

From: keiths@ori.org  
Sent: Wednesday, October 24, 2012 10:10 PM  
To: info@whatworks.ed.gov  
Subject: IES WWC Website: Contact Us: Publications, Intervention Reports,  
Reference ID Number: 1186237386

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

From: keiths@ori.org

Message: Dear WWC,

I am writing today about the recent review of Reading Mastery for students with learning disabilities, which I found to be quite problematic. I have some experience with WWC and Reading Mastery (RM). The review of Gunn, Biglan, Smolkowski, and Ary (2000), which met WWC evidence standards, led the WWC to recommend RM for English learners. Indeed, it was one of the top-rated programs. Because of the very different results from the review of RM for students with learning disabilities, I took a closer look.

While talking about this report with my colleagues at the National Institute for Direct Instruction as well as other researchers, I also discovered some additional discrepancies with the WWC procedures. In an effort to improve the process, I have been encouraged by everyone to forward my concerns to you. Below I raise my concerns with the RM review as well as issues that may lead to inconsistent and potentially inadequate reviews of other programs.

The review of RM for students with learning disabilities suffers from serious flaws. In particular, the report relies on two studies. Cooke et al (2004) compares RM to its very similar sister program, Horizons. Herrera et al. (1997) compares RM to a comparison condition that included RM plus a supplemental program called the Stabilized Learning System (SLS). The review concluded that RM did not perform better than Horizons. This is what one would expect given that they are highly aligned and designed by the same people. The second paper is not a valid comparison, as it compares RM to RM plus SLS and erroneously concludes that RM may have potentially negative effects.

If RM turns out to be an effective way to teach children with learning disabilities and if schools, instead, turn to less effective programs, the WWC is then actively producing change toward less effective programs in schools. I believe RM is an effective instructional program. In our research on RM, we included many more students than English learners. The students with the lowest performance on literacy and preliteracy and exposed to RM made either equal or greater gains than control students (see Gunn, Smolkowski, Biglan, Black, & Blair, 2005, in J. of Special Education). Our poor readers were selected because they performed in the bottom 15% of students in their schools, and although we did not classify students by disability status, 17.1% of our total sample either received special education or Chapter services for reading. When we examined poor-reader status as a moderator, we found that our poor readers in the RM condition made greater gains than control students. The difference between conditions was either the same or greater than that for students who began the year below grade level but were not designated as poor readers. This appeared to be true regardless of language status. From our research with RM, then, it appears that it is likely an effective approach, yet WWC is currently recommending to educators nationwide that they remove the program from their shelves based on a flawed review.

The WWC RM report appeared to miss many papers on the RM. Granted, not all of them would qualify for this review, but list of references at the end of the review is missing numerous papers. Books such as Engelmann and Carnine's (2011) *Could John Stuart Mill have Saved Our Schools* and Adams and Engelmann's (1996) *Research on Direct Instruction* include many papers not cited in the literature reviewed. As I mentioned, not all of those would meet criteria for evaluation, but the reference list on their page 8 and 9 should have included more papers.

The limited search for papers raises another question. When I spoke with my

colleagues at the National Institute for Direct Instruction, I asked if they had been contacted with a request to provide all relevant papers. They did not believe any attempt at contact had been made. For other interventions, however, WWC representatives have contacted evaluators and program authors to ensure that they have all relevant papers before they begin their review. For example, for a recent review of First Step to Success, WWC contacted Hill Walker and other colleagues to ensure they included all published sources. With respect to the RM, it appears that the review began without a complete literature search.

Interestingly, for the review of First Step, the WWC representatives refused to include presentations and IES project reports. One researcher mentioned that he had included a draft paper with his annual report to IES. He also suggested that WWC examine several presentations. They were told that WWC does not accept presentations or annual progress reports for grants. This, however, directly contradicts Jill Constantine's response to an earlier inquiry that I made. In my prior inquiry, I specifically asked about progress reports and publications and was told that "every study, regardless of prior publication, is subject to the full WWC review." There appears to be some inconsistency with how WWC literature searches are conducted and with what is considered worthy of review.

The WWC procedures also state that the review will not include research that compares a program to itself or to one that is substantially similar. WWC procedures specify that two reviewers trained in WWC standards independently review the study design and that any differences in reviews are reconciled with a senior WWC reviewer. It appears that all of these reviewers appeared to miss the fact that the two papers included in the WC review violate the prohibition against comparing a program to another substantially similar program. I am surprised that with just two papers it was not determined that the review relied on inappropriate comparisons.

The WWC report also excluded key details, especially for the Herrera paper. For example, I could not find in the report any mention that the Stabilized Learning System (SLS) was a supplemental program that added 90 minutes per day of instruction. The WWC report makes it sound like the study compared SLS to RM, which is clearly incorrect. Herrera et al. (1997) explain that SLS "was implemented forty-five minutes per school day as a supplemental to the usual and customary curriculum" (p. 78). They go on to say that the comparison condition also received "the usual and customary prescribed curriculum for the school day, Reading Mastery's explicit methodology" (p. 78). Hence the study found that RM did not perform as well as RM plus an extra 45 minutes of daily supplemental instruction. The report leaves out these critical details.

The WWC review apparently allowed only studies in which 50% of students were labeled with learning disabilities. This appears to be an arbitrary standard that allows the review to include a study of 20 total students and 12 with LD yet exclude a study with 1000 total students and 480 with LD. The second study is clearly superior, offering substantially improved precision and statistical power. Why would WWC exclude such studies from their review? From my experience with other WWC reviews, however, I am not sure that this or similar criteria have been consistently applied.

As you know, LD identification is notoriously problematic. A recent discussion on the Society for the Scientific Study of Reading e-mail server outlined many problems. States, districts, and schools use different definitions and procedures, and those procedures have changed with time and the adoption of RTI. This can lead to considerably different samples of students from paper to paper. Our study by Gunn et al. (2005), reviewed by WWC, included students with learning disabilities, but we generally ignored the label due to the problems just mentioned. We selected students reading below the 15th percentile within each school and labeled them "poor readers." The WWC report does not identify the criteria for LD, and it appears that studies are included if the report used the label "learning disabilities" but not if the study failed to use those terms.



When the WWC encounters effects that are not statistically significant (e.g., WWC RM report, p. 5), they compute effect sizes. The challenge with this procedure is that the statistically nonsignificant test means that the effect is not sufficiently stable to interpret. Effect sizes were often recommended for research reports not to make statements about statistically nonsignificant findings but rather to quantify the clinical value of fairly small effects from large studies. Many large studies might be able to detect tiny effects (e.g.,  $g = .02$ ), and by recommending that authors report such effect sizes, readers could determine if such a small effect was meaningful. The use of effect sizes to quantify nonsignificant, unreliable results extends an unfortunate myth about statistical analysis. The effect mentioned on page 6 of the RM report can only be substantively important if it is also statistically reliable.

Stanford statistician Helena Kraemer and her colleagues (2006, in Archives of General Psychiatry) discuss the misuse of effect sizes from underpowered pilot studies and the unintended consequences in some detail. Nonsignificant differences imply large confidence bounds on effect sizes. The cited effect size,  $-0.33$ , may well have a confidence interval from  $-1.0$  to  $+0.33$ , although such information is not reported. Such problematic reporting has led a number of statisticians to call for even higher standards for what counts as statistically significant (e.g.,  $p < .001$  or  $.0001$ ; see Ioannidis, 2005, in PLoS Medicine; Simmons, Nelson, & Simonsohn, 2011, in Psychological Science; Sterne & Smith, 2001, in the British Medical Journal). "Ultimately, a study with low power has very little chance of advancing scientific knowledge" (Fitzmaurice, 2002, p. 290, in Nutrition). In the present case, the WWC is trying to make a definitive statement about the value of a program but using substandard criteria for that decision. The WWC should instead use high standards in its attempt to advance the field of education.

Having talked with representatives the National Institute for Direct Instruction, I learned that they have asked that the WWC remove the report and

that their request has been declined. This irresponsible lack of action has led some educators to begin to question a potentially valuable program based on a very poor review. 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. Lewandowsky and colleagues (2012, in Psychological Science in the Public Interest) discuss the tremendous effort required to correct misinformation once it becomes disseminated to the public. Indeed, their review is quite consistent with a principal among Direct Instruction proponents: it takes 5 to 10 times as much instruction to reteach a misrule as it does to teach a rule correctly to begin with. Citing over a dozen studies, Lewandowsky et al. report that "research using this paradigm has consistently found that retractions rarely, if ever, have the intended effect of eliminating reliance on misinformation, even when people believe, understand, and later remember the retraction" (p. 114). The WWC review of RM has already been reference by ERIC, and educators have started to react negatively toward the program. In time, no retraction or public statement on the WWC web page will undo the damage.

