. Education and Treatment of Children 3: 323-329. (h)

Carnine, D. and R. Gersten. 1983. Effectiveness of Direct Instruction in teaching selected reading comprehension skills: Preliminary Draft. Paper presented at American Educational Research Association (April 11-15, 1983). (b)

DuPree, T. J. 1976. Brief history of Cherokee schools, 1804-1976. BIA Education Research Bulletin 4: 3-11. (ERIC Document Retrieval Service No. ED 127051). (a)

Gersten, R., W.C. Becker, T. J. Heiry, and W. A. T. White. 1984. Entry IQ and yearly academic growth of children in Direct Instruction programs: A longitudinal study of low SES children. Educational Evaluation and Policy Analysis 6: 109-121. (e)

Gersten, R. and A. Maggs. 1982. Teaching the general case to moderately retarded children: Evaluation of a five-year project. Analysis and Intervention in Developmental Disabilities 2: 329-343. (i)

Gersten, R., A. Brockway, and N. Henares. 1983. The Monterey DI program for students with limited English (ESL). Direct Instruction News, 2(4): 8-9. (a)

Gordan, M.B. (ed.) 1971. DISTAR instructional system: Summaries of case studies on the effectiveness of the DISTAR instructional system. Chicago: Science Research Associates. (a)

Gregory, R. P. and B. G. Warburton (1983). DISTAR Reading and remedial children in an infant school. School Psychology International 4: 169-172. (i)

Haring, N.G. and D. A. Krug. 1975. Evaluation of a program of systematic instructional procedures for extremely poor retarded children. American Journal of Mental Deficiency 79: 627-631. (b) (i)

Kastner, S. and M. Hollingshead. 1973. An evaluation of ESEA Title I Programs, Community School District 15. New York: New York University, Center for Educational Research and Field Services, School of Education. (ERIC Document Reproduction Service No. ED 087842). (a)

Kaufman, M. 1973. The effect of the DISTAR Instructional System: An evaluation of the 1972-1973 Title I program of Winthrop, Massachusetts. (ERIC Document Reproduction Service No. ED 110171). (a)

Kaufman, M. 1974. The effect of the DISTAR Instructional System: An evaluation of the 1973-1974 Title I program of Winthrop, Massachusetts. (ERIC Document Reproduction Service No. ED 110170). (a)

Lewis, A. 1982. An experimental evaluation of direct instruction programme with remedial readers in a comprehensive school. Educational Psychology 2: 121-135. (g)

Lloyd, J., D. Cullinan, E.D. Heins, and M.H. Epstein. 1980. Direct instruction: Effects on oral and written language comprehension. Learning Disabilities Quarterly 3: 70-76. (g)

McCabe, T.A. 1974. The DISTAR Reading and Language Program: Study of its effectiveness as a method for the initial teaching of reading. Doctoral dissertation, University of Massachusetts, 1974. (ERIC Document Retrieval Service No. ED 102498).

Meyer, L. A. 1983. Long-term academic effects of Direct Instruction Follow Through. Technical report No. 299. Champaign: University of Illinois at Urbana-Champaign, Center for the Study of Reading. (ERIC Document Retrieval Service No. 237932). (a)

Newark (NJ) Board of Education. 1974. Program to improve the informational processoing of children with reading and learning problems. ERIC 106 826. (b)

Ogletree, E. J. 1976. A comparative study of the effectiveness of DISTAR and eclectic reading methods for inner-city children. Chicago: Chicago State University. (ERIC Document Reproduction Service No. ED 146544). (a) (b)

Ogletree, E.J. 1977. Does DISTAR meet the reading needs of inner-city kindergarten pupils? ERIC 146 303. (b)

Rawl, R. K. and F. S. O'Tuel. 1982. A comparison of three prereading approaches for kindergarten students. Reading Improvement 19: 205-211. (a) (b)

Richardson, E., B. Dibenedetto, A. Christ, M. Press, and B. Winsbert. 1978. An assessment of two methods for remediating reading deficiencies. Reading Improvement 15: 82-95. (a) (b) (d) (g) (i)

Sassenrath, J. M. and R. E. Maddux. 1974. Language instruction, background, and development of disadvantaged kindergarten children. California Journal of Educational Research 25: 61-68. (a)

Serwer, B. L., B. J. Shapior, and P. P. Shapiro. 1973. Comparative effectiveness of four methods of instruction on the achievement of children with specific learning disabilities. Journal of Special Education 7: 241-249. (b) (d)

Singer, B. 1973. The effects of structured instruction on kindergarten pupils: Final report. Washington, D.C.: U.S. Office of Education. (ERIC Document Reproduction Service No. ED 087564). (a)

Stebbins, L. B., R.G. Pierre, E. C. Proper, R. B. Anderson, and T. R. Cerva. 1976. Education as experimentation: A planned variation model (Vols. 3A-3B). Cambridge, MA: Abt Associates. (ERIC Document Reproduction Service No. ED 148489). (a)

Stein, C. L. and J. Goldman. 1980. Beginning reading instruction for children with minimal brain dysfunction. Journal of Learning Disabilities 13: 52-55. (b) (g) (i)

