Research Syntheses of Direct Instruction Outcomes:

A "Tertiary" Review

NATIONAL INSTITUTE FOR DIRECT INSTRUCTION

June 2011

Cristy Coughlin

This technical report is a pre-publication draft of a chapter to be published in
Lloyd, J., Carnine D., Slocum, T., Watkins, C. (Eds.) *Does Direct Instruction Deserve Status as an Evidence-Based Practice?* ADI Press, 2011.

Research Syntheses of Direct Instruction Outcomes:

INTRODUCTION

Research on the effectiveness of Direct Instruction (DI) has spanned over 40 years, consistently providing support for the assertion that all children can learn. Studies involving Direct Instruction curricula have been conducted with a wide variety of populations, in different settings, within numerous subject areas. One of the most significant demonstrations of the effectiveness of Direct Instruction programs was provided in Project Follow Through. With the intent to obtain evidence about the most effective educational approaches for teaching economically disadvantaged children, the United States Office of Education funded and implemented Project Follow Through from 1968 and eight years onward, comparing effects of various instructional models implemented across 170 communities in the U.S. The Direct Instruction model was the only model in which significant, positive outcomes were demonstrated in the areas of basic skills, problem solving and thinking skills, as well as affective measures of self-esteem. While the intent of the project was to identify educational interventions that worked, the results of Project Follow Through failed to affect policy on a large scale, and resulted in no push by the federal government to adopt those interventions that were shown to be most effective, including Direct Instruction (Engelmann, 2007). Project Follow Through was the largest educational experiment the country has undertaken, and no other study with a similar scope has been conducted to address the question of which instructional models are most effective in teaching children. However, the field of education continues to be concerned about identifying *evidence-based practices* in schools, and support for this movement has garnered increased support over the last decade not only from practitioners and scholars, but has also been propelled forward by national policies and laws, such as No Child Left Behind Act (U.S. Department of Education, 2002), that have been put into place to accelerate the adoption of practices and curricula that are “scientifically-based”. Yet, there still remains a lack of consensus on what it takes for a practice to be considered evidence-based. To guide decision-making at the national, state, and local level, it is necessary for educators and policy makers to try to make sense of the evidence attesting to the effectiveness of instructional practices, which can be an arduous task to undertake. Because it is infeasible to conduct another experiment with the same magnitude as Project Follow Through, syntheses of available research are typically used to guide conclusions about the effectiveness of practices. Fortunately, as Direct Instruction programs are characterized by a long history, wide implementation, and extensive research, the effectiveness of DI curricula is well-suited to be scrutinized through the lens of rigorous research syntheses.

THE ROLE OF RESEARCH SYNTHESES

Adoption and dissemination of educational interventions and curricula has been historically driven by theories on how children learn, the political climate of schools and districts, and the popularity of faddish approaches. Though unsupported by substantial evidence, schools have been quick to implement a program before it is evident that the program will likely impart change in their students’ achievement. However, due to recent educational reform efforts and the enactment of federal policies, a shift towards identifying and promoting the use of "evidence-based" or "research-validated" practices and programs in schools has taken place. Legislation such as the No Child Left Behind Act and reauthorization in 2004 of the Individuals with Disabilities Education Act has served as a catalyst for these efforts, and today schools are held more accountable for student achievement than ever before. While this accountability is designed with American children in mind, pressure on schools and educators has intensified and, given the influence of limited funding and a multitude of available curriculum materials claiming to be "research-based", making sense of the evidence on the effectiveness of a program and deciding on which curriculum to implement is a daunting task.

Research syntheses are one method of integrating the empirical evidence on educational interventions' effects, making findings and conclusions more palatable and accessible to a wider audience. The medical field has embraced the use of research syntheses to inform practice, going so far as to establish a consortium of researchers assigned with analyzing the cumulative evidence existing in hundreds of areas within medicine and public health (Cochrane Collaboration, 2008). Similarly, the American Psychological Association organized the Presidential Task Force on Evidence-Based Practice in 2005, called on to undertake "…the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (American Psychological Association's Presidential Task Force on Evidence-Based Practice, 2006). The field of education has sought to create councils to achieve similar goals by distinguishing effective educational interventions from those that are less effective, including, for example, the National Reading Panel and the National Mathematics Advisory Panel. Relying on the use of systematic reviews and meta-analyses, these organizations are crucial to the identification and dissemination of research-validated practices.

Types of Research Syntheses

While research syntheses intend to review and integrate the cumulative findings of individual studies within the same topic area, there are different approaches a researcher can take to achieve this goal.

Systematic reviews. A *systematic review* is a broad term used here to describe a literature review seeking to collect and code variables of all primary research conducted in a given area in an explicit and standardized manner. Conducting a systematic review of literature requires researchers to apply the same level of rigor as when conducting primary research studies. This involves researchers applying techniques to minimize the difference in results obtained from retrieved studies and the results of studies that were not found and consistently applying specific criteria to determine which studies will be included and excluded from the review, regardless of whether or not results support the researchers' hypotheses. Systematic reviews are widely used in the medical fields to guide decision and policy-making about evidence-based treatments, integrating the results of randomized trials of healthcare interventions and treatments. To conduct a systematic review, researchers proceed though a sequence of steps, much like the stages that characterize primary research studies, generally including the following steps: establishing a research question, searching the literature, gathering information from studies, evaluating the quality of studies, and analyzing the outcomes (Cooper, 2010). There have been several systematic reviews of the Direct Instruction research base, synthesizing research on the effectiveness of Reading

Mastery (Schieffer, Marchand-Martella, Martella, Simonsen, & Waldron-Soler (2002), Corrective Reading (Prychodzin-Havis et al., 2005) and DI math programs (Przychodzin, Marchand-Martella, Martella, & Azim, 2004), in addition to a synthesis of the effectiveness of using DI with special education populations (Kinder, Kubina, & Marchand-Martella, 2005).

Meta-analyses. Like systematic reviews, a *meta-analysis* necessitates a thorough, comprehensive review of all of the research completed within a specific topic area and calls for a systematic approach to coding and evaluating the results of this research. Typically utilized in conjunction with systematic reviews, meta-analysis involves quantitatively integrating the results of numerous studies, shifting away from the use of statistical significance for judging program effects, typical of primary research studies, to emphasizing examination of the magnitude and direction of results through an effect size estimate. Meta-analysis is useful for interpreting a research base wherein the effect of an intervention has been examined across a broad range of settings, populations, and methodologies, when contradictory findings may exist, and when it is desirable to explore hypothesized mediators and moderators of effects (Rosenthal, 1984); moreover, especially relevant to the identification of evidence-based practices in education, "…meta-analytic renderings of research domains move the decision process beyond the false assumption that a single study, no matter how 'perfect', can provide the basis for sound decision" (Kavale, 2001, pg. 262), and therefore, by quantitatively analyzing the *cumulative* results of a series of individual studies, more confident conclusions can be drawn from syntheses of research (Cowan, 2004).