Thank you for considering these issues.

**From:** WhatWorks  
**Sent:** Thursday, October 25, 2012 11:19 AM  
**To:** 'keiths@ori.org'  
**Subject:** What Works Clearinghouse (WWC 3917)

Hello,

Thank you for contacting the What Works Clearinghouse (WWC). We have received your email below. WWC staff are reviewing your request and will prepare a response.

#### What Works Clearinghouse

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

#### -----Original Message-----

From: keiths@ori.org [mailto:keiths@ori.org]  
Sent: Wednesday, October 24, 2012 10:10 PM  
To: info@whatworks.ed.gov  
Subject: IES WWC Website: Contact Us: Publications, Intervention Reports, Reference ID Number: 1186237386

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

From: keiths@ori.org

Message: Dear WWC,

I am writing today about the recent review of Reading Mastery for students with learning disabilities, which I found to be quite problematic. I have some experience with WWC and Reading Mastery (RM). The review of Gunn, Biglan, Smolkowski, and Ary (2000), which met WWC evidence standards, led the WWC to recommend RM for English learners. Indeed, it was one of the top-rated programs. Because of the very different results from the review of RM for students with learning disabilities, I took a closer look.

While talking about this report with my colleagues at the National Institute for Direct Instruction as well as other researchers, I also discovered some additional discrepancies with the WWC procedures. In an effort to improve the process, I have been encouraged by everyone to forward my concerns to you. Below I raise my concerns with the RM review as well as issues that may lead to inconsistent and potentially inadequate reviews of other programs.

The review of RM for students with learning disabilities suffers from serious flaws. In particular, the report relies on two studies. Cooke et al (2004) compares RM to its very similar sister program, Horizons. Herrera et al. (1997) compares RM to a comparison condition that included RM plus a supplemental program called the Stabilized Learning System (SLS). The review concluded that RM did

not perform better than Horizons. This is what one would expect given that they are highly aligned and designed by the same people. The second paper is not a valid comparison, as it compares RM to RM plus SLS and erroneously concludes that RM may have potentially negative effects.

If RM turns out to be an effective way to teach children with learning disabilities and if schools, instead, turn to less effective programs, the WWC is then actively producing change toward less effective programs in schools. I believe RM is an effective instructional program. In our research on RM, we included many more students than English learners. The students with the lowest performance on literacy and preliteracy and exposed to RM made either equal or greater gains than control students (see Gunn, Smolkowski, Biglan, Black, & Blair, 2005, in *J. of Special Education*). Our poor readers were selected because they performed in the bottom 15% of students in their schools, and although we did not classify students by disability status, 17.1% of our total sample either received special education or Chapter services for reading. When we examined poor-reader status as a moderator, we found that our poor readers in the RM condition made greater gains than control students. The difference between conditions was either the same or greater than that for students who began the year below grade level but were not designated as poor readers. This appeared to be true regardless of language status. From our research with RM, then, it appears that it is likely an effective approach, yet WWC is currently recommending to educators nationwide that they remove the program from their shelves based on a flawed review.

The WWC RM report appeared to miss many papers on the RM. Granted, not all of them would qualify for this review, but list of references at the end of the review is missing numerous papers. Books such as Engelmann and Carnine's (2011) *Could John Stuart Mill have Saved Our Schools* and Adams and Engelmann's (1996) *Research on Direct Instruction* include many papers not cited in the literature reviewed. As I mentioned, not all of those would meet criteria for evaluation, but the reference list on their page 8 and 9 should have included more papers.

The limited search for papers raises another question. When I spoke with my colleagues at the National Institute for Direct Instruction, I asked if they had been contacted with a request to provide all relevant papers. They did not believe any attempt at contact had been made. For other interventions, however, WWC representatives have contacted evaluators and program authors to ensure that they have all relevant papers before they begin their review. For example, for a recent review of First Step to Success, WWC contacted Hill Walker and other colleagues to ensure they included all published sources. With respect to the RM, it appears that the review began without a complete literature search.

Interestingly, for the review of First Step, the WWC representatives refused to include presentations and IES project reports. One researcher mentioned that he had included a draft paper with his annual report to IES. He also suggested that WWC examine several presentations. They were told that WWC does not accept presentations or annual progress reports for grants. This, however, directly contradicts Jill Constantine's response to an earlier inquiry that I made. In my prior inquiry, I specifically asked about progress reports and publications and was told that "every study, regardless of prior publication, is subject to the full WWC review." There appears to be some inconsistency with how WWC literature searches are conducted and with what is considered worthy of review.

The WWC procedures also state that the review will not include research that compares a program to itself or to one that is substantially similar. WWC procedures specify that two reviewers trained in WWC standards independently review the study design and that any differences in reviews are reconciled with

a senior WWC reviewer. It appears that all of these reviewers appeared to miss the fact that the two papers included in the WC review violate the prohibition against comparing a program to another substantially similar program. I am surprised that with just two papers it was not determined that the review relied on inappropriate comparisons.

The WWC report also excluded key details, especially for the Herrera paper. For example, I could not find in the report any mention that the Stabilized Learning System (SLS) was a supplemental program that added 90 minutes per day of instruction. The WWC report makes it sound like the study compared SLS to RM, which is clearly incorrect. Herrera et al. (1997) explain that SLS “was implemented forty-five minutes per school day as a as a supplemental to the usual and customary curriculum” (p. 78). They go on to say that the comparison condition also received “the usual and customary prescribed curriculum for the school day, Reading Mastery’s explicit methodology” (p. 78). Hence the study found that RM did not perform as well as RM plus an extra 45 minutes of daily supplemental instruction. The report leaves out these critical details.

The WWC review apparently allowed only studies in which 50% of students were labeled with learning disabilities. This appears to be an arbitrary standard that allows the review to include a study of 20 total students and 12 with LD yet exclude a study with 1000 total students and 480 with LD. The second study is clearly superior, offering substantially improved precision and statistical power. Why would WWC exclude such studies from their review? From my experience with other WWC reviews, however, I am not sure that this or similar criteria have been consistently applied.

As you know, LD identification is notoriously problematic. A recent discussion on the Society for the Scientific Study of Reading e-mail server outlined many problems. States, districts, and schools use different definitions and procedures, and those procedures have changed with time and the adoption of RTI. This can lead to considerably different samples of students from paper to paper. Our study by Gunn et al. (2005), reviewed by WWC, included students with learning disabilities, but we generally ignored the label due to the problems just mentioned. We selected students reading below the 15th percentile within each school and labeled them “poor readers.” The WWC report does not identify the criteria for LD, and it appears that studies are included if the report used the label “learning disabilities” but not if the study failed to use those terms.

When the WWC encounters effects that are not statistically significant (e.g., WWC RM report, p. 5), they compute effect sizes. The challenge with this procedure is that the statistically nonsignificant test means that the effect is not sufficiently stable to interpret. Effect sizes were often recommended for research reports not to make statements about statistically nonsignificant findings but rather to quantify the clinical value of fairly small effects from large studies. Many large studies might be able to detect tiny effects (e.g.,  $g = .02$ ), and by recommending that authors report such effect sizes, readers could determine if such a small effect was meaningful. The use of effect sizes to quantify nonsignificant, unreliable results extends an unfortunate myth about statistical analysis. The effect mentioned on page 6 of the RM report can only be substantively important if it is also statistically reliable.

Stanford statistician Helena Kraemer and her colleagues (2006, in Archives of General Psychiatry) discuss the misuse of effect sizes from underpowered pilot studies and the unintended consequences in some detail. Nonsignificant differences imply large confidence bounds on effect sizes. The cited effect size,  $-0.33$ , may well have a confidence interval from  $-1.0$  to  $+0.33$ , although such information is not reported. Such problematic reporting has led a number of statisticians to call for even higher standards for what counts as statistically significant (e.g.,  $p < .001$  or  $.0001$ ; see Ioannidis, 2005, in PLoS

Medicine; Simmons, Nelson, & Simonsohn, 2011, in *Psychological Science*; Sterne & Smith, 2001, in the *British Medical Journal*). “Ultimately, a study with low power has very little chance of advancing scientific knowledge” (Fitzmaurice, 2002, p. 290, in *Nutrition*). In the present case, the WWC is trying to make a definitive statement about the value of a program but using substandard criteria for that decision. The WWC should instead use high standards in its attempt to advance the field of education.