Summerell, S. and G. G. Brannigan. 1977. Comparison of reading programs for children with low levels of reading readiness. Perceptual and Motor Skills 44: 743-746. (a) (b) (d) (g)

Williamson, F. 1970. DISTAR reading: Research and experiment. (ERIC Document Reproduction Service No. ED 045318). (a) (b)

Note: Letters in parentheses refer to the source(s) providing the citation.

a. Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.
b. Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.
c. McGraw-Hill, 2002. "Appendix: Research supporting instruction in Reading Mastery: A Selected Annotated Bibliography." Results with Reading Mastery. ERIC ED 469 523.
d. Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.
e. Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.
f. NIFDI research staff found this citation independently.
g. Grossen, Bonnie. n.d. "References for Reading Mastery Reading Review," The Research Base for Reading Mastery, SRA, Eugene, Oregon: University of Oregon. (http://darkwing.uoregon.edu/~adiep/rmref.htm).
h. National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washing ton, D.C.: National Institute of Child Health and Human Development.
i. Kinder, D., R. Kubina, N.E. Marchand-Martella. 2005. Special education and Direct Instruction: An effective combination. Journal of Direct Instruction 5: 1-36.

Appendix C
Studies of Reading Mastery Published 1985 or Later But Not Included in the WWC Review

Note: The letters in parentheses at the end of a citation refer to the source from which it was obtained. These sources are listed at the end of this appendix.

Ball, E. and B. Blachman. 1991. Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? Reading Research Quarterly 26: 49-66. (d)

Bateman, B. 1991. Teaching word recognition to slow-learning children. Reading, Writing, and Learning Disabilities 7: 1-16. (b)

Brent, G. and N. DiObilda. 1993. Effects of curriculum alignment versus direct instruction on urban children. Journal of Educational Research 86: 333-338. (a)

Brett, A., L. Rothlein, and M. Hurley. 1996. Vocabulary acquisition from listening to stories and explanation of target words. Elementary School Journal 96: 415-422. (d)

Bryne, B. and R. Fielding-Barnsley. 1991. Evaluation of a program to teach phonemic awareness to young children. Journal of Educational Psychology 83: 451-455. (d)

Butler, P.A. 2003. Achievement outcomes in Baltimore City Schools. Journal of Education for Students Placed at Risk 8: 33-60. (e)

Chamberlain, L.A. 1987. Using DI in a Victoria, B.C. resource room. ADI News 7 (1): 7-8. (i)

Cole, K.N. Dale, P.S. and P. E. Mills. 1991. Individual differences in language delayed children's responses to direct and interactive preschool instruction. Topics in Early Childhood Special Education 11: 99-124. (i)

Cole, K. N., P.S. Dale, P.E. Mills, and J. R. Jenkins. 1993. Interaction between early intervention curricula and student characteristics. Exceptional Children 60: 17-28. (i)

Cooke, N.L., S.L. Gibbs, M. L. Campbell, and S. L. Shalvis. 2004. A comparison of Reading Mastery Fast Cycle and Horizons Fast Track A-B on the reading achievement of students with mild disabilities. Journal of Direct Instruction 4: 139-151. (i)

Cunningham, A. 1990. Explicit versus implicit instruction in phonemic awareness. Journal of Experimental Child Psychology 50 :249-444. (d)

Dale, P.S. and K.N. Cole. 1988. Comparison of academic and cognitive programs for young handicapped children. Exceptional Children 54: 439-447. (i)

Darch, C. and E. Kameenui. 1987. Teaching critical reading skills to learning disabled children. Learning Disability Quarterly 10: 82-92. (g)

Diobilda, N. and G. Brent. 1986. Direct Instruction in an urban school system. Reading Instruction Journal 29: 2-5. (e)

Dowdell, T. 1996. The effectiveness of Direct Instruction on the reading achievement of sixth graders. (ERIC Document Retrieval Service No. ED 396268). (a) (d)

Fredrick, L.D., M.C. Keel, and J. H. Neel. 2002. Making the most of instructional time: Teaching reading at an accelerated rate to students at risk. Journal of Direct Instruction 2: 57-63. (f)

Gersten, R. and D. Carnine. 1986. Direct Instruction in reading comprehension. Educational Leadership 44: 68-78. (b)

Gersten, R., C. Darch, and M. Gleason. 1988. Effectiveness of a Direct Instruction academic kindergarten for low-income students. The Elementary School Journal 89: 227-240. (e) (h)

Gersten, R., T. Keating, and W. Becker. 1988. The continued impact of the direct instruction model: Longitudinal studies of follow through students. Education and Treatment of Children 11: 318-327. (h)

Gunn, B., A. Biglan, K. Smolkowski, and D. Ary. 2000. The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. Journal of Special Education 2: 90-103. (c) (d)

Haskell, D. W., B. R. Foorman, and P. R. Swank. 1992. Effects of three orthographic/phonological units on first-grade reading. Remedial and Special Education 13: 40-49. (c) (d)

Herrera, J. A., C. H. Logan, P. G. Cooker, D. P. Morris, and D. E. Lyman. 1997. Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children. Learning disability or learning difference? Reading Improvement 34: 71-89. (d)

Jenkins, J. R., B. Matlock, and T. A. Slocum. 1989. Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. Reading Research Quarterly 24: 215-235. (d)

