

WHAT WORKS CLEARINGHOUSE

Reading Recovery: How Do We Rank?

Robert M. Schwartz, Oakland University

In September 2017, I noticed a change in the What Works Clearinghouse (WWC) website.¹ An initiative of the U.S. Department of Education's Institute of Education Sciences (IES), the WWC reviews and assesses research evidence for educational programs, products, practices, and policies. For the first time, they were ranking the 220 literacy interventions they had reviewed in order of their effectiveness. This ranking seems like an attempt to make the WWC research reviews more helpful for school districts and administrators looking for evidence-based practices, as called for in the Every Student Succeeds Act² (ESSA) or the previous No Child Left Behind legislation.

Because an organization like the WWC cannot advocate for any particular program, it isn't apparent in their reviews or rankings how exceptional and strong the evidence is for the effectiveness of Reading Recovery.[®] My goal in this article is to help education decision makers see that Reading Recovery is one of the very few interventions with sufficient evidence of effectiveness to justify implementation. To accomplish this, I detail the WWC criteria for ranking research-based interventions, show how the most recent Reading Recovery research (May, Sirinides, Gray, & Goldsworthy, 2016) would influence these rankings, and provide additional effectiveness evidence that goes beyond these rankings.

When I first noticed the WWC rankings, Success for All[®] was at the top of their list and Reading Recovery was ranked fifth. I was surprised by these rankings since I had recently read and compared (Schwartz, 2016) the independent evaluations of the Investing in Innovation (i3) scale-up grants awarded to these two interventions. I also knew that the WWC was aware of how much stronger the Reading Recovery research evidence was than that for Success for All, since they had published single study reviews of both grants' final reports. The Reading Recovery review³ showed a large effect on standardized measures of word reading skill and comprehension, while the Success for All review⁴ showed a small effect on word reading and no impact on students' comprehension.

Because WWC rankings did not match my reading of the research, I contacted the WWC help desk to ask how

they determined their rankings. They quickly responded to my question, referring me to the *WWC Procedures and Standards Handbook, Version 3.0* (Appendix B).⁵ Unfortunately, these procedures describe how WWC prioritizes interventions for review or updating of previous intervention reports, but not how they rank the effectiveness of interventions they have reviewed. The next day, when I clarified that I was interested in how WWC ranked the effectiveness of the interventions listed under the Literacy topic area, they responded that their staff were considering my question and would prepare a response.

Incorporating the results of the What Works Clearinghouse single study review of the i3 final report would move Reading Recovery to the first spot in WWC ranking of effective interventions.

This took a bit longer. A month later, I followed up on my inquiry. They responded that they were still looking into it and would get back to me as soon as possible. Another month passed. No word. When I checked the WWC website at the end of November, however, the rankings had changed. Reading Recovery was ranked third, and Success for All had moved down the list to tenth.

Since I still did not know the basis for these ranking, I contacted WWC again. Here is their December 1, 2017 email response:

Thank you for following up. The WWC has resolved the issue with how interventions are organized in the Find What Works tool. The list of interventions is sorted by the number of outcome domains with positive effects. If there is a tie between two interventions (that is, the evidence for both interventions demonstrates positive effects in the same number of outcome domains), then the tie is broken

by considering the following factors (in order): 1) The number of domains with positive or potential positive effects; 2) The number of studies meeting WWC design standards; and 3) The number of students in studies meeting WWC design standards.

The WWC Procedures Handbook Version 4.0⁶ describes the process for developing intervention reports. We encourage you to read the handbook to learn more about this process. Periodically, the WWC revisits interventions to examine all new research that has emerged since the report’s initial release. The Institute of Education Sciences (IES) and the WWC select which reports to update based on an annual prioritization process. If an intervention report is selected, the WWC screens and reviews new studies. However, we are unable to review all of the interventions with updated research.

After reviewing any additional studies, the WWC will release an updated intervention report. If some of the new research meets design standards, the summary measures (effect size, improvement index, and rating) may change. Please note that new research on an intervention does not guarantee the WWC will update an intervention report.