Effect sizes vs. statistical significance. Because studies use different measures and dependent variables, meta-analyses typically use effect sizes as common metric to compare results across studies. Effect sizes are referenced against a normal distribution and can be interpreted as a conclusion about the effect of one particular intervention when compared to another intervention. Rather than interpreting results in terms of statistical significance, effect sizes convey whether results have practical significance—meaning that the effect of the intervention had a real, important impact on children's learning. Cohen's *d,* which is the standardized difference between two means, is often calculated to estimate effect size. Cohen (1988) established guidelines for interpreting the magnitude of these effect size estimates, so that values of 0.20 are considered "small", 0.50 are "medium" and 0.80 are "large". Within the field of education, an effect size of 0.25 is generally considered as "educationally significant" (Tallmadge, 1977).

Alternatively, statistical significance conveys that the obtained results are not likely to have occurred by chance alone and that differences between groups are unlikely to be a product of sampling error. Whether or not differences are statistically significant relies largely on sample size; the larger the sample size, the closer differences approach the level of statistical significance. However, when the intent is to integrate results from multiple studies and because sample size can vary so much within a literature base—which is very true for studies on Direct Instruction programs, with samples ranging from under 10 students up to more than 30,000 students—relying on statistical significance to judge effects is biased against those studies with fewer participants. Additionally, unlike effect size, statistical significance provides little insight into the magnitude or importance of a program's effect. Still, some research syntheses attempt to synthesize results based on statistical significance, utilizing simple polling or vote-counting, where the number of statistically significant differences reported across all studies included in the synthesis is tallied. Specific to the DI literature base, most of the systematic reviews discussed in this chapter employ this vote-counting approach of quantitatively synthesizing results, whereas the three oft-cited meta-analyses of DI research that will be summarized, Adams & Engelmann (1996), White (1988), and the comprehensive school reform meta-analysis conducted by Borman, Hewes, Overman, & Brown (2003), take advantage of effect size estimates to capture program effects.

Best Evidence Syntheses. Finally, the third method of synthesizing results of DI program implementation is based on the approach coined *best evidence synthesis* by Slavin (1986), described as a set of standards for meta-analyses that give more weight to high-quality studies, or studies that employ rigorous methodologies, minimize bias, and have high internal and external validity (Slavin, 1995). In this type of synthesis, program effects are categorized by level of evidence, incorporating effect size estimates with a judgment of study quality, typically accomplished by excluding studies that fail to meet minimum standards of methodological quality. Not without controversy, best evidence syntheses have been recently used at the national level in efforts to synthesize research on educational programs, spearheaded by organizations sponsored by the U.S. Department of Education, such as the What Works Clearinghouse (WWC) and Best Evidence Encyclopedia, as well as other independent organizations, including the American Institutes for Research. The mission of these organizations is to provide recommendations about the strength of evidence supporting the effectiveness of educational programs, taking into account both the magnitude of an intervention's effects and the quality of methods employed by primary studies to determine these effects.

The differentiation between these three categories of research syntheses is articulated in order to provide a structure for exemplifying the breadth and variety of research available on Direct Instruction, and is an artifact only for organizational purposes of this chapter. For more in-depth discussion of the fundamentals of the various approaches to synthesizing research, it is advisable to consult more comprehensive sources, such as Cooper (2010), Hedges & Olkin (1985) and Slavin (1986)/(1995). The purpose of this chapter is to serve as a sort of "tertiary review", summarizing and discussing the results of several oft-cited research syntheses, including systematic reviews, meta-analyses, and – a product of more recent efforts nationwide to identify effective programs – "best evidence syntheses" (see Table 1). Further, differences and similarities in the conclusions drawn from multiple research syntheses, which are distinguished by the various approaches and methodologies used to integrate research findings, will be highlighted.

Stages of Research Syntheses and Procedural Variations

Several specific models for conducting research syntheses have been developed (e.g., Jackson, 1980; Cooper, 2010), but generally follow the same basic steps, mirroring the progression of activities in primary research studies. As Cooper (2010) points out, every decision made along each step of the research synthesis process has a potential impact on the conclusions that are drawn upon completion. As methodological decisions are made along the way, it is possible for variation to occur in the conclusions drawn by different researchers. Exemplifying this influence, we can examine some of the procedural variations identified by Cooper (2010) and Shadish, Cook, and Campbell (2002) and the accompanying stages that decisions about these variations might take place. First, at the initial stage of conducting a research synthesis, a research question is formulated, and variables and relationships of interest are identified and defined. The greater extent to which these definitions vary across syntheses, the more likely differences in conclusions will result. For example, a researcher who defines Direct Instruction as "a set of effective teaching practices" would draw very different conclusions from a researcher who defines Direct Instruction as "any program developed and authored by Siegfried Engelmann".

Once a research question is fully defined, a comprehensive literature search follows, wherein a researcher will seek to locate as many relevant studies as possible. At this stage of the synthesis, a researcher's selection of which search terms to use and what sources to reference can result in a different set of studies synthesized, therefore leading to different conclusions. One synthesis which, for example, utilizes a thorough search through databases containing dissertation studies and other unpublished work, and another synthesis which restricts inclusions to studies only published in peer-reviewed journals might result in conclusions drawn from two different bodies of literature. Once all relevant studies are obtained, necessary information is extracted from each study, as trained individuals classify and report study features and findings. When obtaining information from studies, the intent is to retrieve data relevant to the research question, including any features that are hypothesized to influence outcomes (i.e., mediators and moderators of an effect). It is therefore necessary for researchers to thoughtfully select which characteristics will be coded, keeping in mind that if too little data is retrieved, there exists the possibility that important influences on cumulative effects will be missed, while balancing the need to classify variables into categories that are as discrete as possible. Given this task, the impact of variation in study characteristics selected for coding, how these characteristics are conceptualized and distinguished, and differences in training for individuals responsible for coding studies may influence cumulative conclusions drawn from the synthesis. Even if the exact same set of studies was retrieved when the literature base was searched, small differences in the way that participants in these studies were described in the coding scheme, for example, with one synthesis capturing disability status in many categories and the other collapsing students with disabilities into one category, could lead to the first synthesis capturing an important influence on effects and the second synthesis altogether missing this influential factor.