Having talked with representatives the National Institute for Direct Instruction, I learned that they have asked that the WWC remove the report and that their request has been declined. This irresponsible lack of action has led some educators to begin to question a potentially valuable program based on a very poor review. 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. Lewandowsky and colleagues (2012, in *Psychological Science in the Public Interest*) discuss the tremendous effort required to correct misinformation once it becomes disseminated to the public. Indeed, their review is quite consistent with a principal among Direct Instruction proponents: it takes 5 to 10 times as much instruction to reteach a misrule as it does to teach a rule correctly to begin with. Citing over a dozen studies, Lewandowsky et al. report that “research using this paradigm has consistently found that retractions rarely, if ever, have the intended effect of eliminating reliance on misinformation, even when people believe, understand, and later remember the retraction” (p. 114). The WWC review of RM has already been reference by ERIC, and educators have started to react negatively toward the program. In time, no retraction or public statement on the WWC web page will undo the damage.

Thank you for considering these issues.

**From:** Doug Carnine <dcarnine@uoregon.edu>  
**Sent:** Thursday, October 25, 2012 9:10 PM  
**To:** Joy.Lesnick@ed.gov  
**Cc:** WhatWorks; jstockard@nifdi.org; Keith Smolkowski  
**Subject:** Reading Mastery report by WWC--Comments by Keith Smolkowski

10-10-12

Dear Dr. Lesnick:

I am following up on my 10-10-12 email to you. Although I have no financial interest in Reading Mastery, I have close ties to the authors and was involved with Direct Instruction for my entire career. This link to Direct Instruction could bias my reading of the WWC reports. Keith Smolkowski provides the perspective of well-regarded researcher who has no ties to Reading Mastery. His letter to WWC appears below; I have cc'ed him on this email. I hope his comments will spur the re-evaluation of the WWC's report on Reading Mastery that I assume is still active on their web site.

Sincerely,

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

Dear WWC,

I am writing today about the recent review of Reading Mastery for students with learning disabilities, which I found to be quite problematic. I have some experience with WWC and Reading Mastery (RM). The review of Gunn, Biglan, Smolkowski, and Ary (2000), which met WWC evidence standards, led the WWC to recommend RM for English learners. Indeed, it was one of the top-rated programs. Because of the very different results from the review of RM for students with learning disabilities, I took a closer look.

While talking about this report with my

colleagues at the National Institute for Direct Instruction as well as other researchers, I also discovered some additional discrepancies with the WWC procedures. In an effort to improve the process, I have been encouraged by everyone to forward my concerns to you. Below I raise my concerns with the RM review as well as issues that may lead to inconsistent and potentially inadequate reviews of other programs.

The review of RM for students with learning disabilities suffers from serious flaws. In particular, the report relies on two studies. Cooke et al (2004) compares RM to its very similar sister program, Horizons. Herrera et al. (1997) compares RM to a comparison condition that included RM plus a supplemental program called the Stabilized Learning System (SLS). The review concluded that RM did not perform better than Horizons. This is what one would expect given that they are highly aligned and designed by the same people. The second paper is not a valid comparison, as it compares RM to RM plus SLS and erroneously concludes that RM may have potentially negative effects.

If RM turns out to be an effective way to teach children with learning disabilities and if schools, instead, turn to less effective programs, the WWC is then actively producing change toward less effective programs in schools. I believe RM is an effective instructional program. In our research on RM, we included many more students than English learners. The students with the lowest performance on literacy and preliteracy and exposed to RM made either equal or greater gains than control students (see Gunn, Smolkowski, Biglan, Black, & Blair, 2005, in J. of Special Education). Our poor readers were selected because they performed in the bottom 15% of students in their schools, and although we did not classify students by disability status, 17.1% of our total sample either received special education or Chapter services for reading. When we examined poor-reader status as a moderator, we found that our poor readers in the RM condition made greater gains than control students. The difference



between conditions was either the same or greater than that for students who began the year below grade level but were not designated as poor readers. This appeared to be true regardless of language status. From our research with RM, then, it appears that it is likely an effective approach, yet WWC is currently recommending to educators nationwide that they remove the program from their shelves based on a flawed review.

The WWC RM report appeared to miss many papers on the RM. Granted, not all of them would qualify for this review, but list of references at the end of the review is missing numerous papers. Books such as Engelmann and Carnine's (2011) *Could John Stuart Mill have Saved Our Schools* and Adams and Engelmann's (1996) *Research on Direct Instruction* include many papers not cited in the literature reviewed. As I mentioned, not all of those would meet criteria for evaluation, but the reference list on their page 8 and 9 should have included more papers.

The limited search for papers raises another question. When I spoke with my colleagues at the National Institute for Direct Instruction, I asked if they had been contacted with a request to provide all relevant papers. They did not believe any attempt at contact had been made. For other interventions, however, WWC representatives have contacted evaluators and program authors to ensure that they have all relevant papers before they begin their review. For example, for a recent review of First Step to Success, WWC contacted Hill Walker and other colleagues to ensure they included all published sources. With respect to the RM, it appears that the review began without a complete literature search.

Interestingly, for the review of First Step, the WWC representatives refused to include presentations and IES project reports. One researcher mentioned that he had included a draft paper with his annual report to IES. He also suggested that WWC examine several presentations. They were told that WWC does not accept presentations or annual progress reports for

grants. This, however, directly contradicts Jill Constantine's response to an earlier inquiry that I made. In my prior inquiry, I specifically asked about progress reports and publications and was told that "every study, regardless of prior publication, is subject to the full WWC review." There appears to be some inconsistency with how WWC literature searches are conducted and with what is considered worthy of review.

The WWC procedures also state that the review will not include research that compares a program to itself or to one that is substantially similar. WWC procedures specify that two reviewers trained in WWC standards independently review the study design and that any differences in reviews are reconciled with a senior WWC reviewer. It appears that all of these reviewers appeared to miss the fact that the two papers included in the WC review violate the prohibition against comparing a program to another substantially similar program. I am surprised that with just two papers it was not determined that the review relied on inappropriate comparisons.

The WWC report also excluded key details, especially for the Herrera paper. For example, I could not find in the report any mention that the Stabilized Learning System (SLS) was a supplemental program that added 90 minutes per day of instruction. The WWC report makes it sound like the study compared SLS to RM, which is clearly incorrect. Herrera et al. (1997) explain that SLS "was implemented forty-five minutes per school day as a as a supplemental to the usual and customary curriculum" (p. 78). They go on to say that the comparison condition also received "the usual and customary prescribed curriculum for the school day, Reading Mastery's explicit methodology" (p. 78). Hence the study found that RM did not perform as well as RM plus an extra 45 minutes of daily supplemental instruction. The report leaves out these critical details.

The WWC review apparently allowed only studies in which 50% of students were labeled with learning disabilities. This appears to be an arbitrary

standard that allows the review to include a study of 20 total students and 12 with LD yet exclude a study with 1000 total students and 480 with LD. The second study is clearly superior, offering substantially improved precision and statistical power. Why would WWC exclude such studies from their review? From my experience with other WWC reviews, however, I am not sure that this or similar criteria have been consistently applied.

As you know, LD identification is notoriously problematic. A recent discussion on the Society for the Scientific Study of Reading e-mail server outlined many problems. States, districts, and schools use different definitions and procedures, and those procedures have changed with time and the adoption of RTI. This can lead to considerably different samples of students from paper to paper. Our study by Gunn et al. (2005), reviewed by WWC, included students with learning disabilities, but we generally ignored the label due to the problems just mentioned. We selected students reading below the 15th percentile within each school and labeled them "poor readers." The WWC report does not identify the criteria for LD, and it appears that studies are included if the report used the label "learning disabilities" but not if the study failed to use those terms.

