

***Examining the Inaccuracies and Mystifying Policies and Standards of the What Works Clearinghouse:  
Findings from a Freedom of Information Act Request***

Technical Report 2014-5



Timothy W. Wood

Office of Research and Evaluation, NIFDI

October 14, 2014

## Table of Contents

	Page
Executive Summary	iii
Text Body	1
Reasons for Quality Reviews and Changes Made	2
WWC Reviews and the Issue of Implementation Fidelity	4
Version 1.0 (May 2008)	5
Reviews of Reports on Read Naturally	
Review of Wilson Reading System	
Review of Accelerated Reader Program	
Version 2.0 (December 2008) and 2.1 (September 2011)	12
Review of the LETRS Program	
Review of Read Naturally	
Version 3.0 (February 2013, Final Approval March 2014)	17
Additional Areas of Concern	18
A Note on the WWC Response to the FOIA Request	20
Summary and Conclusion	22
Appendix A: FOIA Request	25
Appendix B: Organizations, Study Authors, Program Developers, and Education Researchers Who Requested Quality Reviews	26
Appendix C: Programs and Studies That Were the Focus of Quality Reviews	28
Appendix D: Reactions to the WWC Reports and Quality Review Process	30
Appendix E: Quality Reviews and WWC Conclusions	34
References	40

## Executive Summary<sup>1</sup>

The What Works Clearinghouse (WWC) was established in 2002 by the Institute for Education Sciences (IES) at the U.S. Department of Education. Their mission is “to be a central and trusted source of scientific evidence for what works in education” (WWC Procedures and Standards Handbook Version 3.0, pp. 1). The WWC conducts analyses on available research, covering multiple subjects and populations, to determine the effectiveness of specific instructional programs.

In the fall of 2013, in response to a Freedom of Information Act (FOIA) request, the National Institute for Direct Instruction (NIFDI) Office of Research received 3,721 pages of emails regarding What Works Clearinghouse’s Quality Reviews that had been conducted up to that point in time. From 2007 to the fall of 2013, 62 Quality Reviews were performed in response to the concerns of 54 organizations, study authors, program developers, teachers, and education researchers. Forty-one different instructional programs and study reviews were examined in these Quality Reviews. This report examines the Quality Reviews conducted by the WWC, the different versions of the *WWC Procedures and Standards Handbook*, how the WWC inconsistently applies their standards and policies, and the range of problems documented in the Quality Reviews.

The first section of this report summarizes why these Quality Reviews were requested and the changes the WWC made as a result of the reviews. This information shows the wide-ranging nature of the concerns forwarded to the WWC. The second section focuses on reviews involving concerns with one area: the way in which the WWC treated issues of low fidelity of implementation. This focus was chosen because high fidelity is a crucial element of guaranteeing strong internal validity of a study and accurate reports of the impact of an intervention. The third section provides a synopsis of other areas of concern expressed by those requesting a review, showing that issues were raised in a variety of areas. In review of the documents provided it became evident that a significant amount of information and documentation was withheld. This missing information and the insight it would have provided are discussed more extensively in the fourth section of the paper. The final section includes a brief summary and discussion of concerns raised within the analysis.

Quality Reviews were conducted in response to a variety of concerns. The major concerns documented in these reports included the misinterpretation of study findings, inclusion of studies where programs were not fully implemented, exclusion of relevant studies from review, inappropriate inclusion of studies, concerns over WWC policies and procedures, incorrect information about a program developer and/or publisher, and the classification of

---

<sup>1</sup> The author gratefully acknowledges the assistance and guidance of Jean Stockard in the analysis of the discussed documents. All conclusions and opinions in this document are, however, the sole responsibility of the author.

programs. As a result of these Quality Reviews, the WWC concluded revisions were necessary 37 times. The WWC conclusions about eight of the Quality Reviews were not included in the documents sent to the NIFDI Office of Research. By removing these eight Quality Reviews when calculating the percentage of reports that were changed, 37 of 54 (69%) of the WWC publications were revised in response to the concerns of inquirers.

The changes made were conducted to provide greater clarity on the review process and the selection of studies, provide greater clarity on the classification of studies reviewed, change the classification ratings of a study or studies, reclassify a program, provide greater clarity on the description of a study, provide greater clarity on the level of implementation in a study, remove a study or studies from consideration in an intervention report, add a study or studies in the review of a program, and change a summary judgment of a program.

Despite the changes made in response to these Quality Reviews, multiple errors remained. Among the most serious errors was the exclusion of relevant research, the inclusion of inappropriate studies, and the misinterpretation of study findings. The WWC inclusion of inappropriate studies became evident in the concerns of numerous Quality Reviews. Inquirers were greatly concerned about the inclusion of studies that misrepresented the design of a study and the effectiveness of an intervention. Inquirers were greatly concerned about how the inclusion of these studies created an inaccurate review of a program as well as the way in which they promoted misinformation about the design of a program.

The most shocking and severe errors documented in these Quality Reviews were the WWC's repeated misinterpretation of study findings. Multiple inquirers documented how the WWC made conclusions about study findings that did not align with the authors' conclusions, and in some instances reported totally different conclusions. Over 80 percent of the requests for Quality Reviews involved concerns with misinterpretations of study findings, which appears to indicate that this example is far from unique. Misinterpretation of study findings could occur from procedural errors of individual reviewers. Yet, numerous Quality Reviews indicated that errors resulted from WWC policies, specifically, the WWC decision not to consider the fidelity of implementation when determining the effectiveness rating of an intervention.

The WWC's policy on examining the impact of implementation fidelity in relation to the effectiveness of a program is unclear and troubling. This lack of attention to fidelity is especially disturbing given the key role of fidelity to internal validity of research designs. If a research study does not implement a program as it was designed or intended to be used, the results of the study are clearly invalid. It is impossible to tell if an effect (or lack of an effect) is related to the intervention as designed or to the changes made with alternative implementations. Moreover, with poor fidelity of implementation less effective programs will be deemed to be more effective while high quality programs will be determined to be less

effective (see Fixsen, et al., 2005; Stockard, 2010 for discussions of the importance of fidelity of implementation).

Multiple individuals and organizations were concerned that the WWC misinterpreted study findings and inappropriately included studies that only partially implemented an intervention, thus having, by definition, low fidelity to the intervention model. To better understand the WWC's wavering stance on the importance of implementation fidelity, and their inconsistent application of their own policies, the WWC's Procedures and Standards Handbooks were examined in relation to the WWC publications that were the basis for the 62 Quality Reviews.

The errors documented in these Quality Reviews contrast sharply with the WWC's statement that they "provide accurate information on education research" (<http://ies.ed.gov/ncee/wwc/aboutus.aspx>). In fact, the number and variety of errors could be seen as severe and alarming. The errors range from simple mistakes of mischaracterizing an instructional program to including inappropriate studies in the review of an instructional program, resulting in an inaccurate rating of effectiveness. These Quality Reviews have clearly documented the multitude of errors in WWC reports and their negative consequences.

As alarming as the errors documented in these Quality Reviews are, it is equally worrisome that the WWC policies, procedures, and standards are not clear and consistently applied. With the information provided from the FOIA request and the publicly available information, three conclusions appear clear. 1) The WWC suffers from a lack of transparency in their policies and guidelines, 2) the conclusions they create in their reports can be misleading, and 3) the reports are potentially damaging to program developers and ultimately the success of students.

# Examining the Inaccuracies and Mystifying Policies and Standards of the What Works Clearinghouse: Findings from a Freedom of Information Act Request

---

The What Works Clearinghouse (WWC) was established in 2002 by the Institute for Education Sciences (IES) at the U.S. Department of Education. Their mission is “to be a central and trusted source of scientific evidence for what works in education” (WWC *Procedures and Standards Handbook* Version 3.0, pp. 1). The WWC conducts analyses on available research, covering multiple subjects and populations, to determine the effectiveness of specific instructional programs. Additional reports are conducted to review the findings from individual studies. Quality Reviews are conducted when individuals contact the WWC about concerns they have about the findings of WWC reports. For these reviews the WWC Quality Review Team examines the original report and determines if there are any errors and if the WWC followed their procedures and standards during the creation of the report. However, the WWC does not provide any information on their website about the extent to which such Quality Reviews have been requested or the outcomes of the processes.

In the fall of 2013, in response to a Freedom of Information Act (FOIA) request, the National Institute for Direct Instruction (NIFDI) Office of Research received 3,721 pages of emails regarding What Works Clearinghouse’s Quality Reviews that had been conducted up to that point in time. The letter from the NIFDI research office to the FOIA office, listing the requested information, is included in Appendix A. This report summarizes the findings and concerns developed from a review of these documents.

From 2007 to the fall of 2013, 62 Quality Reviews were performed in response to the concerns of 54 organizations, study authors, program developers, teachers, and education researchers. Appendix B lists all of these organizations and individuals. Forty-one different instructional programs and study reviews were examined in these Quality Reviews. Appendix C lists all of these programs and study reviews. This report will examine the Quality Reviews conducted by the WWC, the different versions of the WWC *Procedures and Standards Handbook*, how the WWC inconsistently applies their standards and policies, and the range of problems documented in the Quality Reviews. Appendix E lists all of the Quality Reviews, the programs and or studies they reviewed, and the conclusions of the WWC following the completion of the review (when provided).

The first section of this report summarizes why these Quality Reviews were requested and the changes the WWC made as a result of these reviews. This information shows the wide-ranging nature of the concerns forwarded to the WWC. The second section focuses on reviews involving concerns with one area: the way in which the WWC treated issues of low

fidelity of implementation. As explained below, this focus was chosen because high fidelity is a crucial element of guaranteeing strong internal validity of a study and accurate reports of the impact of an intervention. The third section provides a synopsis of other areas of concern expressed by those requesting a review, showing that issues were raised in a variety of areas.