After studies are collected and coded, the researcher must next apply exclusion criteria, separating studies that will be analyzed within the synthesis from those that don't meet all requirements for inclusion. In meta-analyses, researchers will typically adhere to some criteria for study design and methodology when excluding studies, so that the type of methodology employed in a study is congruent with the larger research question of the synthesis. For example, if the question of interest is whether Curriculum X is more effective than Curriculum Y, then it is logical for a researcher to only be interested in findings from studies with a comparison group, thus excluding studies that only examine change in achievement before and after implementation of Curriculum X, without comparing these changes to students who had received instruction with a different curriculum. Other exclusionary criteria that may be applied at this stage of the synthesis might be related to sample characteristics (e.g., excluding studies of students in a certain age group if the research question is only interested in effects on elementary-aged students), details about the implementation of the intervention (e.g., only including studies that report data on fidelity of implementation), or accessibility of necessary quantitative information (i.e., studies that report requisite means, standard deviations, and sample size information to calculate effect sizes). Finally, when comparing and contrasting results across studies, a researcher has to make decisions about which procedures to use when analyzing and integrating cumulative findings and how to interpret and discuss the importance of these findings. As noted earlier, authors of research syntheses on Direct

Instruction programs have used both effect size estimates, vote-counting of statistical significance, and simple tallying of positive findings to quantitatively integrate results; these different approaches are not just a matter of preference, but are designed to capture distinct aspects of the literature base, namely, the average magnitude of a program's effects, the likelihood of differences between experimental and comparison groups being attributed to chance alone, and the proportion of studies reporting general positive findings for the experimental group. Application and interpretation of these various techniques will likely prompt divergences in the collective conclusions made.

These procedural variations are just a snapshot of the range of decisions that a researcher makes when conducting a research syntheses. When reflecting on conclusions asserted from a synthesis of research, it is important to consider how these conclusions are a product of accumulating variations, and that each decision made by the researcher has contributed to their overall interpretation of findings amassed. If conclusions across syntheses conflict, a review of the procedural decisions made by authors of both syntheses may provide insight about factors that may have contributed to these differences, bringing to light decisions that may be questionable and areas where further investigation may be warranted.

Within the Direct Instruction literature base, systematic reviews, meta-analyses, and best evidence synthesis have been repeatedly used to integrate and make sense of the hundreds of studies conducted. Conclusions drawn from these syntheses have been generally very consistent, with systematic reviews and meta-analyses resulting in very similar conclusions about the positive effects of DI programs. The following section will summarize these conclusions, discussing results as to correspond to various populations, settings, and subject areas studied, and exploring the procedural variations that may have led to discrepancies from one synthesis to another.

DIRECT INSTRUCTION AND SPECIAL EDUCATION

Arguably the most extensively researched population within the DI literature base, there have been hundreds of studies published that have involved students with disabilities and students in special education settings. This can largely be attributed to the alignment of DI programs with the diverse needs of students comprising this population. With the reauthorization of the Individuals with Disabilities Education Act (IDEA) in 2004, schools are now required to provide specially designed instruction to students eligible to receive special education services. This "specially-designed" instruction refers to the adaptation of content or delivery of instruction to meet the unique needs of students eligible for special education services. These modifications might include adjusting the pace of instruction to ensure student mastery of concepts taught, matching instructional level to student skill level, identifying and prioritizing the most important concepts to teach, providing ample opportunity for student responding and teacher feedback, among others (Carnine, Silbert, Kameenui, & Tarver, 2004). Packaged together, these effective teaching practices, along with careful attention to the design of instruction and organization structure of material, provide the framework for all DI programs. Direct Instruction programs are not only characterized by effective teaching practices, but also by the efficiency of these practices, wherein the goal for students is to learn material in the minimum amount of time. Students are taught concepts and ideas that can be best generalized and applied to the widest possible range of examples that a student might encounter. Examples and non-examples are carefully selected and material is sequenced so that minimal confusion and a high rate of student success in acquiring new concepts and information are demonstrated. Relatively fast pacing, choral responding of students, teacher signaling, and scripted presentation contribute to rapid student progress through lessons. These features distinguish Direct Instruction programs as an efficient approach to teaching and learning, wherein students can acquire skills relatively quickly that will allow them to apply generalizable strategies to a range of possible items, including those not directly taught within the lessons (Watkins & Slocum, 2003). Because these features of DI curricula are inherent in the structured program design, their applicability to students with special needs is obvious, and when implemented in the context of special education, research has demonstrated that student achievement is consistently accelerated.

Adams & Engelmann conducted a meta-analysis of the literature on the effectiveness of DI programs in 1996, and analyzed data from 34 research studies that met their criteria for inclusion. A positive effect size was demonstrated in 32 of these 34 studies, with a mean effect size of .97 and over half of the studies included in the meta-analysis showing a large effect. Included in their meta-analysis were studies that examined DI within a variety of subject areas with participation from both general education and special education students. Breaking these results down by type of students, the mean effect size was .90 for special education students and .82 for general education students, indicating that on average, special education students receiving instruction with DI curricula scored nine tenths of a standard deviation higher than those students in the comparison group. As noted before, Cohen's (1988) guidelines for interpreting the relative magnitude of an effect size estimate can be used as one method for referencing effect sizes reported in meta-analyses, and an effect size of 0.90 exceeds Cohen's benchmarks of a "large" effect size of 0.80. Similarly, White (1988) conducted a meta-analysis on research examining effects of DI programs on special education populations only, and reported an average effect size of 0.84, also exceeding Cohen's benchmark of 0.80. This consistency in conclusions is especially impressive, given that only about half of the 25 studies that were included in White's meta-analysis were also synthesized in the Adams and Engelmann meta-analysis. Despite the span of 8 years lapsed between these two meta-analyses and differences in exclusionary criteria for determining eligibility of studies, the cumulative conclusion about the average effect of DI programs on special education students in the 25 studies integrated in White's synthesis differed by only .06 from the average effect size specific to special education students calculated based on the body of literature retrieved by Adams and Engelmann. To put these effect sizes in context, we can refer to Forness, Kavale, Blum, and Lloyd (1997), which report the results of a "mega-analysis" of meta-analyses within the field of special education and related services. Forness et al. reviewed 18 meta-analyses that integrated the effectiveness of various strategies, approaches, and programs in special education instruction, referencing White (1988) to obtain effect size information for Direct Instruction. Of the 18 interventions reviewed, DI's average effect size of 0.84 was the fourth highest reported, and was classified by Forness et al. as one of only seven "interventions that work".

Outside of the meta-analyses focused on Direct Instruction, other researchers have undertaken systematic reviews to capture effects of DI programs on students with disabilities and those in special education settings. Conducting a systematic review, Kinder et al. (2005) provide an overview of the overwhelmingly positive effects produced when DI programs are used with special education populations. The authors discuss the research base on students with high-incidence and low-incidence disabilities separately, describing a total of 45 studies across student disability categories. In order to be eligible for review, studies had to appear in peer-reviewed educational journals only, which excluded unpublished studies and those considered as grey literature, including dissertations, technical reports, and conference presentations. Criteria for experimental design were very inclusive, resulting in case studies, one-group pretest-posttest designs, and single-subject methodology represented in the articles selected for review. Rather than illustrating the magnitude of an intervention's impact by calculating effect sizes, Kinder et al. provide a narrative review of characteristics and findings of the studies selected through their systematic search of the literature base.