When the WWC encounters effects that are not statistically significant (e.g., WWC RM report, p. 5), they compute effect sizes. The challenge with this procedure is that the statistically nonsignificant test means that the effect is not sufficiently stable to interpret. Effect sizes were often recommended for research reports not to make statements about statistically nonsignificant findings but rather to quantify the clinical value of fairly small effects from large studies. Many large studies might be able to detect tiny effects (e.g.,  $g = .02$ ), and by recommending that authors report such effect sizes, readers could determine if such a small effect was meaningful. The use of effect sizes to quantify nonsignificant, unreliable results extends an unfortunate myth about statistical analysis. The effect mentioned on page 6 of the

RM report can only be substantively important if it is also statistically reliable.

Stanford statistician Helena Kraemer and her colleagues (2006, in Archives of General Psychiatry) discuss the misuse of effect sizes from underpowered pilot studies and the unintended consequences in some detail. Nonsignificant differences imply large confidence bounds on effect sizes. The cited effect size, -0.33, may well have a confidence interval from -1.0 to +0.33, although such information is not reported. Such problematic reporting has led a number of statisticians to call for even higher standards for what counts as statistically significant (e.g.,  $p < .001$  or  $.0001$ ; see Ioannidis, 2005, in PLoS Medicine; Simmons, Nelson, & Simonsohn, 2011, in Psychological Science; Sterne & Smith, 2001, in the British Medical Journal). "Ultimately, a study with low power has very little chance of advancing scientific knowledge" (Fitzmaurice, 2002, p. 290, in Nutrition). In the present case, the WWC is trying to make a definitive statement about the value of a program but using substandard criteria for that decision. The WWC should instead use high standards in its attempt to advance the field of education.

Having talked with representatives the National Institute for Direct Instruction, I learned that they have asked that the WWC remove the report and that their request has been declined. This irresponsible lack of action has led some educators to begin to question a potentially valuable program based on a very poor review. 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. Lewandowsky and colleagues (2012, in Psychological Science in the Public Interest) discuss the tremendous effort required to correct misinformation once it becomes disseminated to the public. Indeed, their review is quite consistent with a principal among Direct Instruction proponents: it takes 5 to 10 times as much instruction to reteach a misrule as it does

to teach a rule correctly to begin with. Citing over a dozen studies, Lewandowsky et al. report that "research using this paradigm has consistently found that retractions rarely, if ever, have the intended effect of eliminating reliance on misinformation, even when people believe, understand, and later remember the retraction" (p. 114). The WWC review of RM has already been reference by ERIC, and educators have started to react negatively toward the program. In time, no retraction or public statement on the WWC web page will undo the damage.

Thank you for considering these issues.

Best,

-Keith

--

Keith Smolkowski, Ph.D.      keiths@ori.org  
Research Scientist      p 541.484.2123  
Oregon Research Institute      f 541.484.1108  
1715 Franklin Boulevard  
Eugene, Oregon, 97403      www.ori.org

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

**From:** Lesnick, Joy <Joy.Lesnick@ed.gov>  
**Sent:** Friday, October 26, 2012 10:43 AM  
**To:** Doug Carnine  
**Cc:** WhatWorks; jstockard@nifdi.org; Keith Smolkowski  
**Subject:** RE: Reading Mastery report by WWC--Comments by Keith Smolkowski

Dear Dr. Carnine,

Thank you for your message. I am aware of the matter and know that the WWC contractor is currently looking into it. You will receive additional information once the Quality Review Team has reviewed the inquiry and all of the concerns described below.

The WWC Help Desk has been created to address questions and concerns from researchers, educators, and members of the public. Their role is to handle questions like yours. The Help Desk keeps me informed of questions that arise; please direct future questions to the Help Desk rather than to my email address.

Sincerely,

Joy Lesnick

-----Original Message-----

From: Doug Carnine [mailto:dcarnine@uoregon.edu]  
Sent: Thursday, October 25, 2012 9:10 PM  
To: Lesnick, Joy  
Cc: WhatWorks; jstockard@nifdi.org; Keith Smolkowski  
Subject: Reading Mastery report by WWC--Comments by Keith Smolkowski

10-10-12

Dear Dr. Lesnick:

I am following up on my 10-10-12 email to you.

Although I have no financial interest in Reading Mastery, I have close ties to the authors and was involved with Direct Instruction for my entire career. This link to Direct Instruction could bias my reading of the WWC reports. Keith Smolkowski provides the perspective of well-regarded researcher who has no ties to Reading Mastery. His letter to WWC appears below; I have cc'ed him on this email. I hope his comments will spur the re-evaluation of the WWC's report on Reading Mastery that I assume is still active on their web site. Sincerely,

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

Dear WWC,

I am writing today about the recent review of Reading Mastery for students with learning disabilities, which I found to be quite problematic. I have some experience with WWC and Reading Mastery (RM). The review of Gunn, Biglan, Smolkowski, and Ary (2000), which met WWC evidence standards, led the WWC to recommend RM for English learners. Indeed, it was one of the top-rated programs. Because of the very different results from the review of RM for students with learning disabilities, I took a closer look.

While talking about this report with my colleagues at the National Institute for Direct Instruction as well as other researchers, I also discovered some additional discrepancies with the WWC procedures. In an effort to improve the process, I have been encouraged by everyone to forward my concerns to you. Below I raise my concerns with the RM review as well as issues that may lead to inconsistent and potentially inadequate reviews of other programs.

The review of RM for students with learning disabilities suffers from serious flaws. In particular, the report relies on two studies.

Cooke et al (2004) compares RM to its very similar sister program, Horizons. Herrera et al. (1997) compares RM to a comparison condition that included RM plus a supplemental program called the Stabilized Learning System (SLS). The review concluded that RM did not perform better than Horizons. This is what one would expect given that they are highly aligned and designed by the same people. The second paper is not a valid comparison, as it compares RM to RM plus SLS and erroneously concludes that RM may have potentially negative effects.

If RM turns out to be an effective way to teach children with learning disabilities and if schools, instead, turn to less effective programs, the WWC is then actively producing change toward less effective programs in schools.

I believe RM is an effective instructional program. In our research on RM, we included many more students than English learners. The students with the lowest performance on literacy and preliteracy and exposed to RM made either equal or greater gains than control students (see Gunn, Smolkowski, Biglan, Black, & Blair, 2005, in *J. of Special Education*). Our poor readers were selected because they performed in the bottom 15% of students in their schools, and although we did not classify students by disability status, 17.1% of our total sample either received special education or Chapter services for reading. When we examined poor-reader status as a moderator, we found that our poor readers in the RM condition made greater gains than control students. The difference between conditions was either the same or greater than that for students who began the year below grade level but were not designated as poor readers. This appeared to be true regardless of language status. From our research with RM, then, it appears that it is likely an effective approach, yet WWC is currently recommending to educators nationwide that they remove the program from their shelves based on a flawed review.

The WWC RM report appeared to miss many papers on the RM. Granted, not all of them would qualify for this review, but list of references at the end of the review is missing numerous papers.

Books such as Engelmann and Carnine's (2011) *Could John Stuart Mill have Saved Our Schools* and Adams and Engelmann's (1996) *Research on Direct Instruction* include many papers not cited in the literature reviewed. As I mentioned, not all of those would meet criteria for evaluation, but the reference list on their page 8 and 9 should have included more papers.

The limited search for papers raises another question. When I spoke with my colleagues at the National Institute for Direct Instruction, I asked if they had been contacted with a request to provide all relevant papers. They did not believe any attempt at contact had been made.

For other interventions, however, WWC

representatives have contacted evaluators and program authors to ensure that they have all relevant papers before they begin their review.

For example, for a recent review of First Step to Success, WWC contacted Hill Walker and other colleagues to ensure they included all published sources. With respect to the RM, it appears that the review began without a complete literature search.

Interestingly, for the review of First Step, the WWC representatives refused to include presentations and IES project reports. One researcher mentioned that he had included a draft paper with his annual report to IES. He also suggested that WWC examine several presentations.

They were told that WWC does not accept presentations or annual progress reports for grants. This, however, directly contradicts Jill Constantine's response to an earlier inquiry that I made. In my prior inquiry, I specifically asked about progress reports and publications and was told that "every study, regardless of prior publication, is subject to the full WWC review."

There appears to be some inconsistency with how WWC literature searches are conducted and with what is considered worthy of review.