Juel, C. 1988. Learning to read and write: A longitudinal study of 54 children from first through fourth grade. Journal of Educational Psychology 80: 437-447. (d)

Juel, C. and C. Minden-Cupp. 2000. Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly 35: 458-492. (d)

Kamps, D.M., H.P. Wills, C.R. Greenwood, S. Thorne, J.F. Lazo, J.L. Crockett, J.M. Akers, and B. OL. Swaggart. 2003. Curriculum influences on growth in early reading fluency for students with academic and behavioral risk. Journal of Emotional and Behavioral Disorders 11: 211-224. (f)

Kaiser, S. K. Palumbo, R.C. Bialozor, and T. F. McLaughlin. 1989. Effects of Direct Instruction with rural remedial students: A brief report. Reading Improvement 26: 88-93. (b) (c) (g)

Keel, M.C., L. D. Frederick, T. A. Hughes, and S.H. Owens. 1999. Using paraprofessionals to deliver Direct Instruction reading programs. Effective School Practices 18: 16-22. (e)

Kuder, S.J. 1990. Effectiveness of the DISTAR reading program for children with learning disabilities. Journal of Learning Disabilities 23: 69-71. (b) (d) (i)

Kuder, S. J. 1991. Language abilities and progress in a Direct Instruction reading program for students with learning disabilities. Journal of Learning Disabilities 24: 124-127. (b) (d)

Leach, D. and S. Siddall. 1990. Parental involvement in the teaching of reading: A comparison of hearing reading, paired reading, pause, prompt, praise, and direct instruction methods. British Journal of Educational Psychology 60: 349-355. (h)

Ligas, M. R. and D. W. Vaughan. 1999. Alliance of quality schools: 1998-99 evaluation report. Broward , Florida: Broward County Schools. (a)

MacIver, M.A., E. Kedmper, and S. Stringfield. 2000. The Baltimore Curriculum Project: Fourth year report. Report for the Abell Foundation. Baltimore: Johns Hopkins University, Center for Social Organization of Schools. (a)

Marston, D., S. Deno, D. Kim, K. Diment, and D. Rogers. 1995. Comparison of reading intervention approaches for students with mild disabilities. Exceptional Children 62: 20-37. (h) (i)

Mathes, P.G. and T.J. Proctor. 1988. Direct Instruction for teaching “hard to teach” students. Reading Improvement 25: 92-97. (b)

Neely, M. 1995. The multiple effects of whole language, precision teaching, and Direct Instruction on first-grade story-reading. Effective School Practices 14: 33-42. (d)

O’Connor, R. E. and J. R. Jenkins. 1995. Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. The Journal of Special Education 29: 255-275. (i)

O’Connor, R., J. Jenkins, K. N. Cole, and P.E. Mills. 1993. Two approaches to reading instruction with children with disabilities: Does program design make a difference? Exceptional Children 59: 312-323. (b) (d) (i)

Rawl, R. K. and F. O'Tuel. 1982. A comparison of three prereading approaches for kindergarten students. Reading Improvement 19: 205-211. (d)

Reutzel, D. R. and P. M. Hollingsworthy. 1993. Effects of fluency training on second graders' reading comprehension. Journal of Educational Research 86: 325-331. (d)

Ross, S., J. Nunnery, E. Goldfeder, A. J. McDonald, R. Rachor, M. Hornbeck, et al. 2004. Using school reform models to improve reading achievement: A longitudinal study of Direct Instruction and Success for All in an urban district. Journal of Education for Students Placed at Risk 9: 357-389. (e)

Salerno, C. 1992. A comparison of classrooms using a meaning-centered approach and a code-centered approach. ADI News 11 (2): 26-29. (d)

Sexton, C. W. 1989. Effectiveness of the DISTAR Reading I Program in developing first graders' language skills. Journal of Educational Research 82: 289-293. (a) (b) (d) (g)

Smith, S., D. Simmons, M. Gleason, E. Kameenui, S. Baker, M. Sprick, et al. 2001. An analysis of phonological awareness instruction in four kindergarten basal reading programs. Reading and Writing Quarterly 17: 25-50. (d)

Snider, V. E. 1990. Direct Instruction reading with average first-graders. Reading Improvement 27: 143-148. (b) (g) (h)

Stein, M. 1990. *Reading research ... and Reading Mastery*. Chicago: SRA. (b)

Tobin, K. 2000. A comparison between Horizons Fast Track A-B and Silver, Burdett, and Ginn reading curricula in first grade: June 2000 final report. Oregon: National Institute for Direct Instruction. (a) (e)

Torgesen, J., S. Morgan, and C. Davis. 1992. Effects of two types of phonological awareness training on word learning in kindergarten children. Journal of Educational Psychology 84: 364-370. (d)

Traweek, D. and Berninger, V. 1997. Comparisons of beginning literacy programs: Alternative paths to the same learning outcome. Learning Disability Quarterly 20: 160-168. (a) (d) (h)

Umbach, B. T., C. Darch, and G. Halpin. 1987. Teaching reading to low performing first graders: A comparison of two instructional approaches. ERIC 290 130. (b)