For the high-incidence disability population, in only 3 of the 37 studies (or 8% of the sample) did students who received instruction from non-DI methods score better on posttest measures than students receiving DI programs. These studies were conducted between the mid-1970s and 2005, and the majority of students who comprised this population were diagnosed with learning disabilities; 32% of studies involving students with learning disabilities also included students with behavior disorders, mild intellectual disabilities, other health impairments, and/or traumatic brain injury. The remaining studies examining effects of DI curricula on students with high-incidence disabilities included students with language or developmental delays, or students who were solely identified as "eligible" for special education. Six of the 37 studies reviewed took place within high school settings and 8 included preschool and kindergarten children, whereas the majority of studies primarily focused on elementary and/or middle school students. Across these studies, Direct Instruction programs were reported to be effective in increasing skills in areas of reading and spelling. Math and writing program effectiveness were supported with fewer studies, and, because Kinder et al. focused on examining statistical significance, the results obtained from language studies were inconclusive, with most studies showing few, if any, significant differences in post test scores between students receiving DI programs and those receiving instruction through other methods and curricula.

Eight articles involving students with low incidence disabilities were reviewed by Kinder et al. (2005). In half of these studies, students with mental retardation served as participants; the remaining four studies included students with traumatic brain injury, intellectual disabilities and autism, and students labeled as "educationally subnormal". All but one of these studies examined the effectiveness of Direct Instruction reading or language programs. As noted earlier, Kinder et al. (2005) included studies utilizing single subject methodology and less rigorous research designs in their review, including case studies and one-group pretest-posttest designs, both designs representing the methodology employed in all but one of the studies reviewed involving students with low incidence disabilities. The exception to this pattern is Maggs and Morath's (1976) investigation of effects of DISTAR Language, wherein students between the ages of 8-16 with moderate or severe mental retardation were randomly assigned to the DI group or a comparison group utilizing the Peabody Language Program. On all six dependent measures, students in the DI group scored significantly higher than the comparison group, with an average effect size of 1.80 across these measures.

Based on the results of these three syntheses, it is apparent that DI programs have resulted in positive outcomes for special education students. Despite the procedural variations across these three syntheses, including differences in quantitative estimate of effect (i.e., statistical significance, effect size, or general positive results), inclusion criteria for study design, and search procedures used, the results and conclusions drawn from these syntheses are remarkably consistent. For students with disabilities, IDEA's reauthorization requires that schools provide "specially designed" instruction, a mandate that can be accomplished though the use of Direct Instruction curricula; moreover, research on DI program effectiveness addresses the legislative call for schools to use evidence- or scientifically-based instructional practices in special education.

DIRECT INSTRUCTION AND SCHOOL REFORM

While there is ample evidence to the contrary, a wide-held misconception is that DI programs are *only* appropriate for low-achieving and special education students. This myth may have stemmed from the consideration of DI as a "basic skills" model in Project Follow Through. Or, because DI programs have been shown time and again to accelerate the learning of lower-performing students, who have presumably fell behind in other programs, they are often exclusively used with remedial and special education populations (Adams & Engelmann, 1996). Yet, when examining the research base investigating effects of DI programs with general education populations, it is apparent that this assertion has no basis. Studies conducted since the 1980s have shown that DI programs are, in fact, very effective when used with average- and high-achieving students (e.g., Ashworth, 1999; Gersten, Becker, Heiry, & White, 1984; Ginn, Keel, & Fredrick, 2002; Robinson & Hesse, 1981; Sexton, 1989; Tarver & Jung, 1995; Vitale & Joseph, 2008). Moreover, as recent attention has been directed toward evaluating school reform efforts, evidenced by the creation of the Comprehensive School Reform (CSR) Program, a component of the No Child Left Behind Act of 2001, the effectiveness of Direct Instruction implementation as a school reform model also refutes the false notion that DI is only useful within the realm of special education. Numerous studies have supported the implementation of DI school-wide (e.g., Carlson & Francis, 2002; Cross, Rebarber, & Wilson, 2002; MacIver & Kemper, 2002; O'Brien & Ware, 2002), and a meta-analysis conducted by Borman, Hewes, Overman and Brown (2003) concluded that research supports DI as an effective school reform model.

In this synthesis of research, Borman et al. sought to "…establish the extent to which each of the 29 models is supported by scientifically based research", considering the quality and quantity of supporting research as well as the statistical significance and positivity of findings. Studies employing a design in which a comparison could be used to calculate an effect were eligible for review, including quasi-experimental and experimental treatment-comparison group designs, in addition to one-group pretest posttest designs, which ended up characterizing the methodology used to obtain almost half of the outcomes analyzed. Each of the 29 CSR models was classified into one of four categories: *Strongest Evidence of Effectiveness, Highly Promising Evidence of Effectiveness, Promising Evidence of Effectiveness,* and *Greatest Need for Additional Research.* The Direct Instruction model was represented by 49 individual studies – the greatest number of studies included in the meta-analysis for any one CSR model – involving over 40,000 students, and was one of three CSR models to be classified in the *Strongest Evidence of Effectiveness* category. An average effect size of 0.21 was calculated across the 49 studies, which included studies with all types of research designs, and an average effect size of 0.15 was computed based on studies that employed some form of control group (i.e., excluding one-group pretest-posttest designs). To be classified under the *Strongest Evidence of Effectiveness* category, the Direct Instruction model was represented by 10 or more studies, a criterion used to indicate generalizability to the population of schools likely to implement CSR models, and was shown to have statistically significant and positive effects on student achievement in studies using comparison groups against which to reference gains, including at least 5 third-party studies not conducted by the model's developer.

The findings of the Borman et al. (2003) meta-analysis serve to inform the field of education about the relative effects of recent CSR efforts, but also reiterate conclusions of earlier syntheses of research. In 1998, the American Federation of Teachers identified Direct Instruction as one of six promising school-wide reform programs (American Federation of Teachers, 1998), and a review of research on 24 school reform approaches prompted the American Institutes of Research (AIR) to consider Direct Instruction as one of only three approaches they classified as having "strong evidence of positive effects on student achievement" (American Institutes for Research, 1999). This rating was given to models supported by at least four studies, utilizing rigorous methodologies, which demonstrate positive effects on student achievement, with at least three of these studies reporting statistically significant differences in favor of the model of interest.