The WWC procedures also state that the review will not include research that compares a program to itself or to one that is substantially similar. WWC procedures specify that two reviewers trained in WWC standards independently review the study design and that any differences in reviews are reconciled with a senior WWC reviewer. It appears that all of these reviewers appeared to miss the fact that the two papers included in the WC review violate the prohibition against comparing a program to another substantially similar program. I am surprised that with just two papers it was not determined that the review relied on inappropriate comparisons.

The WWC report also excluded key details, especially for the Herrera paper. For example, I could not find in the report any mention that the Stabilized Learning System (SLS) was a supplemental program that added 90 minutes per day of instruction. The WWC report makes it sound like the study compared SLS to RM, which is clearly incorrect. Herrera et al. (1997) explain that SLS "was implemented forty-five minutes per school day as a supplemental to the usual and customary curriculum" (p. 78). They go on to say that the comparison condition also received "the usual and customary prescribed curriculum for the school day, Reading Mastery's explicit methodology" (p. 78). Hence the study found that RM did not perform as well as RM plus an extra 45 minutes of daily supplemental instruction. The report leaves out these critical details.

The WWC review apparently allowed only studies in which 50% of students were labeled with learning disabilities. This appears to be an arbitrary standard that allows the review to include a study of 20 total students and 12 with LD yet exclude a study with 1000 total students and 480 with LD. The second study is clearly superior, offering substantially improved precision and statistical power. Why would WWC exclude such studies from their review? From my experience with other WWC reviews, however, I am not sure that this or similar criteria have been consistently applied.

As you know, LD identification is notoriously problematic. A recent discussion on the Society for the Scientific Study of Reading e-mail server outlined many problems. States, districts, and schools use different definitions and procedures, and those procedures have changed with time and the adoption of



RTI. This can lead to considerably different samples of students from paper to paper. Our study by Gunn et al. (2005), reviewed by WWC, included students with learning disabilities, but we generally ignored the label due to the problems just mentioned. We selected students reading below the 15th percentile within each school and labeled them "poor readers." The WWC report does not identify the criteria for LD, and it appears that studies are included if the report used the label "learning disabilities" but not if the study failed to use those terms.

When the WWC encounters effects that are not statistically significant (e.g., WWC RM report, p. 5), they compute effect sizes. The challenge with this procedure is that the statistically nonsignificant test means that the effect is not sufficiently stable to interpret. Effect sizes were often recommended for research reports not to make statements about statistically nonsignificant findings but rather to quantify the clinical value of fairly small effects from large studies. Many large studies might be able to detect tiny effects (e.g.,  $g = .02$ ), and by recommending that authors report such effect sizes, readers could determine if such a small effect was meaningful. The use of effect sizes to quantify nonsignificant, unreliable results extends an unfortunate myth about statistical analysis. The effect mentioned on page 6 of the RM report can only be substantively important if it is also statistically reliable.

Stanford statistician Helena Kraemer and her colleagues (2006, in Archives of General Psychiatry) discuss the misuse of effect sizes from underpowered pilot studies and the unintended consequences in some detail.

Nonsignificant differences imply large confidence bounds on effect sizes. The cited effect size,  $-0.33$ , may well have a confidence interval from  $-1.0$  to  $+0.33$ , although such information is not reported. Such problematic reporting has led a number of statisticians to call for even higher standards for what counts as statistically significant (e.g.,  $p < .001$  or  $.0001$ ; see Ioannidis, 2005, in PLoS Medicine; Simmons, Nelson, & Simonsohn, 2011, in Psychological Science; Sterne & Smith, 2001, in the British Medical Journal). "Ultimately, a study with low power has very little chance of advancing scientific knowledge" (Fitzmaurice, 2002, p. 290, in Nutrition). In the present case, the WWC is trying to make a definitive statement about the value of a program but using substandard criteria for that decision. The WWC should instead use high standards in its attempt to advance the field of education.

Having talked with representatives the National Institute for Direct Instruction, I learned that they have asked that the WWC remove the report and that their request has been declined. This irresponsible lack of action has led some educators to begin to question a potentially valuable program based on a very poor review. 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. Lewandowsky and colleagues (2012, in Psychological Science in the Public Interest) discuss the tremendous effort required to correct misinformation once it becomes disseminated to the public. Indeed, their review is quite consistent with a principal among Direct Instruction proponents: it takes 5 to 10 times as much instruction to reteach a misrule as it does to teach a rule correctly to begin with. Citing over a dozen studies, Lewandowsky et al. report that "research using this paradigm has consistently found that retractions rarely, if ever, have the intended effect of eliminating reliance on misinformation, even when people believe, understand, and later remember the retraction" (p. 114). The WWC review of RM has already been reference by ERIC, and educators have started to react negatively toward the program. In time, no retraction or public statement on the WWC web page will undo the damage.

Thank you for considering these issues.

Best,

-Keith

--

Keith Smolkowski, Ph.D.      keiths@ori.org  
Research Scientist      p 541.484.2123  
Oregon Research Institute      f 541.484.1108  
1715 Franklin Boulevard  
Eugene, Oregon, 97403      www.ori.org

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

From: keiths@ori.org  
Sent: Sunday, January 30, 2011 5:22 PM  
To: info@whatworks.ed.gov  
Subject: IES Website: Contact Us: Evidence Standards, Reference ID  
Number:  
1583325209

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

From: keiths@ori.org

Message: Hi. I recently read the WWC report on Pre-K Mathematics, and it raised some concern about the WWC evidence standards. In particular, I suggest that WWC either restrict its use of unpublished reports or provide additional scrutiny.

WWC report on Pre-K Mathematics:

[http://ies.ed.gov/ncee/wwc/reports/early\\_ed/prek\\_math/](http://ies.ed.gov/ncee/wwc/reports/early_ed/prek_math/)

WWC uses unpublished progress reports and presentations to demonstrate evidence in support of programs, which have not been scrutinized by the peer review process nor are unavailable for review by interested researchers. While the peer review process is not perfect, and I have more to say about that below, it serves as a useful primary screener for results, as WWC reviewers can't be expected to capture every issue. The use of literature that has not been reviewed may lead to tenuous conclusions and incorrect reports by the WWC.

In the present case, with Pre-K Mathematics (link below), the WWC includes a reference to an unpublished progress report as "met" and a presentation at the IES conference as "met with reservations." I was unable to find either of these documents. The project that produced the progress reports, however, and perhaps the presentation, appears to include a design flaw, which appears in the investigators' 2008 publication in JREE (Klein, Starkey, Clements, Sarama, & Iyer, 2008, volume 1, issue 3). This paper is cited as a publication from the project that produced the report that met WWC standards.

The JREE paper states that "a total of 40 preschool classroom . . . were randomly assigned" (p. 160). The analyses nested observations within

children and children within classrooms, which appears to be appropriate. The second paragraph in the design discussion, however, also notes that "intervention and control classrooms were located in different buildings" (p. 160). This revelation, then, makes the design a building-randomized trial, not a classroom-randomized trial, and requires that subsequent analyses nest children within buildings, not classrooms. An analysis of subgroups, classrooms in this case, provides an unspecified and inflated Type I error rates (Murray, Hannan, & Baker, 1996).

In this case, the effect sizes and F-values are quite large and may still provide statistically significant effects, but we cannot determine the level of significance. The results depend on the proportion of variation at the building level. If teachers within buildings worked synergistically on the implementation of the Pre-K Mathematics program, then the ICCs may have been high for building level and could potentially produce nonsignificant findings. Furthermore, the test would need to specify degrees of freedom based on the (unspecified) number of buildings in the study. Actually, had Klein and colleagues randomly assigned classrooms, the JREE paper still incorrectly bases the degrees of freedom for their tests of condition on individual students (e.g.,  $df = 270$ , p. 168) rather than classrooms (Feng, Diehr, Peterson, & McLerran, 2001). But classrooms should not have been the unit of analysis.

As the progress report used by WWC is not publicly available, one can only assume that the design-analysis mismatch evident in the JREE manuscript was repeated in the progress report. This raises questions, however, about the review process and the use of unpublished and unavailable documents and presentations as evidence for WWC. Granted, in this case, JREE reviewers also missed the design-analysis flaw, and they possibly did so because Klein and colleagues misrepresented their design when they stated that classrooms rather than buildings were randomly assigned to condition.