In review of the documents provided it became evident that a significant amount of information and documentation was withheld. This missing information and the insight it would have provided are discussed more extensively in the fourth section of the paper. The final section includes a brief summary and discussion of concerns raised within the analysis. Additionally, in review of these documents, it became apparent that multiple individuals were significantly dissatisfied, frustrated, and even angry over the content of the WWC reports and their treatment by the employees of the WWC. Appendix D lists excerpts from the Quality Reviews that describe their experiences working with the WWC during the request for and process of the Quality Reviews.

### **Reasons for Quality Reviews and Changes Made**

The WWC publishes three types of reports: intervention reports, single study reviews, and quick reviews. The Quality Reviews obtained through the FOIA request were conducted in response to concerns over the information presented in all three types. Multiple Quality Reviews related to the same concerns and WWC publications. The major concerns documented in these reports included the misinterpretation of study findings, inclusion of studies where programs were not fully implemented, exclusion of relevant studies from review, inappropriate inclusion of studies, concerns over WWC policies and procedures, incorrect information about a program developer and/or publisher, and the classification of programs. The occurrences of these concerns are documented in Table 1, showing how often these issues were addressed and what percentage of the Quality Reviews dealt with these issues. Note that over 80 percent of the requested reviews dealt with misinterpretations of the findings of a study.

As a result of these Quality Reviews, the WWC concluded revisions were necessary 37 times. The WWC conclusions about eight of the Quality Reviews were not included in the documents sent to the NIFDI Office of Research. By removing these eight Quality Reviews when calculating the percentage of reports that were changed, 37 of 54 (69%) of the WWC publications were revised in response to the concerns of inquirers. The changes made were conducted to provide greater clarity on the review process and the selection of studies, provide greater clarity on the classification of studies reviewed, change the classification

ratings of a study or studies, reclassify a program, provide greater clarity on the description of a study, provide greater clarity on the level of implementation in a study, remove a study or studies from consideration in an intervention report, add a study or studies in the review of a program, and change a summary judgment of a program<sup>2</sup>. A summary of the changes made for each review is in Appendix D.

Table 1

Reasons for Requesting a Quality Review of WWC Findings, 2007 to Mid-2013

<u>Reasons for requesting a Quality Review</u>	<u>Number of reports addressing this issue</u>	<u>Percentage of Reports</u>
Misinterpretation of study findings	53	85%
Inclusion of studies where programs were not fully implemented	15	24%
Exclusion of relevant studies from review	23	37%
Inappropriate inclusion of studies	24	39%
Concerns over WWC policies and procedures	23	37%
Incorrect information about a program and/or developer	6	10%
Classification of a program	2	3%

Note: Percentages are calculated from a base of 62 Quality Reviews.

Despite the changes made in response to these Quality Reviews, multiple errors remained. Among the most serious errors was the exclusion of relevant research, the inclusion of inappropriate studies, and the misinterpretation of study findings. Multiple Quality Reviews documented how the WWC did not include relevant research on specific programs that would have aided the WWC in determining a more accurate summary judgment. Sometimes the WWC would justify their actions based on their established criteria for accepting studies for review.<sup>3</sup> However, the WWC also explicitly ignored publications supplied to them that as part of the Quality Review Process (See QR2011-001, QR2012-012, and QR2012-014).

<sup>2</sup> A summary judgment was only changed for the intervention report on *Reading Mastery* for students with learning disabilities.

<sup>3</sup> Two companion NIFDI Technical Reports (2014-3 and 2014-4) document numerous logical errors in these criteria, the extent to which they contradict standard methodological practices, and, using statistical analyses, how they provide no value added to estimates of the effectiveness of interventions.



Even though they were provided with the relevant literature, the WWC did not acknowledge these studies at all, violating their own stated procedures for conducting a comprehensive literature search.<sup>4</sup> There is no indication that the WWC expanded the literature base for a given report as a result of a Quality Review. By not including these publications in the revised intervention report, the Quality Review failed to achieve its goal of correcting clear errors.

The WWC inclusion of inappropriate studies became evident in the concerns of numerous Quality Reviews. Inquirers were greatly concerned about the inclusion of studies that misrepresented the design of a program and the effectiveness of an intervention. In some instances, studies did not appropriately reflect the program, using only elements of the program or not implementing it as designed. Inquirers were greatly concerned about how the inclusion of these studies created an inaccurate review of a program as well as the way in which they promoted misinformation about the design of a program.

The most shocking and severe errors documented in these Quality Reviews was the WWC's repeated misinterpretation of study findings. Multiple inquirers documented how the WWC made conclusions about study findings that did not align with the authors' conclusions, and in some instances reporting totally different conclusions. For greater detail on the dangers of these misinterpretations see Stockard and Wood (2013), which documented how the WWC ignored evidence in a study that demonstrated strong positive effects of an intervention, and instead determined it had no discernable effects. The fact that over 80 percent of the requests for Quality Reviews involved concerns with misinterpretations of study findings indicates that this example is far from unique.

Misinterpretation of study findings could occur from procedural errors of individual reviewers. Yet, numerous Quality Reviews indicated that errors resulted from WWC policies, specifically, the WWC decision not to consider the fidelity of implementation when determining the effectiveness rating of an intervention. The next section examines the reviews related to this issue in more detail.

### **WWC Reviews and the Issue of Implementation Fidelity**

The WWC's policy on examining the impact of implementation fidelity in relation to the effectiveness of a program is unclear and troubling. This lack of attention to fidelity is

---

<sup>4</sup> The WWC should have noted these recommended research studies in their initial reports, whether or not the studies would meet their standards. Even if the studies would not meet WWC standards they should have been included in the report's reference list, noting that they were identified and examined.

especially disturbing given the key role of fidelity to internal validity of research designs. If a research study does not implement a program as it was designed or intended to be used, the results of the study are clearly invalid. It is impossible to tell if an effect (or lack of an effect) is related to the intervention as designed or to the changes made with alternative implementations. Moreover, with poor fidelity of implementation less effective programs will be deemed to be more effective while high quality programs will be determined to be less effective (see Fixsen, et al., 2005; Stockard, 2010 for discussions of the importance of fidelity of implementation).<sup>5</sup>

Multiple individuals and organizations were concerned that the WWC misinterpreted study findings and inappropriately included studies that only partially implemented an intervention, thus having, by definition, low fidelity to the intervention model. To better understand the WWC's wavering stance on the importance of implementation fidelity, and their inconsistent application of their own policies, the WWC's Procedures and Standards Handbooks were examined in relation to the WWC publications that were the basis for the 62 Quality Reviews.

The WWC has published multiple versions of their *Procedures and Standards Handbook*. The first publicly available version was released in May of 2008 and the most recent edition was finalized in March 2014. Each version has addressed the importance of implementation fidelity to varying degrees. In examining the documents provided through the FOIA request, these handbooks were reviewed to verify that the WWC was following policies in effect at the time of the review, and whether individual statements about WWC policy were supported by the handbook. The following discussion examines each version of the handbook and Quality Reviews raising questions regarding fidelity and conducted while these handbooks were in use. Each handbook is examined in relation to the policies on the fidelity of implementation and then specific Quality Reviews are reviewed to examine how the WWC followed, or did not follow, its policies.

### Version 1.0 (May 2008)

In version 1.0 (May 2008) of the handbook, the WWC considers the measurement of implementation fidelity in the assessment of the strength of the evidence that a study

---

<sup>5</sup> As would be expected, studies have found that the gap between students in DI programs, which have been found to be highly effective, and those in traditional programs is greater for students of teachers who implemented DI with higher fidelity (Gersten, Carnine, Zoref, & Cronin, 1986; Stockard, 2010). Similarly, studies that have focused only on students receiving Direct Instruction have found the highest achievers in classrooms or schools that have higher levels of fidelity of implementation (Benner, Nelson, Stage, & Ralston, 2010; Carlson & Francis, 2002; Gersten, Carnine, & Williams, 1982; Ross et al., 2004).

provides for the intervention's effectiveness. In this version the WWC classified "well-designed and implemented randomized controlled trials" as meeting evidence standards. Studies that were "quasi-experiments with equating and no severe design or implementation problems" (pp. 5) were classified as meeting evidence standards "with reservations." To what level implementation fidelity was examined is not clear, but it appears that the WWC did make a decision on whether an intervention had any problems with implementation. Additionally, under the section Evidence Base and Heuristic Rules, the handbook discusses how they characterize the quality of the research design generating the effect estimates. A strong design is classified as "designs that meet the WWC's evidence standards, which are RCTs without severe design or implementation problems." A weak design is classified as "designs that meet WWC's evidence standards with reservation, which include RCTs with severe design or implementation problems, and QEDs with equating and without severe design or *implementation problems*" (pp.19, emphasis added). Once again the handbook states that the WWC makes a decision on whether an intervention has any implementation problems. What constitutes a problem or how this is determined is, however, not discussed.

In Quality Review 2007-003<sup>6</sup>, the inquirer asked "How does WWC guard against drawing conclusions about a program based on lack of fidelity to the model rather than flaws in the model itself" (pp. 5).<sup>7</sup> The WWC responded

"In education research in general - and in WWC reviewed studies in particular - absolute fidelity to the program model is rare. However, we do take steps on [sic] WWC to ensure that we are reviewing an appropriate manifestation of the model. As you will see in our intervention reports, we engage in extensive efforts to outline the program model and to determine whether the intervention is appropriately labeled. Differences in implementation are outlined in our reports. That said, as you can imagine, it is extremely difficult to establish criteria by which we can determine that the program examined in one study is implemented according to the model, while the program examined in another study is implemented outside of the scope of the model. Instead, our approach is to accept the authors' claim that the program studied is consistent with the model, but then in our reports, highlight any apparent differences between the program studied and the model" (2007-003, pp. 77).