Umbach, B., C. Darch, and G. Halpin. 1989. Teaching reading to low performing first graders in rural schools: A comparison of two instructional approaches. Journal of Instructional Psychology 16: 23-30. (h)

Urdegar, S. M. 1998. Evaluation of the Success for All Program, 1997-98. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research. (a)

Urdegar,k S. M. 2000. Evaluation of the Success for All Program, 1998-99. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research. (a)

Weinstein, G. and N. L. Cooke. 1992. The effects of two repeated reading interventions on generalization of fluency. Learning Disability Quarterly 15: 21-28. (d)

Varela-Russo, C., K. A. Blasik, and M. Ligas. 1998. Alliance of quality schools evaluation report. Ft. Lauderdale, FL: School Board of Broward County. (a)

Yu, L. and R. Rachor. 2000. The two-year evaluation of the three-year Direct Instruction program, in an urban public school system. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 441831). (a) (e)

Note: Letters in parentheses refer to the source(s) providing the citation.

a. Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. Review of Educational Research 73: 125-230.
b. Adams, Gary L. and Siegfried Engelmann. 1996. Research on Direct Instruction: 25 years beyond DISTAR. Seattle: Educational Achievement Systems.
c. McGraw-Hill, 2002. "Appendix: Research supporting instruction in Reading Mastery: A Selected Annotated Bibliography." Results with Reading Mastery. ERIC ED 469 523.
d. Schieffer, Cheryl, Nancy E. Marchand-Martella, Ronald C. Martella, Flint L. Simonsen, and Kathleen M. Waldron-Soler. 2002. An analysis of the reading mastery program: Effective components and research review. Journal of Direct Instruction 2: 87-199.
e. Comprehensive School Reform Quality Center. 2006. CSRQ Center Report on Elementary School Comprehensive School Reform Models. Washington, D.C. American Institutes for Research.
f. NIFDI research staff found this citation independently.
g. Grossen, Bonnie. n.d. "References for Reading Mastery Reading Review," The Research Base for Reading Mastery, SRA, Eugene, Oregon: University of Oregon. (http://darkwing.uoregon.edu/~adiep/rmref.htm).
h. National Reading Panel. 2000. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washing ton, D.C.: National Institute of Child Health and Human Development.
i. Kinder, D., R. Kubina, N.E. Marchand-Martella. 2005. Special education and Direct Instruction: An effective combination. Journal of Direct Instruction 5: 1-36.

Appendix D
The Impact on Summary Ratings of Ignoring Fidelity of Implementation

As noted in the text of this report (p. 11), the WWC has indicated that they have chosen to ignore ratings of fidelity of implementation and, instead, to rely on "replicated findings, which ensures that any one study in which fidelity issues may have arisen are averaged with findings from other studies" (Dynarski, 2008, p. 3, in Appendix A of this report). As noted in the text, this assumption is seriously flawed. In fact, for both exemplary programs and for programs that are ineffective, poor implementation of a program would, very likely, produce results that are systematically biased. This has very serious consequences for consumers, leading to minimizing the positive impact of good programs as well as the negative impact of poor programs.

To illustrate these results the tables below examine the possible implications of inadequate fidelity of implementation for two different hypothetical curricular programs: one that is more effective than a control group and one that is less effective than the control group. We use effect sizes (the difference between the mean scores of experimental and control groups divided by the common standard deviation) as the metric of comparison. Basic to our analysis is the assumption of "regression toward the mean," the well established statistical phenomenon where those with high scores (or low scores) on a measure will tend to have scores that are closer to the mean (lower for those who are high and higher for those who are low) at later testing periods. Simply through regression toward the mean we would expect those who score lower than a mean at pretest to score higher (and closer to the mean) at posttest. For sake of illustration we will assume that the scores involved are normal curve equivalent (NCE) scores with a control group mean of 50 and a standard deviation of 21.

Consider first a program that is effective. The first column of Table D-1 gives four possible values of the experimental group that we will assume are the "true" values and the second column gives the value of Cohen's d that results when comparing these experimental means with those of the control group. It can be seen that the d values vary from .24, when the experimental mean is 55, to .95, when the experimental mean is 70. The higher values in this column are, in fact, similar to those often obtained in studies of Direct Instruction, especially when it is implemented with fidelity.

The other columns in part A of Table D-1 examine the likely result if the effective program is implemented with less than optimal fidelity, but the impact is simply the "random" effects that the WWC assumes will occur. In these calculations the means of the experimental group do not change, as implied by the WWC assumption. However, the standard deviation is assumed to change, for with lower levels of fidelity, even if the means stay constant, greater variability would be expected. This is at the basis of the WW notion of "averaging out." Three different values of the standard deviation are given, and it can be seen that in all situations the value of the effect size becomes lower than the "true" value, often substantially so. (Compare the "true" values in the second column with the other values of d going across each row.)

Panel B of Table D-1 presents results that are probably more likely if an effective program is implemented with less than adequate fidelity. In such a situation it would be logical to expect that performance would be lower – that the mean value would decline. As with Panel

A, the first two columns of Panel B give the “true” experimental mean and effect size for comparison purposes. The next two columns present the effect sizes that would occur if the average value of the experimental group declined by either 5 points (column 3) or by 10 points), but assuming no change in the standard deviation. Again, all of the effect sizes are smaller. Finally, the calculations in the last three columns of Panel B assume that lower fidelity with an effective program has two effects: the average score is lower than it would otherwise be and the standard deviation is larger, as in Panel A. Again, of course, the effect sizes are substantially lower and the declines are marked.