In general, it seems problematic to publish strong conclusions about the efficacy or effectiveness of programs based on unpublished reports and presentations without providing additional scrutiny and making them public. This experience has unfortunately reduced my confidence in all WWC reviews,

which I had previously assumed to be of a higher standard.

Best,

Keith Smolkowski, Ph.D.  
Research Scientist  
Oregon Research Institute  
1715 Franklin Boulevard  
Eugene, Oregon, 97403

**From:** WhatWorks  
**Sent:** Monday, January 31, 2011 10:25 AM  
**To:** 'keiths@ori.org'  
**Subject:** What Works Clearinghouse (WWC 2762)

Dear Dr. Smolkowski,

Thank you for contacting the What Works Clearinghouse (WWC). We have received your email below. WWC staff are reviewing your request and will prepare a response.

#### What Works Clearinghouse

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

#### -----Original Message-----

From: keiths@ori.org [mailto:keiths@ori.org]  
Sent: Sunday, January 30, 2011 5:22 PM  
To: info@whatworks.ed.gov  
Subject: IES Website: Contact Us: Evidence Standards, Reference ID Number: 1583325209

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

From: keiths@ori.org

Message: Hi. I recently read the WWC report on Pre-K Mathematics, and it raised some concern about the WWC evidence standards. In particular, I suggest that WWC either restrict its use of unpublished reports or provide additional scrutiny.

WWC report on Pre-K Mathematics:  
[http://ies.ed.gov/ncee/wwc/reports/early\\_ed/prek\\_math/](http://ies.ed.gov/ncee/wwc/reports/early_ed/prek_math/)

WWC uses unpublished progress reports and presentations to demonstrate evidence in support of programs, which have not been scrutinized by the peer review process nor are unavailable for review by interested researchers. While the peer review process is not perfect, and I have more to say about that below, it serves as a useful primary screener for results, as WWC reviewers can't be expected to capture every issue. The use of literature that has not been reviewed may lead to tenuous conclusions and incorrect reports by the WWC.

In the present case, with Pre-K Mathematics (link below), the WWC includes a reference to an unpublished progress report as "met" and a presentation at the IES conference as "met with reservations." I was unable to find either of these documents. The project that produced the progress reports, however, and perhaps the presentation, appears to include a design flaw, which appears in the investigators' 2008 publication in JREE (Klein, Starkey, Clements, Sarama, & Iyer, 2008, volume 1, issue 3). This paper is cited as a publication from the project that produced the report that met WWC standards.

The JREE paper states that “a total of 40 preschool classroom . . . were randomly assigned” (p. 160). The analyses nested observations within children and children within classrooms, which appears to be appropriate. The second paragraph in the design discussion, however, also notes that “intervention and control classrooms were located in different buildings” (p. 160). This revelation, then, makes the design a building-randomized trial, not a classroom-randomized trial, and requires that subsequent analyses nest children within buildings, not classrooms. An analysis of subgroups, classrooms in this case, provides an unspecified and inflated Type I error rates (Murray, Hannan, & Baker, 1996).

In this case, the effect sizes and F-values are quite large and may still provide statistically significant effects, but we cannot determine the level of significance. The results depend on the proportion of variation at the building level. If teachers within buildings worked synergistically on the implementation of the Pre-K Mathematics program, then the ICCs may have been high for building level and could potentially produce nonsignificant findings. Furthermore, the test would need to specify degrees of freedom based on the (unspecified) number of buildings in the study. Actually, had Klein and colleagues randomly assigned classrooms, the JREE paper still incorrectly bases the degrees of freedom for their tests of condition on individual students (e.g.,  $df = 270$ , p. 168) rather than classrooms (Feng, Diehr, Peterson, & McLerran, 2001). But classrooms should not have been the unit of analysis.

As the progress report used by WWC is not publicly available, one can only assume that the design-analysis mismatch evident in the JREE manuscript was repeated in the progress report. This raises questions, however, about the review process and the use of unpublished and unavailable documents and presentations as evidence for WWC. Granted, in this case, JREE reviewers also missed the design-analysis flaw, and they possibly did so because Klein and colleagues misrepresented their design when they stated that classrooms rather than buildings were randomly assigned to condition.

In general, it seems problematic to publish strong conclusions about the efficacy or effectiveness of programs based on unpublished reports and presentations without providing additional scrutiny and making them public. This experience has unfortunately reduced my confidence in all WWC reviews, which I had previously assumed to be of a higher standard.

Best,

Keith Smolkowski, Ph.D.  
Research Scientist  
Oregon Research Institute  
1715 Franklin Boulevard  
Eugene, Oregon, 97403

**From:** WhatWorks  
**Sent:** Wednesday, February 09, 2011 4:13 PM  
**To:** 'keiths@ori.org'  
**Subject:** RE: What Works Clearinghouse (WWC 2762)

Dear Dr. Smolkowski,

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

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

Thank you,

What Works Clearinghouse

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

---

**From:** WhatWorks  
**Sent:** Monday, January 31, 2011 9:25 AM  
**To:** 'keiths@ori.org'  
**Subject:** What Works Clearinghouse (WWC 2762)

Dear Dr. Smolkowski,

Thank you for contacting the What Works Clearinghouse (WWC). We have received your email below. WWC staff are reviewing your request and will prepare a response.

What Works Clearinghouse

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

-----Original Message-----



From: keiths@ori.org [mailto:keiths@ori.org]

Sent: Sunday, January 30, 2011 5:22 PM

To: info@whatworks.ed.gov

Subject: IES Website: Contact Us: Evidence Standards, Reference ID Number: 1583325209

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

From: keiths@ori.org

Message: Hi. I recently read the WWC report on Pre-K Mathematics, and it raised some concern about the WWC evidence standards. In particular, I suggest that WWC either restrict its use of unpublished reports or provide additional scrutiny.

WWC report on Pre-K Mathematics:

[http://ies.ed.gov/ncee/wwc/reports/early\\_ed/prek\\_math/](http://ies.ed.gov/ncee/wwc/reports/early_ed/prek_math/)

WWC uses unpublished progress reports and presentations to demonstrate evidence in support of programs, which have not been scrutinized by the peer review process nor are unavailable for review by interested researchers. While the peer review process is not perfect, and I have more to say about that below, it serves as a useful primary screener for results, as WWC reviewers can't be expected to capture every issue. The use of literature that has not been reviewed may lead to tenuous conclusions and incorrect reports by the WWC.

In the present case, with Pre-K Mathematics (link below), the WWC includes a reference to an unpublished progress report as "met" and a presentation at the IES conference as "met with reservations." I was unable to find either of these documents. The project that produced the progress reports, however, and perhaps the presentation, appears to include a design flaw, which appears in the investigators' 2008 publication in JREE (Klein, Starkey, Clements, Sarama, & Iyer, 2008, volume 1, issue 3). This paper is cited as a publication from the project that produced the report that met WWC standards.

The JREE paper states that "a total of 40 preschool classroom . . . were randomly assigned" (p. 160). The analyses nested observations within children and children within classrooms, which appears to be appropriate. The second paragraph in the design discussion, however, also notes that "intervention and control classrooms were located in different buildings" (p. 160). This revelation, then, makes the design a building-randomized trial, not a classroom-randomized trial, and requires that subsequent analyses nest children within buildings, not classrooms. An analysis of subgroups, classrooms in this case, provides an unspecified and inflated Type I error rates (Murray, Hannan, & Baker, 1996).

In this case, the effect sizes and F-values are quite large and may still provide statistically significant effects, but we cannot determine the level of significance. The results depend on the proportion of variation at the building level. If teachers within buildings worked synergistically on the implementation of the Pre-K Mathematics program, then the ICCs may have been high for building level and could potentially produce nonsignificant findings. Furthermore, the test would need to specify degrees of freedom based on the (unspecified) number of buildings in the study. Actually, had Klein and colleagues randomly assigned classrooms, the JREE paper still incorrectly bases the degrees of freedom for their

tests of condition on individual students (e.g.,  $df = 270$ ,  $p = .168$ ) rather than classrooms (Feng, Diehr, Peterson, & McLerran, 2001). But classrooms should not have been the unit of analysis.