---

<sup>6</sup> Documentation related to all of the Quality Reviews is available on the NIFDI website at [http://nifdi.org/resources/downloads/cat\\_view/265-what-works-clearinghouse-wwc](http://nifdi.org/resources/downloads/cat_view/265-what-works-clearinghouse-wwc)

<sup>7</sup> Page numbers correspond to documentation on these Quality Reviews, provided by the WWC and available on the NIFDI website.

The decision to rely on an author's claim that the program is consistent with the model is not clearly substantiated by the WWC handbook used at that time. Furthermore, this policy is not consistently used by the WWC in their reports and Quality Reviews. Multiple Quality Reviews were conducted solely because inquirers were concerned the WWC did not accept the authors' claim that they did **not** use an appropriate manifestation of the intervention. In other words, even though the authors explicitly said that they did not fully implement an intervention, the WWC chose to treat the study as a valid test of the program. Not only did WWC employees cite policies clearly not documented in the WWC handbook, they did not consistently apply these policies in multiple reports. Three examples of the WWC not relying on the author's claims about a study design while version 1.0 of the handbook was in effect are given below.

**Reviews of Reports on Read Naturally** – In Quality Review 2008-011, the inquirer questioned the inclusion of a study that used passages from the instructional program *Read Naturally*, but did not use the *Read Naturally* strategy. Furthermore, the inquirer argued the author did not conduct the study to evaluate *Read Naturally* and, in support of this point, included emails from the study author that described the design and purpose of the study. In an email to the inquirer, the study author wrote

“Per our phone conversation this morning, I wanted to clarify a couple of issues regarding my 2002 dissertation titled, ‘Accelerated Reading Trajectories: The Effects of Dynamic Research-Based Instruction.’ While I used *Read Naturally* materials, I did not fully implement the *Read Naturally* strategy and my study was not intended to evaluate the *Read Naturally* strategy. Rather, the purpose was to determine the impact of ongoing supplemental fluency practice on 2<sup>nd</sup> grade students’ rates of learning to read” (pp. 1-2).

In response to this email from the study author, WWC Project Director Becki Herman wrote to the inquirer

“...the WWC does not look at WHY a study author chose to test an intervention. If a study author looks at the effects of an intervention, we would consider that study eligible for review (pending review of other design issues). Further, as we have also discussed earlier, the WWC does not consider implementation except in the case of removing studies that have zero implementation (e. g., there were no elements whatsoever of the intervention used in the study). In this case, although the study author may not have implemented the intervention as you had originally intended, it is clearly not a

case of zero implementation. Although we have heard and considered your concerns, the Department, the Beginning Reading team, and the WWC core team agree that this study should remain in the review” (pp. 2).

Becki Herman’s statements are inconsistent with the policies and standards published in the WWC handbook in use at that time (1.0). In this summary of the Quality Review, Becki Herman stated that the WWC will only remove a study if it has “zero implementation,” rather than determining if there were problems related to implementation. The reference to zero implementation is not found in the handbook at all. Furthermore, the WWC did not address the study author’s claim about not determining the effectiveness of the intervention and not fully implementing the program. The study author clearly stated that she did not intend her study to be an evaluation of the *Read Naturally* program and strategy, and did not fully implement the program as it was designed. The WWC ignored this information in the initial decision to include this information in the report and reinforced this decision during the Quality Review. Becki Herman’s statements about WWC policy are distinctly different from the published policies related to this issue as well as previous comments by the WWC in Quality Review 2007-003, described above.

As a result of this Quality Review (2008-011) the WWC did communicate with the study author about her study to gather additional information about its goal and design. In an email to the WWC, the study author wrote

“As we discussed, I used much of the *Read Naturally* material and process for my dissertation in 2002/3. However, the exact process was not followed. The focus of the study was to look only at fluency (not vocabulary or comprehension). Therefore, the vocabulary (pre-reading) and comprehension (question answering and retell) parts of the *Read Naturally* process were not emphasized. While I did have students answer questions before moving on to the next story/level, it was not an identified implementation procedure of the study itself. I can see where the *Read Naturally* folks are concerned with my study’s inclusion in the WWC as the program protocol was not followed in its entirety” (pp. 8).

The study author clearly documented how she only implemented portions of the *Read Naturally* program and strategy, and the study was not a primary analysis of the intervention. Concluding their report of this Quality Review, WWC Director Mark Dynarski wrote

"We have concluded that, while some components of the intervention were not implemented, the Hancock study should remain in the review of *Read Naturally*. Because there is no standard metric for treatment fidelity, WWC intervention reports include studies with variation in fidelity but note aspects of implementation that may affect the interpretation of findings. We will update the WWC report on *Read Naturally* to clarify which aspects of *Read Naturally* were not implemented in the Hancock study" (pp. 50).

Once again the WWC conclusions were not consistent with the version of the WWC handbook that was effective at the time. Mark Dynarski acknowledged problems with implementation, but determined this does not affect the effectiveness rating of the *Read Naturally* program. Dynarski notes that the level of implementation may affect the interpretation of the findings, but does not affect WWC conclusion on the effectiveness rating. Apparently the WWC believes individuals should determine the importance of implementation fidelity in examining the effectiveness of an intervention, but not the WWC.<sup>8</sup> At the time this Quality Review was conducted, WWC policy on the importance of fidelity of implementation to ratings was unclear and appeared to vary between the handbook and judgments of individuals writing and reviewing these reports. Using the criteria in the handbook, one could reasonably suggest that this study did not provide an appropriate test of the *Read Naturally* model and should not have passed evidence standards due to problems with implementation.<sup>9</sup>

**Review of Wilson Reading System** –The developers of *Read Naturally* were not the only inquirers who were concerned about how the WWC treated the issue of fidelity of implementation in their reports and how they presented this information. In Quality Review 2009-002, the inquirer was concerned about how the WWC listed their report findings. On July 5<sup>th</sup> 2007, the inquirer wrote

"I've just finished reading the report published by the WWC on the Wilson Reading System. I thought you did a fine job in describing the program, and I

---

<sup>8</sup> The clarification mentioned by Dynarski appears to have resulted in one sentence added to an appendix to the report that states that the "study excluded *Read Naturally*'s pre-reading vocabulary instruction component and the *Read Naturally* placement system to individualize instruction" (p. 17 of the 2013 updated report). However, there is no mention that this contradicts the design of the program or any reflection of the author's contention that this study did not test the efficacy of the program.

<sup>9</sup> Dynarski's statement that there is no standard metric for measuring implementation fidelity is inaccurate. Measures of implementation fidelity are, by definition, program specific. Moreover, having a simple dichotomous variable of problems noted or not would be far better than simply ignoring the issue. See an extended discussion in Stockard (2008)

also noted that you clearly spelled out that the study we did of Wilson and several other intervention programs used a modified, or limited version of the program, in order to test a specific hypotheses [sic] about the impact of word level instruction vs. instruction that provided both word-level and vocabulary/comprehension instruction. The Wilson program was asked to participate in the word-level only part of the study, and they carefully complied with the study protocol, and thus did not include the vocabulary and comprehension aspects of their program. You have said this in your report, but I'm concerned that the report will still be misleading to educational practitioners for two reasons. First, in the only highlighted text in the document that describes the findings, you used these words; **The WWC found Wilson Reading Systems to have potentially positive effects on alphabetics and no discernable effects on fluency and comprehension.** It seems to me, as the PI on this study, it would have been more fair to Wilson, and also more correct, to say something like: **'The WWC found the modified version of the Wilson Reading System used in this study to have...'** Also, in the summary section (which is the take away message a lot of folks will leave with, you need a final qualifier after the sentence ending '... no discernable effects in fluency or comprehension' that would read something like, 'It should be clearly noted, however, that the comprehension and vocabulary components of the Wilson Reading System were not used at the request of the researchers conducting the study.' I want to communicate as clearly as I can both that the Wilson people complied very precisely with the study protocol, and they also expressed some reservations at the time about focusing only on the word level components of their program. At this point in time, I regret implementing 'named programs' in only limited versions because of the fact that it does not produce a fair evaluation of their program as it is marketed and implemented in schools. However, we had to go with the 'two intervention classes' in our design in order to meet power requirements. I'm hoping that you can be even clearer than you presently are in your report about the limited version of the Wilson System used in this study in the interests of both fairness to the Wilson people and in the interests of good science" (pp. 4).

WWC Deputy Director Scott Cody determined the suggested rewording of the WWC would be against WWC policy. On March 13<sup>th</sup> 2009, he wrote



“WWC topic reports (and by extension, the WWC website ‘Create Your Own Summary’ search tool) are intended to present summary statistics from individual intervention reports. Similarly, the ‘Rating of Effectiveness’ tables in individual intervention reports are intended to summarize across individual studies. It is WWC policy not to annotate these summary tables with information about individual studies” (pp. 6).

The unfortunate consequence of this decision to reject the author’s explanation of the study and to maintain this phrasing is that the WWC conclusions can be used out of context, are clearly misleading, and have the potential to be damaging to the programs and their success in educating children. Multiple other Quality Reviews have documented how the WWC findings are misleading and how the findings are displayed can be damaging.<sup>10</sup> Furthermore, the policy Scott Cody is referring to is not documented in version 1.0 of the handbook, which was the current handbook when the report was published in 2007. Reference to this policy is also not documented in version 2.0 (released in December 2008), which was the current handbook when Scott Cody responded to the Quality Review request over 18 months after the initial publication of the report.