Please let us know if you have additional questions about the process of developing and updating intervention reports. (personal communication)

Updating the Evidence

In Reading Recovery’s July 2013 intervention report⁷ we had positive ratings in the Beginning Reading outcome domains of alphabetics and reading achievement, and potentially positive ratings in the two other Beginning Reading domains — comprehension and reading fluency. Figure 1, a partial table from page 25 of the *Procedures Handbook, Version 4.0* shows their criteria for a positive or potentially positive rating. To determine how the Reading Recovery scale-up research (May et al., 2016) would influence our domain ratings and intervention ranking, I applied the WWC criteria. WWC’s single study review found May et al.’s study to meet their design standards without reservations and have a significant positive effect on the Iowa Test of Basic Skills (ITBS) comprehension measure. The significant positive effect size in this domain would increase our rating in comprehension from potentially positive to positive. This one change would move Reading Recovery to the first spot in WWC ranking of effective interventions.

The i3 research (May et al., 2016) has many other elements that would establish Reading Recovery’s top ranking in the WWC list of effective interventions. For example, the report includes a subgroup analysis for the 19% of the sample classified as English Language Learners (ELL). May et al. report large impact estimates for this group on the Observation Survey Total Score, the ITBS Total Score, ITBS Reading Words, and Comprehen-

Figure 1. Partial Table of WWC Effectiveness Criteria

Table IV.3. Criteria Used to Determine the WWC Rating of Effectiveness for an Intervention

| | |
|---------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Positive effects: Strong evidence of a positive effect with no overriding contrary evidence. | <ul style="list-style-type: none"> • Two or more studies show statistically significant positive effects, at least one of which meets WWC group design standards without reservations, AND • No studies show statistically significant or substantively important negative effects. |
| Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence. | <ul style="list-style-type: none"> • At least one study shows statistically significant or substantively important positive effects, AND • Fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects, AND • No studies show statistically significant or substantively important negative effects. |

sion subscales. WWC last reviewed Reading Recovery in the ELL topic area in December 2009. At that time, they reported “no studies of Reading Recovery® were found that fell within the scope of the ELL review protocol and met WWC evidence standards.”⁸

Whether this new information results in positive or potentially positive ratings for alphabets, comprehension, and reading achievement in the ELL topic area depends on issues related to replication. May et al. could be considered as one study or as four replications of the effectiveness findings. This multiyear study reports results on large independent random samples of first-grade students taught by different Reading Recovery teachers from different schools in each of four school years. As indicated in the WWC rating criteria, a positive rating requires two or more studies, which in turn depends on whether each replication is considered a study. Replication of research findings across independent samples is rare in fields like medicine, psychology, or education (Open Science Collaboration, 2015). It should be valued in the WWC ratings (IES, 2016). Additional positive ratings in the ELL topic area would be important because WWC adds the number of positive ratings across topic areas in determining their ranking of an intervention’s effectiveness.

The reading fluency domain has the largest effect size in the Reading Recovery 2013 WWC intervention report. Despite this large effect, the fluency domain is rated as potentially positive because WWC found only one study that met their standard, Schwartz (2005). The fluency

measure they used from this study was the Text Reading Level subtest from *An Observation Survey of Early Literacy Achievement* (Clay, 2013). A number of studies reviewed by WWC presented data on Text Reading Level, but WWC was concerned that this measure might not have the measurement characteristics necessary to calculate an effect size from the usual parametric formula. Schwartz was able to provide the raw score on this measure to WWC so they could do a non-parametric estimate of the effect size. This analysis produced the large effect and potentially positive rating in the Reading Recovery 2013 intervention report.

D’Agostino, Rodgers, and Mauck (2017) address issues that have been problematic in previous research using Clay’s Observation Survey. They provide a Rasch transformation for both the Total Score and the Text Reading Level measure to ensure an equal interval scale. May et al. (2016) used this transformation in their analysis of the Observation Survey Total Score, but they don’t report results on the subscores that WWC previously considered as measures in the alphabets, comprehension, reading achievement, and fluency domains.