As the progress report used by WWC is not publicly available, one can only assume that the design-analysis mismatch evident in the JREE manuscript was repeated in the progress report. This raises questions, however, about the review process and the use of unpublished and unavailable documents and presentations as evidence for WWC. Granted, in this case, JREE reviewers also missed the design-analysis flaw, and they possibly did so because Klein and colleagues misrepresented their design when they stated that classrooms rather than buildings were randomly assigned to condition.

In general, it seems problematic to publish strong conclusions about the efficacy or effectiveness of programs based on unpublished reports and presentations without providing additional scrutiny and making them public. This experience has unfortunately reduced my confidence in all WWC reviews, which I had previously assumed to be of a higher standard.

Best,

Keith Smolkowski, Ph.D.  
Research Scientist  
Oregon Research Institute  
1715 Franklin Boulevard  
Eugene, Oregon, 97403

**From:** What Works  
**Sent:** Thursday, September 29, 2011 4:33 PM  
**To:** 'keiths@ori.org'  
**Subject:** What Works Clearinghouse (WWC 2011001)  
**Attachments:** QRT 2011001.pdf

Dear Dr. Smolkowski,

Attached is a response to the questions you raised in your January 30 message to the What Works Clearinghouse (WWC).

Thank you,

What Works Clearinghouse

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

---

**From:** WhatWorks  
**Sent:** Wednesday, February 09, 2011 4:13 PM  
**To:** 'keiths@ori.org'  
**Subject:** RE: What Works Clearinghouse (WWC 2762)

Dear Dr. Smolkowski,

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

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

Thank you,

What Works Clearinghouse

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

---

**From:** WhatWorks  
**Sent:** Monday, January 31, 2011 9:25 AM  
**To:** 'keiths@ori.org'  
**Subject:** What Works Clearinghouse (WWC 2762)

Dear Dr. Smolkowski,

Thank you for contacting the What Works Clearinghouse (WWC). We have received your email below. WWC staff are reviewing your request and will prepare a response.

#### What Works Clearinghouse

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

-----Original Message-----

From: keiths@ori.org [mailto:keiths@ori.org]  
Sent: Sunday, January 30, 2011 5:22 PM  
To: info@whatworks.ed.gov  
Subject: IES Website: Contact Us: Evidence Standards, Reference ID Number: 1583325209

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

From: keiths@ori.org

Message: Hi. I recently read the WWC report on Pre-K Mathematics, and it raised some concern about the WWC evidence standards. In particular, I suggest that WWC either restrict its use of unpublished reports or provide additional scrutiny.

WWC report on Pre-K Mathematics:  
[http://ies.ed.gov/ncee/wwc/reports/early\\_ed/prek\\_math/](http://ies.ed.gov/ncee/wwc/reports/early_ed/prek_math/)

WWC uses unpublished progress reports and presentations to demonstrate evidence in support of programs, which have not been scrutinized by the peer review process nor are unavailable for review by interested researchers. While the peer review process is not perfect, and I have more to say about that below, it serves as a useful primary screener for results, as WWC reviewers can't be expected to capture every issue. The use of literature that has not been reviewed may lead to tenuous conclusions and incorrect reports by the WWC.

In the present case, with Pre-K Mathematics (link below), the WWC includes a reference to an unpublished progress report as "met" and a presentation at the IES conference as "met with reservations." I was unable to find either of these documents. The project that produced the progress reports, however, and perhaps the presentation, appears to include a design flaw, which appears in the investigators' 2008 publication in JREE (Klein, Starkey, Clements, Sarama,

& Iyer, 2008, volume 1, issue 3). This paper is cited as a publication from the project that produced the report that met WWC standards.

The JREE paper states that “a total of 40 preschool classroom . . . were randomly assigned” (p. 160). The analyses nested observations within children and children within classrooms, which appears to be appropriate. The second paragraph in the design discussion, however, also notes that “intervention and control classrooms were located in different buildings” (p. 160). This revelation, then, makes the design a building-randomized trial, not a classroom-randomized trial, and requires that subsequent analyses nest children within buildings, not classrooms. An analysis of subgroups, classrooms in this case, provides an unspecified and inflated Type I error rates (Murray, Hannan, & Baker, 1996).

In this case, the effect sizes and F-values are quite large and may still provide statistically significant effects, but we cannot determine the level of significance. The results depend on the proportion of variation at the building level. If teachers within buildings worked synergistically on the implementation of the Pre-K Mathematics program, then the ICCs may have been high for building level and could potentially produce nonsignificant findings. Furthermore, the test would need to specify degrees of freedom based on the (unspecified) number of buildings in the study. Actually, had Klein and colleagues randomly assigned classrooms, the JREE paper still incorrectly bases the degrees of freedom for their tests of condition on individual students (e.g.,  $df = 270$ , p. 168) rather than classrooms (Feng, Diehr, Peterson, & McLerran, 2001). But classrooms should not have been the unit of analysis.

As the progress report used by WWC is not publicly available, one can only assume that the design-analysis mismatch evident in the JREE manuscript was repeated in the progress report. This raises questions, however, about the review process and the use of unpublished and unavailable documents and presentations as evidence for WWC. Granted, in this case, JREE reviewers also missed the design-analysis flaw, and they possibly did so because Klein and colleagues misrepresented their design when they stated that classrooms rather than buildings were randomly assigned to condition.

In general, it seems problematic to publish strong conclusions about the efficacy or effectiveness of programs based on unpublished reports and presentations without providing additional scrutiny and making them public. This experience has unfortunately reduced my confidence in all WWC reviews, which I had previously assumed to be of a higher standard.

Best,

Keith Smolkowski, Ph.D.  
Research Scientist  
Oregon Research Institute  
1715 Franklin Boulevard  
Eugene, Oregon, 97403



# What Works Clearinghouse **WWC**

A central and trusted source of scientific evidence for what works in education.

September 29, 2011

Keith Smolkowski, Ph.D.  
Research Scientist  
Oregon Research Institute  
1715 Franklin Boulevard  
Eugene, Oregon, 97403  
[keiths@ori.org](mailto:keiths@ori.org)

Reference: QR2011001

Dear Dr. Smolkowski:

Thank you for your inquiry concerning the WWC intervention report on Pre-K Mathematics. In response to your email, we conducted an independent quality review. The WWC quality review team responds to concerns raised by study authors, curriculum developers, or other relevant parties about WWC reviews published on our website. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. These quality reviews are one of tools used to ensure that the standards established by the Institute of Education Sciences (IES) are upheld on every review conducted by the WWC.

Your email raised the issue of a design flaw related to the randomization approach in the Starkey and Klein (2005) report. The quality review team examined the study and the WWC records and determined that the randomization strategy was considered carefully in the initial WWC review. In particular, the study authors were asked to provide details on the study design. The authors provided the following information about the randomization approach:

*A total of 40 preschool classrooms... were randomly assigned in the fall of the first year of implementation using block randomization to either the treatment condition (curricular intervention) or the control condition (no curricular intervention)... A sub-sample of eight focal children was randomly selected in each classroom from the total number of consented children, balanced for age and gender. A total of 40 classrooms (37 schools) took part in the Cohort 1 study, and a total of 40 classrooms (36 schools) took part in the Cohort 2 study.*

Based on this information from the authors, the WWC reviewers concluded that the classroom was the unit of randomization.

The quality review team agreed with your assessment that the more recent publication on the same study (Klein et al. 2008) suggests some ambiguity about the unit of randomization. As you noted, the fact that the treatment and control classrooms were in different buildings suggests that the building may have been the unit of randomization. In light of the ambiguity in the more recent publication, the quality review team conducted additional calculations and determined that even if the unit of randomization was the building and not the classroom, it would not change the current WWC rating of the study. In assessing studies in which the unit of randomization is not the same as

the unit of analysis, the WWC applies a clustering correction to the calculation of statistical significance. In the WWC review of Starkey and Klein (2005), no correction for clustering was applied because the study authors used hierarchical linear modeling (HLM) to analyze data and, in so doing, accounted for clustering of children in classrooms (see appendix table A.3, note 9, in the intervention report). The difference between clustering units of buildings ( $n=37$ ) and classrooms ( $n=40$ ) is not large enough to invalidate statistically significant findings presented in the intervention report. More specifically, it would take 10 buildings (and not 37) to render the effect size of 0.58 non-significant ( $p>0.05$ ). More information on the clustering correction can be found in the WWC Procedures and Standards Handbook, Appendix C (available at <http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&toCId=9>).

