

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 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 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 Reading¹ (*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 $\frac{1}{4}$ 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

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 approach 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 processing 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*. Washington, 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)
- Traweck, 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/~adiiep/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*. Washington, 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.