Table D-1: Effect Sizes of Comparisons with Effective Programs with Different Fidelity Conditions

A. Random Influences of Fidelity Problems

Mean, Experimental Group	"True" d	s.d. = 25	s.d. = 30	s.d. = 35
55	0.24	0.20	0.17	0.14
60	0.47	0.40	0.33	0.29
65	0.71	0.60	0.50	0.43
70	0.95	0.80	0.67	0.57

B. Systematic Influences of Fidelity Problems

		Systematic - Experimental mean smaller, s.d. the same		Systematic - Exp Mean smaller, s.d. larger		
Mean, Experimental Group	“True” d	Mean 5 points less	Mean 10 points less	mean 5 points less, s.d. = 25	mean 10 points less, s.d. = 25	mean down by 10 points, s.d. = 30
55	0.24	0.00	-0.24	0.00	-0.20	-0.17
60	0.47	0.24	0.00	0.20	0.00	0.00
65	0.71	0.47	0.24	0.40	0.20	0.17
70	0.95	0.71	0.47	0.60	0.40	0.33

Note: If the impact of fidelity implementation is random, it is assumed that this affects only the common standard deviation and not the mean. If the impact of fidelity implementation is systematic, both the mean and the standard deviation can be affected. Cohen's d is calculated by subtracting the mean of the experimental group from the mean of the control group and dividing by the common standard deviation. The "true" calculations assume that the standard deviation is 21, the value of the control group = 50 and experimental mean varies as shown in each of the rows of the table.

Now consider the possible implications of poor fidelity of implementation of a program that is, in reality, ineffective and, in fact, produces poorer results than the control group. These results are shown in Table D-2. As with Table D-1, the results with the “true” values are shown

in the first two columns of each panel. It can be seen that the "true" effect sizes vary from -.24 to -.95.

The results in Panel A illustrate what would happen if the fidelity problems produce only random changes: the means would stay the same, but the standard deviations would become larger. It can be seen that, in all cases, the absolute values of the effect sizes become smaller – that is, less negative.

Table D-2: Effect Sizes of Comparisons with Ineffective Programs with Different Fidelity Conditions

A. Random Influences of Fidelity Problems

Mean, Experimental Group	"True" d	s.d. = 25	s.d. = 30	s.d. = 35
45	-0.24	-0.20	-0.17	-0.14
40	-0.47	-0.40	-0.33	-0.29
35	-0.71	-0.60	-0.50	-0.43
30	-0.95	-0.80	-0.67	-0.57

B. Systematic Influences of Fidelity Problems

		Systematic - Experimental mean larger, s.d. the same		Systematic - Experimental Mean higher, s.d. smaller		
Mean, Experimental Group	"True" d	mean 5 points higher	mean 10 points higher	mean 5 points higher, s.d. = 18	mean 10 points higher, s.d. = 18	mean 10 points higher, s.d. = 15
45	-0.24	0.00	0.24	0.00	0.28	0.33
40	-0.47	-0.24	0.00	-0.28	0.00	0.00
35	-0.71	-0.47	-0.24	-0.56	-0.28	-0.33
30	-0.95	-0.71	-0.47	-0.83	-0.56	-0.67

Note: If the impact of fidelity implementation is random, it is assumed that this affects only the common standard deviation and not the mean. If the impact of fidelity implementation is systematic, both the mean and the standard deviation can be affected. Cohen's d is calculated by subtracting the mean of the experimental group from the mean of the control group and dividing by the common standard deviation. The "true" calculations assume that the standard deviation is 21, the value of the experimental mean varies as shown in each of the rows of the table.

The results in Panel B illustrate the results if the impact of poor fidelity is systematic. With programs that are ineffective, it would be expected that poor implementation would lead to higher average scores (through regression to the mean) and smaller standard deviations. The

smaller standard deviations result from having values that are "less ineffective." As can be seen in Table D-2, the result is that the effect sizes are less negative. That is, the extent to which the programs are truly ineffective, as shown by the "true" values of d, is disguised.

The results presented above seem to clearly refute the assumption that guides the WWC's approach to considering the fidelity of implementation in deliberations. Even if the results of poor fidelity of implementation were random, the efficacy of both good and poor programs would be misrepresented. If, as is more likely, the results of poor implementation are not random, the impact would be even greater. Good programs appear less effective and poor programs appear better than they actually are. The WWC's decision regarding consideration of the fidelity of implementation appears to be seriously misguided.

The Impact of Reading Mastery in Kindergarten on Reading Achievement through the Primary Grades: A Cohort Control Group Design from a Rural, Midwestern School

December 28, 2010

Jean Stockard, Ph.D.
Director of Research,
National Institute for Direct Instruction
Eugene, Oregon
jstockard@nifdi.org

The Impact of Reading Mastery in Kindergarten on Reading Achievement through the Primary Grades: A Cohort Control Group Design

Purpose:
This study examined the relationship of exposure to Reading Mastery in kindergarten with growth in reading skills through the primary grades.