Based on their analysis, the quality review team concluded that the intervention report on Pre-K Mathematics does not require revision. However, they recommend that if the intervention report is updated, Klein et al. (2008) should be included in the review and an additional author query should be conducted to address the ambiguity regarding the unit of randomization and, if needed, appropriate cluster corrections should be applied to the estimates.

You also raised the issue of whether the WWC should review unpublished reports. This issue is beyond the scope of the quality review team, whose role is to determine whether WWC standards and procedures were upheld in the review. The established procedures call for an “extensive search of published and unpublished research literature” (WWC Procedures and Standards Handbook, Section II.C, available at <http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&toCId=3#litsearch>).

As Director of the WWC, I would like to explain why we include unpublished reports and how we ensure all studies receive a rigorous, systematic review. The WWC does not require that the study be published because we seek to represent the full body of research evidence publicly available at the time of the WWC review. Some of the most current, high-quality research may be in the publication process at the time of our review, or may be disseminated in other ways. In addition, peer-reviewed publications may be more prone to publication bias such that significant impact findings are more likely to be published than are studies that find no evidence of impact (see Lipsey and Wilson, 2001, “Practical Meta-analysis,” in *Applied Social Research Methods Series*, Volume 49, SAGE Publications: Thousand Oaks). If the WWC relied only on peer-reviewed publications, we would not be conducting a systematic review of all of the research available, and our findings may be more subject to publication bias.

# What Works Clearinghouse **WWC**

A central and trusted source of scientific evidence for what works in education.

The WWC has designed a review process to systematically assess the quality of a study using evidence standards. Every study, regardless of prior publication, is subject to the full WWC review. We have found that the publication of a report, even by a respected peer-reviewed journal, does not ensure that the study will meet WWC evidence standards. Two reviewers who have been trained in WWC standards independently review the study design, outcomes, samples and attrition, and analysis methods. After the reviews are complete, they hold a reconciliation meeting with a senior WWC reviewer to discuss any differences between their reviews and any remaining issues about the study. Remaining issues that are technical in nature are addressed by a statistical or content expert. Questions about the study are posed to the study authors. Further information about the review process and standards can be found in the WWC Procedures and Standards Handbook.

I hope that this letter has addressed your concerns. If you have other concerns, please do not hesitate to contact the WWC through [info@whatworks.ed.gov](mailto:info@whatworks.ed.gov).

Sincerely,

(b)(6)

Jill Constantine  
Director, What Works Clearinghouse

cc: (b)(6)



**From:** What Works  
**Sent:** Friday, December 21, 2012 10:01 AM  
**To:** 'keiths@ori.org'  
**Subject:** What Works Clearinghouse (QR2012014)  
**Attachments:** QRT 2012014.pdf

Dear Dr. Smolkowski,

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

Thank you,

What Works Clearinghouse

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

December 21, 2012

Dr. Keith Smolkowski  
Oregon Research Institute  
1715 Franklin Blvd.  
Eugene, OR 97403

Reference Number: QR2012014

Dear Dr. Smolkowski,

Thank you for your inquiry concerning the WWC intervention report on *Reading Mastery* for students with learning disabilities. In response to your email, we conducted an independent quality review. The WWC Quality Review Team responds to concerns raised by study authors, curriculum developers, or other relevant parties about WWC reviews published on our website. These quality reviews are undertaken when concerned parties present evidence that a WWC review may be inaccurate. When a quality review is conducted, a researcher who was not involved in the initial review undertakes an independent assessment of the study in question. The researcher also investigates the procedures used and decisions made during the original review of the study. These quality reviews are one of the tools used to ensure that the standards established by the Institute of Education Sciences (IES) are upheld on every review conducted by the WWC.

Regarding the study by Cooke et al. (2004), the quality review concluded that the study is eligible for review as a comparison of *Reading Mastery Fast Cycle* with *Horizons Fast Track*. As you noted in your email, the two curricula have some features in common, and the study examines whether the differences between the curricula had different effects on student achievement. The quality review noted that there are differences between the two curricula (as described on page 141 of the study) and that the WWC intervention report properly characterizes the study as “conducted to examine whether the limited substantive differences between the interventions led to different effects” (page 4 of the WWC intervention report). However, the quality review also noted that the intervention report text does not sufficiently highlight that the intervention report is summarizing the effectiveness of *Reading Mastery* relative to another Direct Instruction curriculum with shared features.

Regarding the study by Herrera et al. (1997), the quality review concluded that the study is ineligible for review because both comparison groups may have been taught using *Reading Mastery*. The quality review noted that the study does not explicitly state that both groups received *Reading Mastery* instruction, but this could be inferred from statements regarding the “usual and customary” curriculum for the school day (on page 78 of the study). The quality

review noted that, during the original review process, the WWC queried the study authors regarding this issue but did not receive a response. The original review team interpreted the study to imply that the implicit method replaced the “explicit” method, which the authors describe as Reading Mastery. For example, although the study notes that the implicit method was supplemental, it was integrated without adding time to the school day, so some activity—the team assumed Reading Mastery—would have been replaced. However, the quality review concluded that both comparison groups might have received instruction using Reading Mastery.

You raised a concern that the intervention report does not mention several studies of *Reading Mastery*. The quality review determined that the literature search process was conducted as described in the *Evidence Review Protocol for K–12 Students with Learning Disabilities* (available at <http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=31>). In particular, the WWC contacted McGraw-Hill, the publisher of *Reading Mastery*, to request studies. Studies of *Reading Mastery* that do not fit the eligibility criteria described in the protocol may be screened out of the review process at an early stage and, in that event, would not be included in the reference list. If you are aware of studies that fit the eligibility criteria described in the protocol and are not mentioned in the reference list, please send the reference information to the WWC. The two books that you mentioned in your email, and the references cited therein, will be examined during the search process when the intervention report is updated.

In light of the findings of the quality review, the WWC intervention report on *Reading Mastery* for students with learning disabilities will be revised. The results from the Herrera et al. (1997) study will not be included in the evidence rating, and the study will be listed as ineligible for review, as it is not a test of *Reading Mastery*. The research summary and the description of the remaining study that meets standards will be revised to clearly note the comparison condition for the Cooke et al. (2004) study. Because the revisions affect the ratings, the intervention report will be removed from the WWC website until the revisions are complete, and the revised report will be posted.

In addition to the issues addressed by the quality review, you expressed some concerns with WWC processes. The quality review team considers whether WWC procedures were followed, but does not respond to questions about the procedures themselves. However, I briefly address those concerns here.

- You expressed concern regarding the identification of learning disabilities in students. As noted in review protocol, there is variation in procedures used to diagnose learning disabilities. To address this diversity, the WWC reviewed studies in which the authors claim that the study participants have an identified learning disability based on federal, state, or school district criteria. Studies that report findings for a sample or a subgroup that is at least 50 percent students with learning disabilities are eligible for review.



# What Works Clearinghouse WWC

A central and trusted source of scientific evidence for what works in education.

- You noted that, in a previous letter to you, I explained that the WWC reviews studies regardless of whether they have been published. This continues to be the policy of the WWC. However, the WWC requires that the studies be publicly available; for example, posted to a website as a draft paper. Publicly available documents that accompany a presentation or a grant report would be eligible for review, since they are evidence being presented to the public about the effectiveness of the intervention. For reports submitted as part of a grant, the WWC attempts to confirm with the authors whether the research is publicly available and, if not, whether they would be willing to have it be made available. If a report is not publicly available, then the WWC would not conduct a review on it for public release.
- Finally, you expressed concern with the WWC characterization of “substantively important” effects. WWC reports describe both the magnitude and statistical significance of effects. Interventions with effect sizes of 0.25 or higher are deemed substantively important to indicate that the magnitude of the effect is substantial. We understand that this does not mean the finding is statistically reliable, which is why we use different language to characterize it. We also understand that there are other ways to report on the magnitude and statistical reliability of findings, such as providing point estimates and confidence intervals. The WWC welcomes feedback on standards and reporting, and we will consider your input in future updates of standards and reporting practices.

I hope that this letter has addressed your concerns. If you have other concerns, please do not hesitate to contact the WWC through [info@whatworks.ed.gov](mailto:info@whatworks.ed.gov).

Sincerely,

(b)(6)

Jill Constantine  
Director, What Works Clearinghouse

cc:

(b)(6)