**Review of Accelerated Reader Program** – In Quality Review 2008-012, the WWC’s policy on the fidelity of implementation is called into question. The inquirer was concerned about the inclusion of a study that partially implemented a program and its strategy in a WWC report. The inquirer wrote

“Unfortunately, the Bullock study was not a valid study of the effectiveness of Accelerated Reader because, according to the author, students in the experimental group did not use the program. They may have used the software sparingly but clearly did not implement AR Best Classroom Practices. Specifically, the author cautions that the extremely short 10-week duration and non-implementation are major limitations, reporting that, ‘the duration of the study limits the exposure students had to AR. The number of AR comprehension quizzes taken by the students range from 0 to 14, with an **average of 3.69 per student** in the experimental group. This number of quizzes, combined with the number of participants, may not provide enough

---

<sup>10</sup> Additional information about how WWC reports were misleading and potentially damaging is documented in the following Quality Reviews: QR2008-001, QR2008-011, QR2009-001, QR2009-008, QR2010-003, and QR2010-017. This does not encompass all of the Quality Reviews that documented how the WWC reports were misleading, but rather a selection of different issues.



information for a significant difference to emerge,' p. 37, emphasis added” (pp. 3).

In the Quality Review, the WWC concluded that “...while Bullock did not implement Accelerated Reader using the Accelerated Reader best practices, the study is still eligible for WWC review as a study of Accelerated Reader” (pp. 16). The WWC justified this conclusion, stating

“The WWC is designed to produce a systematic review of literature on the effectiveness of education interventions. To be both informative to educators and comprehensive, the review includes evaluations of interventions in ‘real world’ settings. This means that studies that do not follow all prescribed procedures for an intervention may still be eligible for review if the study's implementation reflects how educators might implement the intervention in actual practice” (pp. 16).

In other words, the WWC's rationale for retaining the original wording was that the study provided an example of how a program could possibly be used in a classroom. Clearly, however, this is not an appropriate approach to determine how effective an instructional program is, for one cannot tell if any effect occurred from the program itself or the modifications. Moreover, the decision does not reflect the WWC policies in their handbook at the time the report was finished. The problem with the WWC rationale can be seen with an analogy to cooking. When preparing a recipe it would be irrational to evaluate the recipe solely based on the list of ingredients and not the process of how those ingredients should be used to create the desired outcome. There are many possible outcomes that could result from combining the ingredients and it is not appropriate to evaluate the original recipe using all of the possible alternative approaches. Many program developers have specifically designed the instructional procedures for their programs to maximize the efficiency and effectiveness of teaching and to assert that an evaluation of parts of these programs also evaluates the whole is logically invalid. (Engelmann 2004)

### **Versions 2.0 (December 2008) and 2.1 (September 2011)**

In December of 2008 the WWC updated their handbook and altered their policies on evaluating implementation fidelity. Under the section on eligibility screening, this handbook (Version 2.0) lists requirements for studies to be eligible for review. One requirement is that the study is a primary analysis of the effect of an intervention (pp. 10). Like the previous version, this handbook still examines the fidelity of implementation when examining the

effectiveness of an intervention. However, under the section Corrections and Adjustments, the handbook states “The WWC makes no adjustments or corrections for variations in implementation of the intervention; however, if a study meets standards and is included in an intervention report, descriptions of implementation are provided in the report appendices to provide context for the findings” (pp. 17). Again, the WWC stance on evaluating implementation fidelity is not clear. There is no mention of the WWC only removing a study from consideration if it has zero implementation, even though the Herman letter noted above stated that this was WWC policy. It would also seem that the judgment of whether a study is a primary analysis of the effect of an intervention is up to the discretion of the reviewers. How the WWC determines that a study is a “primary analysis” is not clearly documented or defined in this handbook. Several Quality Reviews illustrate issues related to this issue and the inconsistent application of criteria in Version 2.0 of the WWC handbook.

***Review of the LETRS Program*** – In Quality Review 2008-013, the inquirer was concerned about the inclusion of a study that they believed did not properly implement the Language Essentials for Teachers of Reading and Spelling (LETRS) program. On December 27<sup>th</sup> 2008, the inquirer wrote

“I am writing to protest the content and nature of the research summary, and to request that it be rewritten to more truthfully reflect what was done by AIR and what was found in the study. What you have is inaccurate, misleading, and undeservedly harmful to the reputation of LETRS” (pp. 1).

The inquirer further explained that

“...we gave the AIR permission to base their training on the LETRS materials, but that we had no control over what was added by their design team, what was imposed in the way of an agenda and time frame, and what was deleted from the course of study that the LETRS materials support. The professional development program was that which was designed by AIR, using our materials but modifying them to suit the emphases that they considered important.”

Based on the previous statement of the WWC in Quality Review 2008-012, described above, it would seem that this situation could be viewed as an example of how the LETRS program could be used in a real world setting, and thus, would be included in the WWC review.

However, on February 18<sup>th</sup> 2009, the WWC concluded

"We have reviewed your concerns and concluded that it is appropriate to modify our description of this report. Rather than describe this study as a study of the Language Essential for Teachers of Reading and Spelling (LETRS) curriculum, we will revise the report to indicate that the study examined the impact of 'a professional development program based on LETRS.' Additionally, we will now refer to the quick review as 'WWC Quick Review of the Report 'The Impact of Two Professional Development Intervention on Early Reading Instruction and Achievement; rather than 'WWC Quick Review: Language Essential for Teachers of Reading and Spelling (LETRS) Study.' We plan to post a revised quick review reflecting these changes in the coming weeks" (pp. 3)

While the WWC appears to have made the appropriate determination by revising their report to more accurately represent the LETRS program and its effectiveness, their conclusion is inconsistent with previous statements related to implementation fidelity and summarized above. Because the WWC did not include their notes about their decisions related to this Quality Review in response to the FOIA request (see discussion of missing information in the next section), it is unclear why this study was revised while others were not. It appears as if decisions regarding such issues are made on a case-by-case basis and do not always clearly reflect the policies given in the *Procedures and Standards Handbook*.

**Review of Read Naturally** – In Quality Review 2010-013 the inquirer was concerned over the inclusion of a study that he argued was not designed to evaluate *Read Naturally*. Furthermore, the study used *Read Naturally* passages, but not the *Read Naturally* strategy (pp. 73). The inquirer claims the author stated the purpose of her study was not to evaluate *Read Naturally*, and provides an excerpt from the study as evidence.

In response to these concerns, Jill Constantine, Director of the WWC, wrote on February 26<sup>th</sup> 2013,

"The WWC followed protocol in choosing to review these four studies... The WWC does not screen based on whether author(s) explicitly intended the study as an evaluation of an intervention, but rather whether the study presents a primary analysis of the effect of an intervention" (pp. 156).

This conclusion reflects the policies described in the current handbook at the time, but what classifies as a primary analysis is still not clear. In response to concerns about the program not being implemented as designed, Constantine responded

"... the WWC includes studies with variation in fidelity and does not evaluate implementation fidelity. The quality review team verified that variations in implementation that are noted in the four studies and that may affect the interpretation of findings were properly included in the WWC publications" (pp. 157).

Again it appears that the WWC followed their stated policies, but the issue remains of what constitutes a primary analysis, for this term remains undefined. If a program were only partially implemented, does the WWC consider this a primary analysis? That seems to be the decision in this case, but not in Quality Review 2008-013 (the LETRS program described immediately above). Once again the WWC's policies appear unclear and their application inconsistent.

The WWC handbook was updated again in September of 2011. While the WWC maintained its policy on not making adjustments or corrections for variations in the implementation of an intervention, they did provide additional information on their treatment of implementation fidelity in determining the effectiveness of an intervention. The revised handbook (Version 2.1) lists five basic steps in the WWC review process, including "Screen studies for relevance and the adequacy of study design, implementation, and reporting" (pp. 6). Once again the WWC notes that implementation is reviewed in the screening process, but the extent of this review is not described. Additionally, under the section Single-Case Design Standards, the handbook states "The PI within each topic area will: (1) define the independent and outcome variables under the investigation, (2) establish parameters for considering fidelity of intervention implementation, and (3) consider the reasonable application of the Standards to the topic area and specify any deviations from the Standards in that area protocol." A footnote is included in reference to intervention implementation that states "Because interventions are applied over time, continuous measurement of implementation is a relevant consideration" (pp. 76). Thus, the WWC handbook appears to evaluate the fidelity of implementation in their review of studies, but the level and nature of this evaluation is not clearly described. Interestingly, the vagueness of the handbook does not reflect the level of detail described by WWC employees when discussing these policies in correspondence with inquirers of Quality Reviews.

An example of this issue is found in Quality Review 2010-013, described above. The inquirer directly addressed the WWC policy that "The WWC does not screen based on whether the author explicitly intended the study as an evaluation of an intervention or whether the developer indicates implementation was acceptable, but rather whether the study presents a primary analysis of the effect of an intervention," writing

"a Read Naturally intervention is defined as Read Naturally materials used in conjunction with the Read Naturally strategy. The materials alone are not an intervention... By using the materials and not the strategy, Hancock's study did not analyze Read Naturally as an intervention. If WWC's policy is truly that 'the study must present a primary analysis of the effect of an intervention, the Hancock study does not qualify... WWC has noted on its website that the Hancock study excluded Read Naturally's pre-reading instruction and placement system, but these notations are not sufficient to clarify that the Read Naturally strategy was not used" (pp. 116).

