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.