Sample and Design:
This study examined reading skills through the primary grades of students in a rural, Midwestern district that occurred after the implementation of a highly structured and explicit reading curriculum (Reading Mastery). The study used a cohort control group (also called a recurrent institutional cycle) design, comparing cohorts of students who had not been exposed to Reading Mastery in kindergarten with cohorts that had such exposure. The research design literature suggests that this is a valid and useful alternative to randomized control trials in organizational settings:

> Many institutions experience regular turnover as one group "graduates" to another level and their place is taken by another group. Schools are an obvious example of this, as most children are promoted from one grade to the next each year….The term cohort designates the successive groups that go through processes such as these. Cohorts are particularly useful as control groups if (1) one cohort experiences a given treatment and earlier or later cohorts do not; (2) cohorts differ in only minor ways from their contiguous cohorts; (3) organizations insist that a treatment be given to everybody, thus precluding simultaneous controls and making possible only historical controls; and (4) an organization's archival records can be used for constructing and then comparing cohorts. (Shadish, Cook, and Campbell, 2002, pp. 148-149, emphasis in original,; see also Cook & Campbell, 1979, pp. 126-127 and Campbell & Stanley, 1963, pp. 56-61).

The data used in this analysis fit these criteria. All members of each cohort received the assigned treatment, and there were no apparent differences between the contiguous cohorts.

There were 790 students in the kindergarten cohorts, 138 in the control cohort (no RM in K and attending K in the 2003-04 school year) and 652 in the treatment cohorts (RM in kindergarten and attending K in the 2004-05 through 2007-08 school years). The control cohorts began RM in first grade. No differences in results were found when the cohort groups were analyzed individually.

Measures and Analysis:
Two fluency measures from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) system were used: nonsense word fluency (NWF) and oral reading fluency (ORF). The DIBELS measures have well-established validity and reliability. ORF scores were translated to Lexiles to provide comparability from one year to another. Mean scores of students in the cohort groups were compared using Student t-tests, effect sizes, repeated measures analyses of variance and linear growth models.

Results:
Two DIBELS measures were used to assess skills at the beginning of kindergarten – letter naming fluency (LNF) and initial sound fluency (ISF) – and results indicated no differences between the cohort groups in these average scores. In other words, the cohorts had equivalent skills at the beginning of kindergarten. (See the first lines of Table 1.)

However, by the middle of kindergarten the cohort groups had significantly different reading skills, with those exposed to RM having significantly higher NWF scores from the middle of kindergarten (the first testing period) through the beginning of second grade (the last testing period). Similar results occurred with Oral Reading Fluency lexile scores, with students who had received RM in kindergarten outscoring those who did not have this exposure through the beginning of third grade. (See Table 1.)

Similar results appeared with multivariate analyses, including repeated measures anovas and linear growth models that controlled for initial skills. In the analyses of growth in both NWF and ORF Lexiles the coefficients associated with cohort membership were statistically significant and substantively strong, slightly larger than the coefficient associated with time. In other words, the difference between the cohort groups was slightly larger than the difference in scores of two testing periods (e.g. fall and winter) for the total group.

Conclusion
The results of this study replicate numerous other studies that have found that students receiving instruction in Reading Mastery in kindergarten have significantly higher reading achievement in later grades than students who study other curricula.

Table 1
DIBELS Scores by Having DI in Kindergarten

	No RM in K			RM in K					
	Mean	SD	N	Mean	SD	N	t	prob.	Cohen's d
Skills at Start of K									
LNF	11	13	138	12	13	652	-0.58	0.56	0.06
ISF	9	8	138	9	8	652	0.29	0.78	-0.03
NWF Scores									
NWF K-2	23	14	139	32	19	636	6.83	<.0001	0.58
NWF K-3	44	20	140	52	28	631	3.83	0.0002	0.32
NWF 1-1	36	20	126	41	27	573	2.22	0.03	0.20
NWF 1-2	57	25	124	72	35	559	5.7	<.0001	0.51
NWF 1-3	65	33	124	83	39	552	5.2	<.0001	0.49
NWF 2-1	63	30	114	73	36	513	2.87	0.005	0.28
ORF Reading Lexiles									
Lex 1-2	-118	269	123	48	304	556	5.6	<.0001	0.58
Lex 1-3	11	281	123	191	292	550	6.22	<.0001	0.63
Lex 2-1	181	227	113	230	239	511	2.01	0.04	0.21
Lex 2-2	404	251	110	500	276	365	3.26	0.001	0.36
Lex 2-3	523	261	108	589	279	362	2.17	0.03	0.24
Lex 3-1	372	251	101	439	259	346	2.31	0.02	0.26

Note: If the two groups had unequal variances, an adjusted t-test was computed.

References

Campbell, D.T. & Stanley, J.C. (1963). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

Cook, T.D. & Campbell, D.T. (1979). Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Shadish, W.R.. Cook, T.D. & Campbell, D.T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin.

From: WhatWorks
Sent: Wednesday, April 09, 2008 4:37 PM
To: 'Jean Stockard'
Subject: RE: Questions on What Works Clearing House Procedures
Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated March 16, 2008 below.