The inquirer provided additional details on how the Hancock study did not test the effectiveness of the *Read Naturally* program. The inquirer wrote

"In an email to Becki Herman of WWC, Carrie Hancock stated the following: 'While I used Read Naturally materials, I did **NOT** fully implement the Read Naturally strategy and my study was **NOT** intended to evaluate the Read Naturally strategy. Rather, the purpose was to determine the impact of ongoing supplemental fluency practice on second grade students' rates of learning to read.' By using the materials and not the strategy, Hancock's study did not analyze Read Naturally as an intervention. If WWC's policy is truly that 'the study must present a primary analysis of the effect of an intervention,' the Hancock study does not qualify" (pp. 116).

As described above, the WWC dismissed this lengthy explanation of problems with the analysis. Again the WWC policy on determining what is a primary analysis is unclear. The correspondence clearly indicates how yet another inquirer is uncertain about the nature of WWC policy and their application of this policy in conducting their reports. The WWC provided no clarification on this issue other than the Quality Review confirming the studies were accurately described.

### **Version 3.0 (February 2013, Final Approval March 2014)**

The Handbook was updated again in February of 2013 (Version 3.0), providing additional information about the treatment of implementation fidelity. New additions to the handbook included greater clarity on the evaluation of the importance of the fidelity of implementation in determining the effectiveness of an intervention. On page 19, the handbook states

"While the WWC documents how the intervention was implemented and the context in which it was implemented for the study sample, it makes no

statistical adjustments or corrections for variations in implementation of the intervention (e.g., relative to an ideal or average implementation). Variations in implementation are to be expected in studies of interventions, since they take place in real-life settings, such as classrooms and schools, and not necessarily under tightly controlled conditions monitored by researchers.”

The handbook also discusses how intervention and single study reports will describe the level of implementation for single-case designs in an appendix. In this version of the handbook, under the category of “Additional Consideration: Areas for Discretion” within the section of pilot standards for the review of single-case designs, the handbook states

“The topic area team leadership will (a) define the independent and outcome variables under the investigation, (b) establish parameters for considering fidelity of intervention implementation and (c) consider the reasonable application of the standards to the area and specify any deviations from the standards in that area protocol” (pp. 61).

The handbook also lists a footnote corresponding to considering fidelity of intervention in single-case designs that states “Because interventions are applied over time, continuous measurement of implementation is a relevant consideration” (pp. 61). There is no such extended discussion related to the group designs that are the major focus of the WWC reports.

In general, throughout the Quality Reviews and handbook guidelines there appear to be inconsistencies on how fidelity of implementation should be measured and interpreted and how variations in implementation fidelity affect the determination of effectiveness. It is unclear if the fidelity of implementation is considered only in the event of zero implementation or if consideration is at the discretion of the topic area team leadership. Furthermore the footnote appears to note the importance of implementation fidelity in determining the effectiveness of an intervention, but this contradicts the statements of WWC employees and some WWC written policies and could, in fact, be interpreted as applying to only the pilot standards for single-case designs.

Additionally, in the handbook under the category of Initial Screening for Eligibility, the WWC lists reasons why a study may be designated as ineligible for WWC review. As in the December 2008 Version (2.0) this handbook states that one of the reasons is if “The study is not a primary analysis of the effect of an intervention” (pp. 8). If the WWC followed this policy it is difficult to understand why studies that only partially implemented a program or

only tested the written materials and not the designed strategy of instruction would be included for review. However, as with earlier iterations of the handbooks and as illustrated in the outcome of Quality Reviews, the meaning of a primary analysis is not clearly defined and the examples of a “primary analysis” vary greatly. Clearly, this policy was not enforced at the time of many of these Quality Reviews. The one exception appears to be associated with Quality Review 2008-013 in which the WWC determined their report did not accurately describe the LETRS program, but rather a modified version based on the LETRS program. This appears to be the only instance of the WWC determining a program was not properly implemented. However, instead of removing the publication with the faulty information the WWC determined it was more appropriate to change the report to a review of a new program based on LETRS rather than a review of a modified version of the LETRS program

### Additional Areas of Concern

While the issue of implementation fidelity is absolutely central to internal validity of studies, the Quality Reviews indicated a number of other areas of concern with WWC policies and procedures. The emails related to these Quality Reviews provide insight into how these WWC reports and the policies that guide them are viewed by program developers, researchers, and educators. The following is a brief synopsis of these issues:

1. The WWC failed to promptly respond to the concerns and questions of individuals who had asked for Quality Reviews. There were multiple instances when the WWC failed to respond to questions within a time frame they provided. Additionally they did not always complete revisions to reports within a reasonable time frame. Delayed responses from the WWC ranged from a few weeks to over a year and a half. Unfortunately this allows for potentially inaccurate information to be posted for a long period of time.
2. The WWC sends courtesy copies of intervention reports and quick reviews to program developers and study authors. They are generally sent 1 day before publication. They are not sent to correct possible errors, but to prepare individuals for the release. In an email from Quality Review 2012-011, the WWC wrote

“... it is WWC policy to send a copy of a report to an author 24 hours before it is released on our website. The report is sent as a courtesy, so that the author is not surprised by the release. The 24-hour period is not designed as a quality assurance check - the report has already been reviewed through a WWC quality assurance process, by IES, and by an independent peer reviewer. However, if the author disagrees with the WWC's findings or characterizations,



and submits those concerns in writing, we have the Quality Review Team in place to review the report in light of those concerns” (pp. 9).

Many of the issues brought up in these Quality Reviews could have been resolved if the WWC had worked with study authors and program designers to ensure studies and programs were appropriately described. If the WWC would work with these individuals, it would help prevent false information from being publicized and published on different websites and blogs, which has occurred in the past. The WWC policy of sending these courtesy copies only 24 hours prior to publication increases the chance that erroneous information will be reported and spread through the education community. If, however, the reports were sent earlier, allowing study authors and program developers to review the reports and request Quality Reviews prior to publication, the chance of widely disseminating erroneous information would be greatly reduced.

3. The WWC will not remove a report from their website unless they believe it contains “clear” errors. However, the WWC’s judgment about what constitute a “clear” error can be questioned. Even when clear errors have been documented multiple times, the WWC did not remove these reports, allowing them to be shared on different websites and blogs.<sup>11</sup> Not only have clear errors been documented, but some of the errors are shocking in terms of their severity and the inability of WWC oversight and procedures to recognize them prior to publication.<sup>12</sup> Additionally they say the report should mention how their conclusions differ from the author(s), but this is not consistently done (e.g., Quality Review 2012-011, pp. 9). The WWC response to errors is disturbing due to their decisions to not remove erroneous errors immediately, not admitting to errors in reports, except in very small type at the end of the documents, and not explicitly saying how their conclusions differ from those of the authors.<sup>13</sup>

---

<sup>11</sup> For instance, in Quality Review 2012-012, the WWC claims SRA/McGraw Hill is the developer of *Reading Mastery*, while, in fact, they are the publisher. This is a clear error that could have easily been avoided (pp. 600). Additional errors include listing studies on *Reading Recovery* in a report on the effectiveness of *Reading Mastery*, incorrect date of publication of an intervention report, and listing an incorrect number of studies reviewed in the creation of an intervention report (Stockard & Wood 2013).

<sup>12</sup> The extremes of these errors include the WWC concluding a study reported positive effects for *Reading Recovery* when the authors of the study made the opposite conclusion. For further information on this error and others see Stockard and Wood (2013).

<sup>13</sup> In correspondence with the WWC regarding their publications the WWC admitted that they do not compare their results with other review organizations and individuals. The WWC was asked if the WWC reviews meta-analyses while conducting their reviews and whether they consider including the differing opinions on the effectiveness of the programs reviewed in the reports. In an email dated February 28th 2014 the WWC responded “the answer is no. As stated previously, other meta-analyses may differ in their inclusion criteria and



4. The WWC will use data from unfinished studies in the creation of their reports. However, in QR-2012-012, a representative of the WWC wrote

"Please note, in all cases, to be reviewed by the WWC, a study must be published research that is deemed "final" by the study authors and publicly available. 'In press' studies are acceptable provided they are publicly available."

Quality Review 2012-013 dealt with the WWC reporting on an unfinished study the author did not want to be reviewed because it was not final (p. 591). The inquirer objected to the inclusion of this study because it was incomplete and a final version was available. The handbook used at the time of the report (Version 2.1) makes no reference to a study needing to be deemed "final" by the study authors. The WWC determined not to revise the report because they reviewed both versions of the study. This is another example of the inconsistency of what the WWC does, what their representatives say, and their written policies.

### A Note on the WWC Response to the FOIA Request

In July 2013, the NIFDI Office of Research and Evaluation submitted a request through the FOIA process for all pertinent information related to WWC Quality Reviews (See Appendix A). A response was received to this request on November 19<sup>th</sup> 2013 in the form of a collection of emails related to each review. However, documentation related to 39 of the 62 Quality Reviews (63%) was missing. This missing pertinent information includes email attachments, emails, and internal communication. It appears that over 100 documents, all of which were referenced in the material provided, are missing from the data supplied by the WWC.

Many more documents would be expected to be included if the WWC made notes about the review process, how they made their decisions, and how they communicated about the issues raised in the requests for Quality Reviews. According to the WWC Procedures and Standards Handbooks, internal communication should have been well documented during the initial development of WWC reports. Even more internal communication would have been created during the Quality Review to ensure the original report followed WWC procedures. The *Procedures and Standards Handbook* (Version 3.0) states

---

standards. We do not report on or interpret the findings from other such reviews. We do list them in our citations so interested readers may find them."