Also interested in the impact of DI programs on the achievement of students in the general education setting, Adams & Engelmann's (1996) meta-analysis provides information about the effectiveness of DI programs compared to other instructional approaches. As noted previously, a mean effect size of .84 was calculated for studies involving students receiving instruction in the general education setting, indicating that, on average, DI curricula results in a "large" effect for general education students. Inclusion criteria related to design type in Adams & Engelmann's meta-analysis resulted in a group of studies that compared instruction with DI programs to instructional with an alternative approach, so that the body of research synthesized was comprised only of quasi-experimental and experimental designs in which effect sizes could be calculated to represent the relative effectiveness of experimental and comparison curricula. Because of this exclusionary provision, the studies that were synthesized in this meta-analysis can be considered as having a relatively high level of rigor in methodology. Further advantageous, Adams & Engelmann reported effect sizes as they correspond to categories organized around population of interest, methodological variables, and subject area, or, in other words, tested the influence of hypothesized moderators and mediators on effects. Discussion on differential effects of these moderating variables is useful for shining light on the situations and environments that have been shown to be most conducive for evidencing positive effects of implementation. Pertinent to policy makers and educators, decisions about curriculum adoption in various subject areas can rely on research syntheses' conclusions about a particular model's success at affecting achievement in a given subject. In addition to Adams and Engelmann's meta-analysis, which provides effect sizes that account for DI's cumulative success in the areas of reading, language, and math, as well as other subjects, several systematic reviews have complemented this analysis of the literature, reviewing research on DI programs in a range of subject areas and with a variety of students.

DIRECT INSTRUCTION AND READING

Compared to other subject areas, DI reading programs have received the most research attention, documenting positive effects on a wide variety of students and evidencing alignment with empirical research on reading instruction. In 2000, the National Reading Panel, a federal effort commissioned to identify elements of effective reading instruction, released a report outlining the conclusions that were made based on an extensive review of research evidence of thousands of empirical studies in the area of reading. Components consistently shown to result in gains in reading achievement were discussed, and recommendations for teaching reading were included. Salient features of Direct Instruction reading programs were among the recommendations provided, including an emphasis on the importance of explicit instruction in phonemic awareness and phonics, ample opportunity for guided reading practice, and corrective feedback (Institute of Child Health and Human Development); these recommendations reinforce the notion that DI curricula are based on empirical research on how best to teach children to read.

In contrast with the NRP report, the Adams & Engelmann (1996) and White (1988) meta-analyses were less concerned with identifying which independent instructional features are closely linked to student achievement, but were rather interested in exploring the magnitude of effects demonstrated when these features are packaged and sequenced into a comprehensive program for teaching reading. Across the studies included in Adams & Engelmann's meta-analysis, an average effect size of 0.69 for reading outcomes was reported, indicating that following instruction with Reading Mastery, Corrective Reading, and other DI reading programs, students scored, on average, 0.69 of a standard deviation higher than students whose reading instruction was based on some other approach. Providing an alternative method to calculating effect sizes as a means of synthesizing results, Schieffer et al. (2002) and Przychodzin-Havis et al. (2005) conducted systematic reviews of research on effects of Reading Mastery/DISTAR Reading and Corrective Reading, employing "vote counting" of statistical significance and mean score differences. Schieffer et al. included a total of 21 studies that compared implementation of Reading Mastery and DISTAR Reading to other reading approaches and reported the percentage of studies that favored the use of DI programs, based on statistically significant differences between DI and comparison programs on measures of reading achievement. This polling resulted in 67% of studies in their sample favoring Reading Mastery/DISTAR Reading, 19% of studies reporting no significant differences, and 14% favoring comparison reading programs.

Analyzing the research base on Corrective Reading, Przychodzin-Havis et al. (2005) simply tallied the number of studies that reported positive findings from using Corrective Reading, including studies representing a whole myriad of experimental designs, from studies that compared DI to other approaches as well as studies that only investigated differences from pre- to post-test with one group of students receiving instruction with Corrective Reading. Przychodzin-Havis et al. concluded that 93% of studies reviewed "found positive results for students who were taught using Corrective Reading" and that "…only one study noted greater effects with another intervention over Corrective Reading." One of the greater contributions of this review was the descriptive discussion of study characteristics, providing an overview of the experimental designs, populations, settings, and type of dependent measures that comprise the range of research conducted on the Corrective Reading curriculum. This information is useful for researchers by identifying areas for future investigation and for educators by exemplifying the conditions under which Direct Instruction programs have been effective. Overall, based on Przychodzin-Havis et al., studies on Corrective Reading were most likely to: employ a pre-experimental design (e.g., one group pretest-posttest design), involve elementary and/or middle school students and students with disabilities as participants, take place in special education settings, and utilize standardized assessments to evaluate outcomes.

In Adams & Engelmann's (1996) meta-analysis, the effect size estimate calculated for reading takes into account studies with participants representing both general education and special education populations; White (1988) provides an effect size estimate, based on 13 studies involving DI reading programs, of 0.85, representing the average magnitude of effect on students with disabilities and students receiving Direct Instruction curricula in special education settings. Further, to capture the impact of instruction on specific reading skills, effect sizes were calculated when outcomes on decoding and comprehension measures were differentiated in primary studies, with effect sizes of 0.64 and 0.54 reported, respectively. Kinder et al. (2005), also focusing on special education students, reviewed a total of 32 studies involved DI reading programs, in which students with high incidence disabilities comprised the sample of 81% of these studies. When DI programs were compared to other approaches to reading instruction, students with high incidence disabilities in the DI group typically outperformed students in the comparison group. In only 10% of Reading Mastery studies did the non-DI group score better on posttest measures, while students receiving Corrective Reading performed significantly better than comparison students in 88% of CR studies. Studies interested in students with low incidence disabilities employed designs that allowed investigators to explore gains from pre- to post-test, but involved no comparison group against which gains could be referenced; in this sample, authors reported that students with intellectual disabilities demonstrated marked improvement in reading ability following instruction with Corrective Reading and DISTAR Reading programs, and that knowledge gains were not limited to sight word recognition alone, but also resulted in acquisition of more sophisticated skills and strategies necessary to decode unfamiliar words. Lastly, Schieffer et al. (2002), specifically examining statistical significance as an indicator of effect, reported inconsistent results based on the studies involving special education students included in their review, with three studies reporting statistically significant differences in favor of the DI group on multiple posttest measures, five studies reporting few or no significant differences, and one study demonstrating statistically significantly higher scores for the comparison group.