Schwartz and Lomax (in preparation) are currently conducting a secondary analysis of these subscores from the i3 data. Table 1 presents the means, standard deviations, and effect size calculations for Text Reading Level raw scores and scale scores across the 4-year i3 study. This table provides the evidence that WWC needs to increase the fluency rating from potentially positive to positive.

Table 1. Effect Size Calculations for Text Reading Level Raw Scores and Scale Scores Across the 4-Year i3 Study

| | 2011-12 | | 2012-13 | | 2013-14 | | 2014-15 | | Pooled | |
|--------------------------------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|
| | Treatment | Control |
| N | 429 | 429 | 725 | 725 | 855 | 855 | 1430 | 1430 | 3439 | 3439 |
| Post-Test | | | | | | | | | | |
| Mean | 10.6 | 5.2 | 10.3 | 5.1 | 10.4 | 5.4 | 10.5 | 5.2 | 10.5 | 5.2 |
| (Standard Deviation) | (4.8) | (3.7) | (4.7) | (4.2) | (4.9) | (3.9) | (4.9) | (4.0) | (4.9) | (4.0) |
| Post-Test Scale Score | | | | | | | | | | |
| Mean | 492.0 | 435.5 | 488.7 | 425.0 | 489.5 | 433.4 | 489.6 | 431.3 | 489.7 | 431.0 |
| (Standard Deviation) | (39.5) | (64.5) | (44.7) | (77.1) | (44.5) | (69.5) | (45.4) | (69.7) | (44.3) | (70.7) |
| Raw Score Effect Size | +1.46 | | +1.24 | | +1.28 | | +1.33 | | +1.33 | |
| Scale Score Effect Size | +0.88 | | +0.83 | | +0.81 | | +0.84 | | +0.83 | |

How Much Evidence is Enough?

In education, as in medicine, there is a strong push to use methods and interventions that have demonstrated effectiveness in clinical trials. Still, evaluating this evidence is complex. The National Cancer Institute provides physicians with levels of evidence related to five different topic areas.⁹ Although the criteria vary across these areas, they usually include some evaluation of the strength of the study design and the outcomes measured. In evaluation of cancer treatments, a measure of mortality reduction is stronger than an indirect measure like tumor response rate. Similarly, in literacy research a measure of text comprehension or general reading achievement should be considered stronger than a measure of phonological awareness that may later lead to reading comprehension gains. When Reading Recovery's intervention report is updated to include the May et al. 2016 study, the evidence will demonstrate positive ratings in all the Beginning Reading domains.

Although WWC uses a count of the number of positive ratings an intervention can demonstrate to rank effectiveness, these ratings and ranking alone are not sufficient to justify implementation decisions. The IES (2016) considered the highest evidence level to be based on an "independent evaluation of a fully-developed education intervention with *prior evidence of efficacy, when implemented by the end user under routine conditions*" (p. 5). This evidence is further strengthened by sufficient diversity in the sample to ensure appropriate generalizability. The May et al. study meets all these scale-up design conditions and replicates the effectiveness findings in each of 4 years using 2 different methodologies.

The findings shown in the May et al. study reflect lessons learned over 40 years of supporting the professional learning of teachers and the implementation efforts of administrators as Reading Recovery has scaled up in five countries. Successfully implementing this early intervention/prevention model for the most at-risk beginning readers requires considerable effort (Briggs & Honchell, 2016). If your school or district is committed to supporting these students, the combination of design elements, replication, effect size estimates, and domain ratings makes Reading Recovery arguably the only early literacy intervention with sufficient scientific evidence to justify adoption decisions.

The combination of design elements, replication, effect size estimates, and domain ratings makes Reading Recovery arguably the only early literacy intervention with sufficient scientific evidence to justify adoption for the most at-risk beginning readers.