“After an inquiry is forwarded to the QRT, a team member verifies that the inquiry meets criteria for a quality review and notifies the inquirer whether a review will be conducted. A reviewer is assigned to conduct an independent review of the study, examine the original review and relevant author and distributor/developer communications, notify the topic area team leadership of the inquiry, and interview the original reviewers. Throughout the process, all actions and conversations are documented and logged” (p. A4).<sup>14</sup>

None of these internal communications and notes regarding the Quality Reviews was included in the documents provided. This missing material is crucial to truly understand how these WWC reports are created, the review process, and how the concerns of the education community are managed. These documents would also provide additional information on whether the WWC consistently follows their own policies.

However, even though a substantial amount of information appears to have been withheld in the response to the FOIA request, there is no indication from the available data that the general conclusions presented in this report would alter substantially if the missing information were provided. In other words, there is no indication that the missing information would provide data to indicate that the concerns expressed by scholars in the Quality Review requests were unfounded. Nor is it likely that the additional information would indicate that the WWC’s response to these concerns differs substantially from the pattern summarized in this analysis.<sup>15</sup>

## Summary and Conclusion

The errors documented in these Quality Reviews contrast sharply with the WWC’s statement that they “provide accurate information on education research” (<http://ies.ed.gov/ncee/wwc/aboutus.aspx>). In fact, the number and variety of errors could be seen as severe and alarming. The errors range from simple mistakes of mischaracterizing an instructional program to including inappropriate studies in the review of an instructional program, resulting in inaccurate ratings of effectiveness. These Quality Reviews have clearly documented the multitude of errors in WWC reports and their negative consequences.

---

<sup>14</sup> This policy was also included in version 2.0, and 2.1 of the WWC *Procedures and Standards Handbook*.

<sup>15</sup> It seems especially unlikely that the problematic issues associated with the WWC review process and judgments that are documented in this technical report would be found to be less serious if the missing information were analyzed.

These errors have tarnished the reputations of the programs and their developers. Revisions to the reports have attempted to make up for these damages, but they cannot fully account for the damage done. In addition to harming the reputations of these programs and individuals, substantial damage arises from providing misleading information on the effectiveness of programs to school districts, potentially leading them to select less effective instructional programs for their students.

In fact, the WWC has commented on how their conclusions have influenced school districts' decisions on what programs to adopt. Quality Review 2010-010 (as well as others) was conducted due to concerns over the WWC not recommending effective programs for K-2 students and the conclusion that *Reading Mastery* had no discernable effects on student achievement.<sup>16</sup> In an email regarding Quality Review 2010-010, reference to the beneficial use of WWC reports in the selection of curricula by school districts was made. Deborah Reed of the WWC Quality Review Team wrote

"Your email could also be read to question the value of a WWC finding that a program has 'no studies meeting WWC standards.' On the contrary, WWC staff learned from school districts that these findings are helpful in decisions about adopting programs" (p. 11).

It appears that she is referring to the WWC intervention report on *Reading Mastery* and how the report has led to school districts adopting another program given a positive rating by the WWC. This demonstrates how the errors in WWC findings can directly affect the sales and reputation of programs like *Reading Mastery*. This is one example of many that demonstrates how the errors in WWC reports can have a negative effect on program developers, school districts looking to select instructional programs based on scientific evidence of effectiveness, and the ability of students to succeed and meet their potential.

As alarming as the errors documented in these Quality Reviews are, it is equally worrisome that the WWC policies, procedures, and standards are not clear nor consistently applied. The WWC has released a series of handbooks with subtle variations that document their standards and procedures, but fail to clearly describe their stance on important issues such as the fidelity of implementation. The vagueness of WWC policy on the fidelity of implementation in their handbook, and the inconsistent application of their policies has created greater difficulty for individuals to understand and work with the WWC to create

---

<sup>16</sup> A very large body of literature has documented the effectiveness of *Reading Mastery (RM)* and the WWC reports on *RM* have been heavily criticized. (see NIFDI Technical Reports 2014-3 and 2014-4 for the most recent analyses.)

studies that will meet their standards. If the WWC does not clearly define their policies they should no longer continue to publish reports that are not firmly supported by their published policies. The vagueness of WWC procedures and policies has prevented the WWC from becoming a transparent organization that can be trusted.

The WWC's relative disregard of the fidelity of implementation is shocking, appalling and confusing given their emphasis on strict experimental controls. The WWC has well documented their efforts to create gold standards for their reviews, but has failed to recognize the importance of fidelity of implementation to determine the effectiveness of a program. The accumulation of accurate data depends on the fidelity of implementation. If the data do not accurately reflect the program studied, what is the use of analyses of inaccurate data? This technical report heavily focused on the issue of fidelity of implementation, but could have analyzed various others issues discovered in these Quality Reviews. The focus on the fidelity of implementation is due to its importance in determining the effectiveness of a program and the overwhelming concern of those who requested Quality Reviews.

To date, the WWC has published 715 reports; and the data analyzed indicate that the WWC has long been aware of issues in the quality of their review process. In 2009, WWC board member Jon Baron requested a Quality Review in response to WWC reviews of 4 studies that he believed were misleading. In reference to an Expert Panel's review of the WWC's overall standards and methodologies, Jon Baron wrote

“I also support the Panel's primary recommendation – that the Department commission a comprehensive review of WWC. In particular, I'd suggest such a review examine how the WWC has applied its overall standards/methodologies to specific interventions and studies – an aspect the Panel did not have a chance to examine (except in a very limited way) given its short time frame. I suggest this because, based on reviewing many of the underlying studies that WWC reviewed, I think the WWC in a number of cases, has missed key flaws in study design and implementation – particularly in studies found to meet evidence standards without reservations” (QR2009-001, pp. 1).

The emails regarding this Quality Review revealed that Baron had previously stated at a board meeting that he believed the WWC had quality control problems that affect perhaps a third of its reviews. This estimate suggests severe issues with WWC standards, policies, and procedures; and that, if additional WWC reports were reviewed, one third of the reports

would contain errors. This error rate is frightening in terms of the resources used to develop these reports and the misinformation publicized. The WWC states their mission is “to be a central and trusted source of scientific evidence for what works in education” (WWC *Procedures and Standards Handbook* Version 3.0, pp. 1). The findings from this FOIA request suggest the WWC is not fulfilling their mission and to do so would require a significant re-evaluation of their standards and procedures for conducting their reports and Quality Reviews.

To truly understand the impact of the WWC reports and the errors within them, further analysis of all reports and the WWC’s notes related to their reviews is needed. However, with the information provided from the FOIA request and the publicly available information, three conclusions appear clear: 1) The WWC suffers from a lack of transparency in their policies and guidelines, 2) the conclusions they create in their reports can be misleading, and 3) the reports are potentially damaging to program developers and ultimately the success of students.

## **Appendix A: FOIA Request**

July 29, 2013

U.S. Department of Education  
Office of Management  
Regulatory Information Management Services  
400 Maryland Avenue, SW, LBJ 2W220

Washington, DC 20202-4536

ATTN: FOIA Public Liaison

Dear FOIA Public Liaison:

This is a request under the Freedom of Information Act.

I request that copies of documents containing information regarding requests sent to the What Works Clearinghouse for Quality Reviews be provided to me. I specifically request copies of 1) all requests for Quality Reviews since the inception of the WWC; 2) information on the disposition of each request; 3) documentation that provided the basis for each of these decisions, including memos, analyses, and correspondence; and 4) additional information about Quality Reviews and their dispositions about which I may be unaware.

I am a scholar and researcher affiliated with the National Institute for Direct Instruction (NIFDI), a nonprofit organization. NIFDI provides administrative and curricular support to schools and districts as they implement Direct Instruction (DI) programs and conducts, promotes and publicizes high-quality research on the effects of DI implementations. I will use the information in analyses of the policies and procedures of the WWC. The analyses will be shared with educational consumers, other researchers, and the developers of the Direct Instruction curriculum.

I request a waiver of all fees for this request. Disclosure of the requested information is in the public interest because it is likely to contribute significantly to public understanding of the operations or activities of the government and is not primarily in my commercial interest. Independent scholarly reviews of WWC reports indicate that they are often inaccurate. The decisions and actions of the WWC impact schools and students throughout the nation and ensuring that such decisions are accurate and transparent is clearly in the interest of the public.

Thank you for your consideration of this request.

## **Appendix B: Organizations, Study Authors, Program Developers, and Education Researchers who Requested Quality Reviews**

American Reading

ASES After School Program

AVID Center

California Charter School Association

Chesapeake Research Associates

Coalition for Evidence-Based Policy

Educational Consultant: Kevin Feldman, Ed. D., Florida Center for Reading Research

HighScope Educational Research Foundation

I Can Learn

Library of Congress

Madison Metropolitan School District

Math Learning Center

Matthew Carr, PhD

MCNY Bronx Extension Center

National Study of Comprehensive Teacher Induction

NBCT

New Century Education Software

New South Wales Department of Education and Training

National Institute for Direct Instruction (NIFDI)

Oregon Department of Education

Oregon Research Institute

Pearson Education

Program author: Siegfried Engelmann

Program author: Louisa C. Moats, author of LETRS (Language Essentials for Teachers of Reading and Spelling)

Program author: Zalman Usiskin, Professor Emeritus of Education, Director, UCSMP, The University of Chicago

Read Naturally

Reading Recovery

Reading & Learning Disability Specialist: Susan Shallenberger, Consultant, Cal Poly State University

Renaissance Learning

Researcher: Doug Carnine

School psychologist

Spouse of teacher

Study author

Success for All

Teacher

Traverse City Education Association

Westat

Wilson Language Training Corporation

ZOOM!