Exploring conclusions drawn from a third approach to integrating the research on DI reading programs, a report on beginning reading instruction by the What Works Clearinghouse, exemplifies a "best evidence synthesis." Giving more weight to studies employing rigorous methodologies, program effects are judged by examining both the quality of a study and the effect size estimate calculated. In the What Works Clearinghouse intervention reports, each DI program is reviewed separately, so that judgments of evidence are relevant only to the curriculum of focus. In the WWC's beginning reading report, *Reading Mastery* was classified under the category of "no studies meeting evidence standards". While 61 studies conducted between 1985 and 2007 were identified and considered for eligibility, none of these studies' features met WWC criteria for inclusion, with study design eliminating almost all of these 61 studies. Because every study was excluded based on methodology, WWC was unable to make any conclusions about the effectiveness of the program. This particular situation exemplifies the dilemma that occurs when researchers identify an abundance of exclusion criteria. While it is true that the research that is accumulated is more homogenous in terms of design characteristics, increasing the validity of conclusions that are drawn from comparing results obtained from these studies, excessive exclusion criteria does a disservice to the consumers of these research syntheses, misrepresenting the extent of research on programs and failing to present the whole picture. Some of the procedures for selecting studies have been called into question, such as the cut date of 1985, consideration of variables as confounding, and the de-emphasis on implementation fidelity, and the decisions made by researchers responsible for WWC reports have been criticized for being arbitrary and preferential to particular types of programs (Engelmann, 2008; Stockard, 2010). As the WWC is called on to identify "what works" in education, some have questioned the emphasis placed on methodological "rigor", which, as evidenced here, may supersede the focus on capturing the *extent* of evidence as the purpose of research synthesis.

The influence of the procedures that researchers utilize when conducting meta-analyses, systematic reviews, and best evidence syntheses is evident, as demonstrated by variation in the conclusions drawn from these syntheses on DI reading programs. As consumers of these syntheses, it is crucial for educators and policy makers to understand the implications of these various procedural variations. Because research syntheses are intended to help make sense of large bodies of research, it is important for persons responsible for disseminating the conclusions drawn from the syntheses to clearly identify what these conclusions represent, how to interpret them in practice, and what limitations exist, continuously keeping in mind the intent of the synthesis in the first place.

DIRECT INSTRUCTION AND OTHER SUBJECT AREAS

Direct Instruction Mathematics Programs

Much like in the area of reading, there has been a recent push by the federal government to identify effective practices instructional programs in the area of mathematics. Examining children's math skills nationwide, in 2001, the National Center for Education Statistics revealed that only 26% of 4th grade students, 27% of 8th grade students, and 17% of 12th grade students were considered to score at the *proficient* level in math. Compared to other countries, the mathematics performance of American children ranked in the bottom half of 41 participating countries (Gonzales et al., 2008). As a result of this startling statistics, the National Mathematics Advisory Panel (NMAP; 2008), modeled after the National Reading Panel, and the National Council of Teachers of Mathematics (NCTM; 2000) have worked over the last 10 years to provide recommendations to educators and policy-makers about how mathematics should be taught in schools across the country, providing guidance through the development of standards for content that should be covered and identifying which curricula are best aligned with these standards and most effective at improving math achievement. The structure and content taught in DI math programs, which include DISTAR Arithmetic, the original DI mathematics program implemented as part of the DI model in Project Follow Through, Corrective Mathematics, Connecting Math Concepts, and, the most recent addition, Funnix, are in accordance with recommendations put forth by NCTM and NMAP, including the call for scientific validation of curriculum impact on student outcomes (Przychodzin et al., 2004).

Consulting research syntheses as indicators of whether Direct Instruction math programs are scientifically-validated, both meta-analyses discuss findings for mathematics, parceling apart studies by topic area and calculating a separate effect size for mathematics-related outcomes. Based on 33 comparisons of math achievement scores across studies included in the Adams & Engelmann (1996) meta-analysis, an effect size of 1.11, in favor of Direct Instruction programs, was reported, with most of these comparisons involving general education students. Well above Cohen's "large" benchmark, an effect size of 1.11 indicates that, on average, students receiving their math instruction with a DI program score over one standard deviation higher than students who receive instruction with a comparison curriculum.

For special education students, the average effect size for math programs is lower, but still indicating a "medium" effect at .50, albeit only four studies on mathematics were included in White's (1988) meta-analysis. Similarly, systematic reviews conducted by Kinder et al. (2005) and Przychodzin et al. (2004) echoed the preponderance of DI math studies focusing on students in general education settings. Przychodzin et al. reviewed results of 11 math studies, with 8 out of these studies exclusively examining the effectiveness of DI math curricula on general education students, 2 studies involving only students with disabilities, and the remaining study, also the only study involving math reviewed by Kinder et al., investigating outcomes of students with and without developmental delays. Overall, only one study in this review reported better gains made by the comparison group; Young et al. (1990) compared instruction with DISTAR Arithmetic to an adaptation of the DI program, incorporating modifications based on Discrimination Learning Theory, and found that the adaptation resulted in greater gains of a small sample of students with moderate intellectual disabilities when assessing effects with a single subject design. As Przychodzin and colleagues' systematic review of Corrective Reading research incorporated research designs with varying degrees of rigor in methodology, inclusion criteria for their review of DI math programs was similar; only two studies sought to achieve group equivalence at pretest and randomly assign participants to groups, four studies included a comparison group, but groups were not manipulated to match on demographic or achievement variables and no random assignment at the student level was employed, and the rest of the studies in the review compared achievement before and after instruction with DI math programs without the use of a control group, either utilizing single subject methodology or a one group pretest-posttest design.

Compared to reading, DI math programs are much less researched, and as shown in these aforementioned research syntheses, the research that has been conducted has typically involved students in general education. Because only a handful of studies have investigated effects of DI math programs with students with disabilities, even though most reported promising results, future research would be helpful to strengthen confidence in overall conclusions about effectiveness with this population.

Direct Instruction Language Programs

The area of language was a focus of earlier research on Direct Instruction programs, and results from Project Follow Through in the area of language illustrate the initial evidence supporting the effectiveness of DI language programs. When results of the DI model was compared to results of the average Follow Through model, effect sizes for language were very large, with scores of students who participated in the DI model one standard deviation higher than the average score of students in all other Follow Through models. Further, students receiving the DI model demonstrated statistically significantly higher performance on all three language measure when compared to *every* other model (Bereiter & Kurland, 1981-1982). Since Follow Through, DI language programs continued to be studied, and in 1996, when Adams & Engelmann completed their meta-analysis of the DI research base, the cumulative effects of these programs were assessed. Based on 39 comparisons calculated from studies included in this meta-analysis, Direct Instruction language programs resulted in an effect size of 0.49, indicating that, on average, students in these studies scored one half of a standard deviation higher than comparison students—an educationally significant difference. As Adams & Engelmann included studies involving both general education and special education students, this effect size is representative of both populations. Referencing White's (1988) meta-analysis, three studies of DI language programs with special education students resulted in an average effect size of 1.21. In contrast to these findings, Kinder et al. (2005), relying on statistical significance to judge effects, concluded that the cumulative research on the effectiveness of DI language programs with students in special education was inconclusive. Based on the studies selected for review, most studies reported few, if any, significant differences in post-test scores between students receiving DI programs and those receiving instruction through other methods and curricula. Kinder et al. (2005) identify that the lack of significant differences between the DI and comparison groups is of particular interest, given the overwhelming evidence of effectiveness of DISTAR Language from Project Follow Through. Kinder et al. (2005) assert that three main differences between participants in the studies included in their review and the sample comprising Project Follow Through contributed to this discrepancy in effectiveness. Students involved in four of the five language-specific studies reviewed by Kinder et al. (i.e., Cole & Dale, 1986; Cole, Dale, & Mills, 1991; Cole, Dale, Mills, & Jenkins, 1993; Dale & Cole, 1988) all had been: 1) identified with disabilities, 2) attended school part-time, and 3) were younger than students in the Project Follow Through sample. As Kinder et al. explains, these characteristics undoubtedly resulted in students receiving less instruction in DISTAR Language than students in Project Follow Through, who attended school full-time, were older, and more likely to be typically developing and conceivably more prepared for instruction in the curriculum.