In response to your first concern, the cut-off date of 20 years is a parameter set to ensure the research reviewed is most-relevant to classrooms as they operate today, and to ensure that the scope of studies to be reviewed is manageable. That said, for every topic reviewed by the WWC, the Principal Investigator (PI) is given the flexibility to include studies published earlier, if they think the expansion is important for the review. In this case, the decision was made to keep the 20 year parameter in place.

In terms of the age-grade parameters established, this is determined by the PI and depends on the topic area and the studies under review. In this case, because the focus of the topic is on early childhood education, the age-grade parameter was set to K-3. Other topic areas have also focused on a subset of age-grade ranges. As the WWC expands, we anticipate expanding these topics to examine outcomes for other age-grade ranges.

The WWC solicited nominations from many sources for the topic areas and prioritized the topic areas based on the following criteria:

- potential to improve important student outcomes;
- applicability to a broad range of students or to particularly important subpopulations;
- policy relevance and perceived demand within the education community; and
- likely availability of scientific studies.

More information about this process and the specific topic areas can be found at http://ies.ed.gov/ncee/wwc/reports/.

Three Direct Instruction interventions were reviewed in two topic areas; the spread across topics was due to the samples included in the studies. Direct Instruction falls under the Early Childhood Education (ECE) area since it includes studies with preschool and kindergarten children where the majority (60% or more) of children in the sample are in preschool. This criterion for inclusion can be found at http://ies.ed.gov/ncee/wwc/pdf/ECE_protocol.pdf. One Direct Instruction intervention is also included in the English Language Learners topic area since it was used to supplement reading instruction for Spanish speaking students in grades K-3.

We hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: Jean Stockard [mailto:jeans@uoregon.edu]
Sent: Sunday, March 16, 2008 11:59 AM
To: WhatWorks
Subject: re: Questions on What Works Clearing House Procedures

Thank you for your reply of February 8 to my query of December 2. I am pleased to hear that you have established a Quality Review Team. Since my initial query I have continued to examine the reviews and have found quite a few that appear very problematic. I, and my colleagues, will be submitting our concerns to you within the coming weeks.

Unfortunately, I'm afraid that your reply failed to address two of my concerns. Your responses simply restated WWC's policies regarding the cut-off year for studies and the age-grade parameters without providing the requested justification. Again, I would be very interested in knowing the scientific and pedagogical justifications for not incorporating the entire corpus of work in your reviews and for limiting your reviews of studies to only those that included children in grades K-3.

My earlier, rather lengthy, explanation of my concerns is included in this e-mail. If you wish me to clarify any part of my query, please let me know. Thank you for your attention to this matter.

Sincerely,

Jean Stockard, Ph.D.
Emerita Professor

At 11:44 PM 2/8/2008, you wrote:

-->

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Please see the response to the questions from your email dated December 2, 2007 below.

The first question you raised concerned the cut-off year for studies that are reviewed by the WWC. To maintain a focus on current research, our reviews examine research conducted in the past 20 years. This cut-off date of 20 years is the standard for every topic reviewed by the WWC. The review for Early Childhood Education (ECE), under which Direct Instruction falls, marked its cut-off date at 1985 (the review began in 2005).

The ECE protocol states that "if sufficient time and resources remain after we have completed our review of research on interventions implemented post 1985, the ECE team will consider reviewing older research on curricula that are still in widespread use." This determination has not been made to date.

In terms of the age/grade range for the review of Direct Instruction research, the ECE team determined the parameters for the review. For further details on inclusion criteria please see the ECE protocol at http://ies.ed.gov/ncee/wwc/reports/early_ed/index.asp

In response to your third question about the qualifications of WWC topic review team members, each team is comprised of several trained professionals, each playing an integral part in the review process. The Principal Investigator (PI) for each topic is a well-known expert in his/her field and is responsible for leadership in conceptualizing the specific topic area, identifying and addressing issues during the review, and developing and reviewing topic and intervention reports developed for the topic. Leadership includes overseeing the quality in the production of the reports and making decisions, based on methodological and substantive expertise, that are not otherwise covered in the WWC protocols and procedures for report production.

The Project Coordinator (PC) is an established education researcher with relevant methodological and substantive expertise. The coordinator oversees the work of the WWC Review Team; manages that specific review; reviews research ratings; and writes and revises the work plan, protocol, and draft and final reports in collaboration with the PI.

Individual reviewers, who prepare initial summaries of studies for the WWC, are professional researchers with experience in research design and methodology. These reviewers undergo a rigorous training and certification process before conducting WWC reviews. .For more information about the staff that make up the WWC team, visit

http://ies.ed.gov/ncee/wwc/overview/index.asp.

The WWC recently created a Quality Review Team to respond to concerns raised by study authors, curriculum developers or other relevant parties about WWC reviews published on our website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. If a quality review concludes that the original review was flawed, a revision will be published. These quality reviews are one of tools used to ensure that the standards established by the Institute of Educational Sciences (IES) are upheld on every review conducted by the What Works Clearinghouse.