## **Appendix C:**

### **Programs and Studies That Were the Focus of Quality Reviews**

100 Book Challenge

Accelerated Reader/Reading Renaissance

AVID

Cognitive Tutor Algebra 1

Connected Mathematics Project

Corrective Reading/Power4 Kids

Direct Instruction

Doors to Discovery

ED's Student Mentoring Program

Failure Free Reading/Power4 Kids

HighScope

I CAN Learn

Kaplan Spell Read/Power4 Kids

Language Essentials for Teachers of Reading and Spelling (LETRS)

New Century Education Software

New Chance

Opening Eyes to Math

PALS program

Phonological Awareness Plus Letter Knowledge

Reading Mastery

Read Naturally

Reading Recovery

Scott Foresman Reading Street program

Success for All

SuccessMaker Enterprise

Wilson Reading/Power4 Kids

WWC Beginning Reading topic area

WWC Intervention Report on the University of Chicago School Mathematics Project

WWC Quick Review of the study "Charter School Performance in Los Angeles Unified School District: A District and Neighborhood Matched Comparison Analysis"

WWC Quick Review of "Head Start Impact Study: Final Report"

WWC Quick Review of "The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA experiment"

WWC Quick Review of "Reading and Language Outcomes of a Five-Year Randomized Evaluation of Transitional Bilingual Education"

WWC Quick Review of "Charter School Performance in New Jersey"

WWC Quick Review of the report "Impacts of Comprehensive Teacher Induction: Results from the First Year of a Randomized Controlled Study"

WWC Quick Review of "The New York City Aspiring Principals Program: A School-Level Evaluation"

WWC Quick Review of "Culture and the Interaction of Student Ethnicity with Reward Structure in Group Learning"

WWC Quick Review of "Charter Performance in New York City"

WWC Quick Review of "The Milwaukee Parental Choice Program Longitudinal Educational Growth Study Third Year Report"

WWC report on "Achievement Effects of Four Elementary School Math Curricula: Findings from First Graders in 39 Schools"

WWC Single Study Review of "Effects of Problem Based Economics on high school economics instruction"

## **Appendix D:**

### **Reactions to the WWC Reports and Quality Review Process**

In the process of reading the emails regarding the Quality Reviews it became evident that many individuals not only disagreed with the WWC's findings, but were outraged by their treatment by the WWC. The inquirers sought clarity relating to the WWC reports and the policies of the WWC. These individuals' frustrations were in response to the WWC not responding within a suitable time frame or within a time frame they provided, the lack of a response to specific issues, the unwillingness to speak via phone or in person, and general outrage over WWC conclusions and their experience working with the WWC. Below is a selection of excerpts from the Quality Reviews that highlight the level of frustration and disappointment individuals expressed during the Quality Review process.

1. Siegfried Engelmann (Program developer) to Paul Decker, President and CEO Mathematica, Policy Research: "The attached paper documents the incredible inadequacies of WWC (both its logic and its execution of what it claims it does). I don't engage in political games, but trust me, I am going to go after What Works with concerted energy, and I am going to take steps to see that the measurement community and the extended scientific community respond to this kind of casuistry you defend in your letter of September 8, 2008" (QR2008-008, pp. 131).
2. Caprice Young (Charter Association) to study author: "We need to find the live humans... who are we dealing with?" (QR2008-009, pp. 8).
3. Eric Stickney (Director of Educational Research, Renaissance Learning) to Mark Dynarski, Director, WWC: "Developers, researchers, and educators all have a stake in ensuring that the WWC produces the most accurate information possible. I strongly encourage you to reconsider the Bullock study in the Accelerated Reader Intervention report because it simply is not a valid study of Accelerated Reader" (QR2008-012, pp. 4).
4. Louisa Moats (author of LETRS instructional program) to Mark Dynarski, Director, WWC: "I am writing to protest the content and nature of the research summary, and to request that it be rewritten to more truthfully reflect what was done by AIR and what was found in the

study. What you have is inaccurate, misleading, and undeservedly harmful to the reputation of LETRS" (2008-013, pp. 1).

5. Clinton Brass (Library of Congress, Congressional Research Service) to the WWC: I think the letter you sent does little or nothing to address the matters my e-mail brought up" (QR2009-008, pp. 9). "It seems to me that a case could be made that WWC is engaged in false advertising, for this Quick Review, if it does not highlight these things. Simply because the WWC Evidence Standards do not address this situation does not seem to be an adequate response. If we go back to what experimental design is all about, then the WWC's lack of standards here makes the WWC's rating and commentary, that the RCT is well-implemented (and therefore reliable as a guide to "what works"), potentially false. Congress and the President are making (and have already made) public policy on the basis of these study findings" (pp. 10).

6. Eric Hurley (study author) to WWC: "I am writing to offer some feedback on the article as it is written because I think it mischaracterizes our findings a bit in the section: what did the authors report. It says: 'Among African-American, students offered a reward based on individual performance had lower scores than those offered no reward.' This language completely misses the point by failing to mention what was present in the learning context in which AA students thrived. The study is after all, an examination of what factors students find motivating in group learning. We found that AA & EA students were motivated by different things. It should say something more like: AA students performed better in the learning context that stressed sharing and cooperation (without extrinsic reward) than in the one which stressed interpersonal competition for an extrinsic reward. Given your mission to provide educators with tools to make informed decisions, I hope that you will be able to make appropriate changes to the document. Otherwise a teacher might be lead to for example, simply omit extrinsic reward without emphasize sharing and cooperation, which could be disastrous" (QR2010-007, pp. 1).

7. Stephanie Metzger (school psychologist) to WWC: "I did not submit a study. I submitted a summary of a study, because the WWC failed to understand that the Iverson and Turner study that met the WWC standards with reservations concluded that Reading Recovery would be effective IF it was less rigid and its proponents willing to incorporated systematic phonics/phonemic awareness activities as part of its methodologies. I submitted the summary because I do not understand how the WWC concluded that Reading Recovery is an 'effective' beginning reading program. Its theoretical underpinnings are the subject of controversy, it does not stay current with research, it relied on whole language methods, and the studies that met the WWC's standards (a total of 4) were all conducted by Reading

Recovery advocates. I have found serious flaws in at least 3 of the studies, and it amazes me that the WWC reached its conclusions about "effectiveness" on these studies. I have asked repeatedly about this, was told that I would get a reply to my questions, and to date, have never received such. I believe, as do most serious reading educators (as well as respected researchers) that the WWC failed to live up to its promise to be a trusted entity examining effectiveness of educational programs" (QR2010-009, pp. 9).

8. John Griffin to WWC: "My goodness, what good does your organization do if none of the Reading Programs studies meet your standards. My wife teaches and loves the results from Reading Mastery. Really, either you folks' standards are impossible or the textbook companies are not interested in having you make a determination. Either way you are of no help... after all of this time you can't give the slightest bit of advice as to which programs work best." (QR2010-010, pp. 1) "After 8 years of study you have not found any regular reading program that fits the requirements of your research criteria that has had a positive impact of student learning? Didja ever think that maybe those requirements are so strict as to make the effort a hopeless cause?... So I am sorry to give you a hard time with this because I doubt if it has been your personal responsibility to do things this way. But you guys are not being paid to study the issue forever, you are being paid to help school districts find those courses that are most effective at teaching kids. At least in the area of beginning reading for the average classroom. You have not provided any of the information that you were set up to provide. Reading recovery is the exception but that does not apply to the regular classroom. Who do I need to talk to? My senators? My congressman? Arne Duncan? Who can change things so you can do the job you were set up to do?... Go ahead, try to find one review on your site of a beginner reading program that works best for average k-2 students" (p. 2).

9. Bridgett Long (study author) to WWC: "There appears to be a HUGE mistake in the report as it suggests we had high rates of sample attrition and that this attrition differed substantially across the treatment and control groups. This is absolutely untrue, and we are puzzled how that conclusion was drawn (of the many who have reviewed the paper, none have come to a similar conclusion). None of our tables or results point to the concern highlighted in the WWC report, and we are baffled at the percentages reported on page 2 concerning sample attrition. How were these even calculated? It is worth noting that the way we track our sample using administrative data prevents such sample attrition. We are extremely concerned that the WWC is about to put out an inaccurate report about our work" (QR2010-011, pp. 1).

10. John Witte (study author) to Mark Dynarski, Director, WWC: "Clearly the reviewers only glanced over our crucial baseline report and failed to see that most of it provides exactly the evidence of equivalence that they claim the study lacks... As to the one-page report, it is hard to believe so many errors could be made in so few words" (2010-016, pp. 3).

11. Keith Smolkowski (Research Scientist, Oregon Research Institute) to WWC: "It took about five minutes reading the papers to see that the reviewers got this one wrong. However, the misinformation that this review presents could potentially undo much of what the WWC strives to achieve" (2011-001, pp. 15).

12. James Cline (Classroom teacher) to Susanne James-Burdumy, Principle Investigator, Methodological Lead - WWC Adolescent Literacy Review Team, Beginning Reading Review Team: "I am writing you with a real concern about the What Works Clearing House website. While I have shared the website with other educators and encouraged them to use this site for RTI research, I can no longer encourage the use of the site while it misrepresents one of my favorite intervention tools, Read Naturally" (2012-006, pp. 2).

13. Susan Shallenberger (Reading Specialist, Learning Disability Specialist, Consultant, Cal Poly State University) to WWC: "I have read your report and I understand that you followed your usual procedures for your studies. But as a serious researcher, and someone who wants to know the truth about programs that can help children, when so few programs have succeeded, I am still not satisfied with your WWC evaluation... Your evaluation that it is not 'what works' based on your inadequate approach to its implementation, is a disservice to all teachers and students who desperately seek meaningful help. If your evaluations lead to teachers not choosing the program, you share in the responsibility for all of the students they will not reach. Please evaluate the program again, using the procedural steps and consistency that are part of following the program" (2012-010, pp. 3).