Given the range of results from these three syntheses, it is difficult to reach conclusions about the relative effects of DI language programs on general education and special education populations. Because of the limited sample size in White's meta-analysis, and the fact that Adams & Engelmann calculated the effect size for language programs by combining studies involving special education and general education populations, it is not possible to discern which populations benefit most from Direct Instruction in language. Yet, it is apparent that, despite the lack of significant differences reported in Kinder et al's review specific to special education populations, these syntheses indicate that DI programs, on average, produce an educationally significant impact on language skills of both typically developing students and students with disabilities.

Direct Instruction Spelling and Writing Programs

Though most of the research on Direct Instruction program effectiveness has primarily been focused in the areas of reading, math, and language, research syntheses have also sought to review research on DI programs in other subject areas. Kinder et al. (2005) identified five studies that examined DI spelling programs, including *Spelling Mastery* and *Morphographic Spelling* (now called *Spelling Through Morphographs*), or writing programs, including *Language for Writing* and *Reasoning and Writing.* Elementary-aged students with high-incidence disabilities, most often learning disabilities, participated in these studies. The only study to use an experimental design, Darch and Simpson (1991) compared effects of Spelling Mastery to spelling instruction incorporating visual imagery and reported that the Spelling Mastery group performed statistically significantly better than the comparison group on all probes, posttest measures, and standardized achievement subtests administered. The remaining four studies employed one group pretest-posttest or single subject designs to assess effects, with all reporting positive results when DI spelling or writing programs were implemented.

Simonsen & Gunter (2001) conducted a systematic review of research on spelling instruction and research examining the effectiveness of DI spelling programs. They concluded that combining components of Direct Instruction delivery, such as systematic error correction and sequenced lessons, with teaching spelling through phonemic (i.e., teaching spelling through letter-sound correspondence), whole word (i.e., explicit instruction and practice in spelling irregular words) and morphographic approaches (i.e., teaching through combining spelling of small units of meanings, or morphographs), typically resulted in meaningful gains in spelling performance. The authors indicate that few comparisons between various approaches to spelling have been researched, and that, of the studies existing, most have compared Spelling Mastery or Spelling Through Morphographs to an alternative approach and have reached the conclusion that DI spelling programs were more effective than these comparison strategies. Additionally, a handful of studies have simply examined spelling gains following instruction with DI programs, reporting positive results for students in general education, elementary and middle school, and for students experiencing difficulties in the area of spelling.

Referring back to the two meta-analyses of the DI literature base, Adams & Engelmann (1996) calculated an effect size of 1.33 in the area of spelling, but reported no information about writing; similarly, White (1988) only reported effect sizes in the area of reading, math, and language. Often, DI spelling and writing programs are researched in conjunction with other DI programs as part of whole school implementation of the Direct Instruction model (e.g., Cross, Rebarber, & Wilson, 2002; Grossen, 2004). As the authors of these research syntheses assert, while results of these studies of DI writing and spelling programs are promising, more research is warranted in both areas.

CONCLUSION

As the movement towards identification and dissemination of "scientifically-validated" or "evidence-based" educational interventions continues to gain momentum, educators and policy makers will be required to make sense of the research that exists on the effectiveness of a spectrum of programs and curricula. Research syntheses prove to be a valuable resource in guiding decision-making at the national, state, and local level, as more confident conclusions can be drawn from examining the cumulative results of studies. Direct Instruction has the advantage of a long history of research on its effectiveness and widespread implementation in American schools. Because of this history, an enormous literature base has amassed, examining effects on a wide variety of students, including students in the general education setting and students with disabilities. Research syntheses have been conducted to integrate the results of these studies, and despite a range of procedural variations that have inherently influenced the cumulative findings of these syntheses, there has been remarkable consistency in conclusions drawn about the effectiveness of Direct Instruction programs. While it is hopeful that this evidence will inform future policy and reform efforts, further research within the field of Direct Instruction is necessary, not only to investigate effects on additional populations, under different conditions, and within different content areas, but also to continue to reinforce the unequivocally solid research base.

REFERENCES

Adams G., & Engelmann, S. (1996). *Research on Direct Instruction: 25 Years beyond DISTAR.* Seattle, WA: Educational Achievement Systems.

American Federation of Teachers. (1998). *Building on the best, learning from what works: Seven promising reading and language arts programs*. Washington, D.C.: American Federation of Teachers.

American Institutes for Research. (1999). *An educator's guide to schoolwide reform.* Arlington, VA: American Association of School Administrators.

American Psychological Association's Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-283.

Ashworth, D.R. (1999). Effects of Direct Instruction and basal reading instruction programs on the reading achievement of second graders. *Reading Improvement*, *26*(4), 150-156.

Bereiter, C., & Kurland, M. (1981-1982). A constructive look at Follow Through results. *Interchange, 12,* 1-22.

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*(2), 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), pp. 141-166

Carnine, D., Silbert, J., Kame'enui, E., & Tarver, S. (2004). Direct instruction reading (4th ed.). Upper Saddle River, NJ: Pearson

Cochrane Collaboration. (2008). *The Cochrane Library: Independent high-quality evidence for health care decision making.* Retrieved December 31, 2010 from http://www.thecochranelibrary.com

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Cole, K. N., & Dale, P. S. (1986). Direct language instruction and interactive language instruction with language delayed preschool children: A comparison study. *Journal of Speech and Hearing Research, 29,* 206–217.

Cole, K. N., Dale, P. S., & Mills, P. E. (1991). Individual differences in language delayed children's responses to direct and interactive preschool instruction. *Topics in Early Childhood Special Education,* 11, 99–124.

Cole, K. N., Dale, P. S., Mills, P. E., & Jenkins, J. R. (1993). Interaction between early intervention curricula and student characteristics. *Exceptional Children, 60,* 17–28.

Cooper H. (2010). *Research synthesis and meta-analysis.* Thousand Oaks, CA: SAGE Publications, Inc.