Endnotes

(All web addresses as accessed January 5, 2018)

- ¹ Finding What Works based on the evidence
<https://ies.ed.gov/ncee/wwc/FWW/Results?filters=,Literacy>
- ² Every Student Succeeds Act (ESSA)
<https://www.ed.gov/essa?src=rn>
- ³ Review of Reading Recovery: An evaluation of the four-year i3 scale-up
<https://ies.ed.gov/ncee/wwc/Study/32027>
- ⁴ Scaling up the Success for All model of school reform: Final report from the Investing in Innovation (i3) evaluation
<https://ies.ed.gov/ncee/wwc/Study/32024>
- ⁵ *What Works Clearinghouse Procedures and Standards Handbook, Version 3.0*
https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf
- ⁶ *What Works Clearinghouse Procedures Handbook, Version 4.0*
https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf
- ⁷ Summary of Evidence for Reading Recovery, Beginning Reading (July 2013)
<https://ies.ed.gov/ncee/wwc/Intervention/209>
- ⁸ Summary of Evidence for Reading Recovery, English Language Learners (December 2009)
<https://ies.ed.gov/ncee/wwc/Intervention/209#ta-6>
- ⁹ PDQ® Levels of Evidence
<https://www.cancer.gov/publications/pdq/levels-evidence>

References

- Briggs, C., & Honchell, B. A. (2016). Effective implementation: What we can learn from the i3 study. *The Journal of Reading Recovery, 16*(1), 40–46.
- Clay, M. M. (2013). *An observation survey of early literacy achievement* (3rd ed.). Auckland, NZ: Pearson.
- D'Agostino, J. V., Rodgers, E., & Mauck, S. (2017). Addressing inadequacies of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly*. doi:10.1002/rrq.181
- Institute of Education Sciences. (2016). *Building evidence: What comes after an efficacy study?* Retrieved from <https://ies.ed.gov/ncer/whatsnew/techworkinggroup/pdf/BuildingEvidenceTWG.pdf>
- May, H., Sirinides, P., Gray, A., & Goldsworthy, H. (2016). *Reading Recovery: An evaluation of the four-year i3 scale-up*. Retrieved from <http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). doi:10.1126/science.aac4716
- Schwartz, R. M. (2005). Literacy learning of at-risk first grade students in the Reading Recovery early intervention. *Journal of Educational Psychology, 97*(2), 257–267.
- Schwartz, R. M. (2016). Effective early intervention: Lessons from the i3 evaluation of Reading Recovery. *The Journal of Reading Recovery, 16*(1), 47–54.

Schwartz, R. M., & Lomax, R (in preparation). *Secondary analysis of the Reading Recovery four-year i3 scale-up*. To be presented at the Society for the Scientific Study of Reading Conference, Brighton, UK, July 2018.

What Works Clearinghouse. (n.d.). *Handbooks and other resources*. Retrieved from <https://ies.ed.gov/ncee/wwcl/Handbooks>

About the Author

Dr. Robert Schwartz is an emeritus professor in the Department of Reading and Language Arts at Oakland University in Rochester, MI. He is a past president of and current research consultant for the Reading Recovery Council of North America. His research interests include self-monitoring in beginning reading, early literacy intervention, research design, and professional development for literacy teachers. In the What Works Clearinghouse 2007 review of 887 studies from 153 beginning reading programs, Dr. Schwartz's Reading Recovery research was one of only 27 studies that met WWC's standards without reservations. He can be reached at rschwartz@oakland.edu.



**Shop Amazon?
Use Smile and
Help Support
the Council!**

amazonsmile
You shop. Amazon gives.

Visit smile.amazon.com, sign in to your regular Amazon account, and designate RRCNA as the charitable organization you'd like to support. Every time you shop, The Amazon Smile Foundation will donate 0.5% of the purchase price from your eligible AmazonSmile purchases.

Go to www.smile.amazon.com and start shopping!