If you have concerns about a published WWC review that you think warrant a quality review, please send those concerns to our Help Desk at info@whatworks.ed.gov. Please identify the study in question, the specific issue(s) that you think were handled incorrectly, and where relevant, explain what you think is the correct interpretation of the study.

I hope you found this information helpful. If you have additional questions, please feel free to contact us again.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

From: Jean Stockard
[
mailto:jeans@uoregon.edu]

Sent: Sunday, December 02, 2007 12:30 PM

To: Mark Dynarski

Subject: Questions on What Works Clearing House Procedures

Dear Dr. Dynarski:

I am a sociologist with many years of experience in quantitative research and recently began working with the National Institute for Direct Instruction (NIFDI). One of my first tasks was to read as much of the background research in the area as I could. Along the way I ran into the "What Works Clearinghouse." In my review I developed several questions that I am hoping you can answer.

First, I noticed that the WWC has listed relatively few of the studies that deal with Direct Instruction. While a few were rejected because they didn't meet some specific standards of design, it appears that many others were rejected because they were published before 1985. I have not been able to find a scientific justification for selecting this date. No studies were cited that suggested children's learning styles had altered after that time or that schools' organizational structures or teachers' actions had changed. In fact, the only research article that I found to justify the methods was actually published in 1974 – 10 years prior to the beginning of your targeted dates. This article dealt with phonological awareness, and had nothing to do with systematic literature reviews and, of course, included nothing to suggest that processes related to phonological awareness have changed over time.

In all the other areas with which I am familiar we use the full range of research articles and, unless there are clear reasons established through the research literature, do not set an arbitrary date for inclusion or exclusion of results. The choice of this date seems to have particularly affected the DI literature because a great deal of the pioneering work was published before the cut-off point. Because I'm new to this particular work, I wondered if you could provide me with any explanation based in the research literature for your choice of this cut-off year. I did find a suggestion that, "if sufficient time and resources remain," studies published before 1985 might be reviewed. Do you know if this step will be taken? From my reading of the literature I believe that your limit on dates of studies may have seriously limited the selection of effective resources that educators will find on your website; and I know that this is a primary aim of your project.

Second, I noticed that a number of studies were rejected because they included children outside the K-3 grade range. I found this somewhat unusual, for the children that are most in need of effective instruction are those who have not learned to read by older ages. Because the effectiveness of DI has been so well demonstrated with regular students at the K-3 grade range, the more recent focus on other ages seems only greater evidence of its worth. This is especially so in the case of special education students and those with behavioral and other issues. Thus, I couldn't understand the rationale of omitting all the studies that targeted these most needy students. Any explanation you might have of this

decision would also be greatly appreciated.

Third, and perhaps most important, as I read studies more closely I found that the Clearinghouse's interpretation of some of the studies didn't seem to accurately reflect the actual content of the reports. I know that these reviews are often done by staff members or graduate students who might not be familiar with all of the intricacies of social research. Thus, I assume that we can send corrections and they will be added.

I look forward to hearing from you, either through e-mail (jeans@uoregon.edu) or by phone (541-346-5005). I'm sure that we share the ultimate goal of helping all students – no matter what their background – develop academic skills and appreciate your appreciate your help.

Sincerely,
Jean Stockard, Professor Emerita

NOTICE:

This message is for the designated recipient only and may contain privileged or confidential information. If you have received it in error, please notify the sender immediately and delete the original. Any other use of this e-mail by you is prohibited.

Jean Stockard, Professor Emerita
Department of Planning, Public Policy, and Management
University of Oregon
Eugene, Oregon 97403-1209
phone: 541-346-5005
fax: 541-346-2040

From: What Works
Sent: Monday, June 29, 2009 11:31 AM
To: 'jstockard@nifdi.org'
Subject: What Works Clearinghouse (WWCPC-1322)

Dear Dr. Stockard,

Thank you for contacting the What Works Clearinghouse (WWC). Your submission of "Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System" was received on October 6, 2008.

Please be advised that the order in which interventions are reviewed by the WWC is based, in part, on the availability of studies examining the effectiveness of an intervention. Thus, one of the reasons this study examining the effects of Direct Instruction mathematics has not yet been reviewed is that there are other interventions with a greater number of studies examining their effectiveness. If there are additional studies about the effectiveness of Direct Instruction mathematics, please submit them to the WWC.

It is important to note that the WWC website does not list all products/interventions that are the subject of research submitted to the WWC, but rather interventions for which we have completed full literature searches and reviewed all eligible studies against evidence standards.

We hope that you will find this information helpful in understanding the study submission and WWC review process.

What Works Clearinghouse

The What Works Clearinghouse was established by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education. For more information, please visit http://ies.ed.gov/ncee/wwc/.

-----Original Message-----
From: jstockard@nifdi.org [mailto:jstockard@nifdi.org]
Sent: Monday, June 22, 2009 3:40 PM
To: info@whatworks.ed.gov
Subject: IES Website: Contact Us: Check on the Status of a Submission,
Reference ID Number: 109747337

info@whatworks.ed.gov, this email was automatically sent through the Contact link on the WWC website.

From: jstockard@nifdi.org

Message: Might you be able to tell me the status of submission # 29487267, please?
It was a study of Direct Instruction mathematics.
Thank you.
Jean Stockard, Ph.D.