Cowan, P. A. (2004). Advancing evidence-based practice through meta-analysis. *Nephrology Nursing Journal, 31(3)*, 343-347.

Cross, R. W., Rebarber, T., & Wilson, S. F. (2002). Student gains in a privately managed network of charter schools using Direct Instruction. *Journal of Direct Instruction, 2*(1), 3-21.

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

Darch, C., & Simpson, R. G. (1991). Effectiveness of visual imagery versus rule-based strategies in teaching spelling to learning disabled students. *Research in Rural Education, 7*(1), 61-70.

Engelmann, S. (2007). *Teaching needy kids in our backwards system.* Eugene, OR: ADI Press.

Engelmann, S. (October 14, 2008). Machinations of What Works Clearinghouse. *Education News.* Retrieved December 12, 2010 from http://ednews.org/articles/machinations-of-what-works-clearinghouse.html

Forness, S. R., Kavale, K. A., Blum, I. M., & Lloyd, J. W. (1997). Mega-analysis of meta-analysis: What works in special education. *Teaching Exceptional Children, 19*(6), 4-9.

Gersten, R. M., Becker, W. C., Heiry, T. J., & White, W. A. T. (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*(2), 109–121.

Ginn, P. V., Keel, M. C., & Fredrick, L. D. (2002). Using Reasoning and Writing with gifted fifth grade students. *Journal of Direct Instruction, 2*(1), 41-47.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context (NCES 2009–001 Revised)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Grossen, B. (2004). Success of a Direct Instruction model at a secondary level school with high-risk students. *Reading & Writing Quarterly*, *20*, 161-178.

Hedges, L. & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando: Academic Press.

Jackson, G. B. (1980). Methods for integrative reviews. *Review of Educational Research, 50,* 438-460.

Kavale, K. A. (2001). Decision making in special education: The function of meta-analysis. *Exceptionality, 9*(4), 245-268.

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

MacIver, 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.

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.

National Center for Education Statistics (2001). *The nation's report card: Mathematics highlights 2000.* Washington, DC: U.S. Department of Education/Office of Educational Research and Improvement.

National Council of Teachers of Mathematics (2000). *Executive summary: Principles and standards for school mathematics.* Retrieved December 15, 2010 from http://www.nctm.org/uploadedFiles/Math_Standards/12752_exec_pssm.pdf\

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel,* Washington, DC: U.S. Department of Education. Retrieved December 15, 2010 from http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

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.

Przychodzin, A. M., Marchand-Martella, N. E., Martella, R. C., & Azim, D. (2004). Direct Instruction mathematics programs: An overview and research summary. *Journal of Direct Instruction, 4*(1), 53-84.

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

Robinson, J. W. & Hesse, K. D. (1981). A morphemically based spelling program's effect on spelling skills and spelling performance of seventh grade students. *Journal of Educational Research, 75*(1), 56-62

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

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.

Slavin, R. E. (1986). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. *Education Researcher, 15,* 5-11.

Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48*(1), 9-18.

Sexton, C. W. (1989). Effectiveness of the DISTAR Reading I program in developing first graders' language skills. *Journal of Educational Research, 82*(5), 289-293.

Shadish, W., Cook, T. & Campbell, D (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin

Simonsen, F., & Gunter, L. (2001). Best practices in spelling instruction: A research summary. *Journal of Direct Instruction, 1*(2), 97-105.

Stockard, J. (2010). An analysis of the fidelity implementation policies of the What Works Clearinghouse. *Current Issues in Education, 13*(4).

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

Tarver, S. G., & Jung, J. S. (1995). A comparison of mathematics achievement and mathematics attitudes of first and second graders instructed with either a discovery-learning mathematics curriculum or a Direct Instruction curriculum. *Effective School Practices, 14*(1), 49-57.

U.S. Department of Education (2002). *No Child Left Behind: A Desktop Reference.* Washington DC: Office of Elementary and Secondary Education.

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 school reform. *Journal of Direct Instruction, 8*(1), pp. 1-18

Watkins, K.L, & Slocum, T.A. (2003). The components of direct instruction*. Journal of Direct Instruction, 3*, 75-110.

What Works Clearinghouse (2008). *Beginning reading. What Works Clearinghouse Topic Report.* Retrieved December 15, 2010 from http://ies.ed.gov/ncee/wwc/pdf/WWC_ReadingMastery_081208.pdf

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.

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.

APPENDICES

Table 1. Research syntheses reviewed

Adams & Engelmann (1996)	Meta-analysis: integrated results of 34 studies involving special education and general education students
Borman, Hewes, Overman, & Brown (2003)	Meta-analysis: integrated results of 232 studies investigating the effectiveness of 29 comprehensive school reform models
Kinder, Kubina, & Marchand-Martella (2005)	Systematic review: summarized findings of 37 studies involving students with high incidence disabilities and 8 studies including students with low incidence disabilities
Przychodzin, Marchand-Martella, & Azim (2004)	Systematic review: reviewed findings of 12 studies examining effects of DI math programs DISTAR Arithmetic, Corrective Mathematics, and Connecting Math Concepts
Przychodzin-Havis, Marchand-Martella, Martella, Miller, Warner et al. (2005)	Systematic review: summarized findings of 28 studies investigating effects of *Corrective Reading*
Schieffer, Marchand-Martella, Martella, Simonsen, & Waldron-Soler (2002)	Systematic review: reviewed findings of 25 studies examining effects of *Reading Mastery* with special education and general education students
Simonsen & Gunter (2001)	Systematic review: summarized findings of studies comparing DI spelling programs, including *Spelling Mastery* and *Spelling Through Morphographs* to comparison approaches to teach spelling
What Works Clearinghouse Beginning Reading Report, *Reading Mastery* (2008)	Best evidence synthesis: sought to review research on the effectiveness of *Reading Mastery;* however, none of 61 studies screened met criteria for inclusion
White (1988)	Meta-analysis: synthesized results of 25 studies involving special education students only

Table 2. Average effect sizes calculated by subject area and population, based on White (1988) and Adams & Engelmann (1996)

	White (1988)	**Adams & Engelmann (1996)**
Overall ES for Special Education Students	.84 (25)	.90 (13)
Overall ES for General Education Students	N/A	.82 (13)
Reading	.85 (13)[1]	.69 (43)[2]
Math	.50 (4)[1]	1.11 (33)[2]
Language	1.21 (3)[1]	.49 (39)[2]
Spelling	N/A	1.33 (27)[2]

Note: Numbers in parentheses refer to the number of studies from which the average ES estimate was calculated; an exception to this is for Adams & Engelmann (1996), as the numbers in parentheses for subject areas only refer to the number of individual comparisons from which the average ES was calculated.

[1]ES by subject area in White (1988) are calculated based on studies involving special education students only

[2]ES by subject area in Adams & Engelmann (1996) are calculated based on studies involving *both* special education and general education students