14. Macke Raymond (study author) to WWC: "So it appears that you err in favor of damaging researchers rather than in giving them their best chance to show you why your report is inaccurate, Nice!" (2012-016, pp. 7).

## Appendix E: Quality Reviews and WWC Conclusions

<u>Quality Review</u>	<u>Changes Made</u>	<u>Program/Study Reviewed</u>
QR2007-001	No	<i>Reading Recovery</i>
QR2007-002	Yes. Report revised to clarify the timeframe of the literature review.	<i>100 Book Challenge</i>
QR2007-003	No	Direct Instruction/ <i>Reading Mastery</i>
QR2008-001	Yes. Report revised to correct previous conclusion about attrition levels.	Roberts & Neal (2004). "Relationship among preschool English language learners' oral proficiency in English, instructional experience and literacy development"
QR2008-002	Yes. Report revised to correct the classification of the program.	<i>Opening Eyes to Math</i>
QR2008-003	No	New Century Education Software
QR2008-004	Yes. Report revised to correct developer and publisher information.	First-Grade Peer Assisted Literacy Strategies
QR2008-005	No	<i>I CAN Learn</i>
QR2008-006	No	<i>Reading Recovery</i>
QR2008-007	No	P. Weisberg (1988). "Direct Instruction in the Preschool "
QR2008-008	No	Direct Instruction/ <i>Reading Mastery</i>

QR2008-009	No	<b>Toney, A., &amp; Murdock, D. (2008).</b> "Charter School Performance in Los Angeles Unified School District: A District and Neighborhood Matched Comparison Analysis"
QR2008-010	Yes. Report revised to clarify descriptions of studies included in the review.	<i>Success for All</i>
QR2008-011	Yes. Report revised to clarify the level of implementation in a study and the description of a study.	<i>Read Naturally</i>
QR2008-012	No	<i>Accelerated Reader/Reading Renaissance</i>
QR2008-013	Yes. Report revised to modify the description of the study to reflect level of implementation.	Garet et al. (2008). "The impact of two professional development interventions on early reading instruction and achievement"
QR2009-001	Yes. Report revised due to error with rating of effectiveness.	<i>New Chance, Reading Recovery, Wilson Reading, Kaplan Spell Read, Corrective Reading, Failure Free Reading</i>
QR2009-002	Yes. Report revised to clarify the level of implementation in a study.	<i>Wilson Reading System</i>
QR2009-003	Yes. Report revised to remove inaccurate description of research on High/Scope Preschool Key Experiences Series, Booklets and Videos.	<i>High Scope</i>
QR2009-004	No	<i>Reading Recovery</i>



QR2009-005	Yes. Report revised to modify conclusions on the attrition rates of two studies.	<i>Doors to Discovery</i>
QR2009-006	No	<i>Reading Recovery</i>
QR2009-007	No	<i>Reading Recovery</i>
QR2009-008	Conclusion not provided.	U.S. Department of Education's Student Mentoring Program
QR2010-001	Conclusion not provided.	<b>Agodini, R. et al. (2009).</b> "Achievement Effects of Four Elementary School Math Curricula"
QR2010-002	Conclusion not provided.	<b>Glazerman, S. et al. (2008).</b> "Impacts of Comprehensive Teacher Induction: Results from the First Year of a Randomized Controlled Study"
QR2010-003	Yes. Report revised to clarify the cutoff date for the literature search.	<i>Scott Foresman Reading Street, SuccessMaker Enterprise, Connected Mathematics Project</i>
QR2010-004	Yes. Report revised to clarify the cutoff date for the literature search.	<i>100 Book Challenge</i>
QR2010-005	Conclusion not provided.	Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2009). "The New York City Aspiring Principals Program: A School-Level Evaluation"
QR2010-006	Conclusion not provided.	Cognitive Tutor Algebra 1

QR2010-007	Conclusion not provided.	<b>Hurley, E. A., Allen, B. A., &amp; Boykin, A. W. (2009).</b> "Culture and the Interaction of Student Ethnicity with Reward Structure in Group Learning"
QR2010-008	Yes. Report revised to note that one study reviewed did use a control group.	<i>Success for All</i>
QR2010-009	No	<i>Reading Recovery</i>
QR2010-010	No	<i>Reading Mastery</i>
QR2010-011	Yes. Report revised, but specific changes are unknown because the WWC did not send their formal response to this issue.	Bettinger, E. P. et al. (2009). "The role of simplification and information in college decisions: Results from the H&R Block FAFSA experiment"
QR2010-012	Conclusion not provided.	Center for Research on Education Outcomes (2010). "Charter School Performance in New York City"
QR2010-013	Yes. 4 reports were revised to more accurately describe the <i>Read Naturally</i> strategy and <i>Read Naturally</i> programs. Reports were also revised to clarify the level of implementation.	<i>Read Naturally</i>
QR2010-014	Yes. Report revised to clarify attrition rates.	Puma, M. et al. (2010). "Head Start Impact Study: Final Report"
QR2010-015	No	<i>Reading Recovery</i>
QR2010-016	Conclusion not provided.	<b>Cowen, J. M. et al. (2010).</b> "The Milwaukee Parental Choice Program Longitudinal Educational Growth Study Third Year Report"

QR2010-017	Yes. Report revised to clarify WWC findings.	Slavin, R. E. et al. (2010) "Reading and Language Outcomes of a Five-Year Randomized Evaluation of Transitional Bilingual Education"
QR2010-018	Yes. Report revised to clarify the description of a study. Specific details not available because this information was not supplied by the WWC.	AVID
QR2011-001	Yes. Report revised to remove a study from the analysis of <i>Reading Mastery</i> and change the summary judgment.	<i>Reading Mastery</i>
QR2011-002	No	<i>University of Chicago School Math Project</i>
QR2012-001	Yes. Report revised to include additional data from a study previously not reported.	<i>I CAN Learn</i>
QR2012-002	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>
QR2012-003	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>
QR2012-004	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>
QR2012-005	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>
QR2012-006	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>
QR2012-007	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>

QR2012-008	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>
QR2012-009	Yes. Report revised to clarify the level of implementation in one study.	<i>Read Naturally</i>
QR2012-010	Yes. Report revised to clarify the timeframe of the literature review.	<i>Read Naturally</i>
QR2012-011	Yes. Report revised to correct errors about study findings and clarify WWC procedures in their analysis.	Finkelstein, N. (2010). "Effects of Problem Based Economics on high school economics instruction"
QR2012-012	Yes. Report revised to remove a study from the analysis of <i>Reading Mastery</i> and change the summary judgment.	<i>Reading Mastery, Reading Recovery</i>
QR2012-013	Yes. WWC database updated to include the preliminary and final version of the study.	Jackson, C. K. (2007). "A little now for a lot later: A look at a Texas advanced placement incentive program."
QR2012-014	Yes. Report revised to remove a study from the analysis of <i>Reading Mastery</i> and change the summary judgment.	<i>Reading Mastery</i>
QR2012-015	Yes. Report revised to provide greater clarity on the study findings.	Slavin, R. E. (2010). "Reading and Language Outcomes of a Five-Year Randomized Evaluation of Transitional Bilingual Education"
QR2012-016	No	Center for Research on Education Outcomes (2012). "Charter School Performance in New Jersey"

QR2013-001	Yes. <i>Reading Mastery</i> intervention report revised to note the report was revised from a previous version that contained errors. <i>Reading Recovery</i> intervention report revised to clarify the study disposition for Baenen et al. (1997).	<i>Reading Mastery, Reading Recovery</i>
QR2013-002	Yes. Report revised to clarify the description of <i>Read Naturally</i> .	<i>Read Naturally</i>

## References

- Benner, G. J., Nelson, J. R., Stage, S. A., & Ralston, N. C. (2010). The influence of fidelity of implementation on the reading outcomes of middle school students experiencing reading difficulties. *Remedial and Special Education Online First*, published February 12, 2010.
- Carlson, C. D., & Francis, D. J. (2002). Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE). *Journal of Education for Students Placed at Risk*, 7(2), 141-166.
- Engelmann, S. (2004). The Dalmatian and its spots: Why research-based recommendations fail logic 101. *Education Week*, 23(20), 28. (Reprint available at <http://zigsite.com/Dalmatian.htm>).
- Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Gersten, R., Carnine, D., Zoref, L. & Cronin, D. (1986). A multifaceted study of change in seven inner-city schools. *The Elementary School Journal*, 86(3), 257-276.
- Gersten, R. M., Carnine, D. W., & Williams, P. B. (1982). Measuring implementation of a structured educational model in an urban school district: An observational approach. *Educational Evaluation and Policy Analysis* 4(1), 67-79.
- Ross, S. M., Nummery, J. A., Goldfeder, E., McDonald, A., Racho, R., Hornbeck, M., & Fleishman, S. (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, 357-368.
- Stockard, J. (2008). The What Works Clearinghouse Beginning Reading Reports and Rating of Reading Mastery: An Evaluation and Comment. NIFDI Technical Report 2008-4.
- Stockard, J. (2010). An Analysis of the Fidelity Implementation Policies of the What Works Clearinghouse, *Current Issues in Education*, 13(4).
- Stockard, J., & Wood, T. W. (2013). The What Works Clearinghouse review process: An analysis of errors in two recent reports. *NIFDI Technical Report 2013-4*.
- WWC Procedures and Standards Handbook Version 1.0. Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_version1\\_standards.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_version1_standards.pdf)
- WWC Procedures and Standards Handbook Version 2.0. Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v2\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_standards_handbook.pdf)

WWC Procedures and Standards Handbook Version 2.1. Retrieved from

[http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v2\\_1\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf)

WWC Procedures and Standards Handbook Version 3.0. Retrieved from

[http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf)