

What Does Science Say About Orton-Gillingham Interventions? An Explanation and Commentary on the Stevens et al. (2021) Meta-Analysis

by Emily Solari, Yaacov Petscher, and Colby Hall

A recent meta-analysis published in *Exceptional Children* (Stevens et al., 2021) looked at the effects of Orton-Gillingham (OG) reading interventions on reading outcomes for students who have word reading difficulties. The results of the study, which showed no statistically significant effect but a practically important effect size (explained further below), have led to questions and lively conversation among practitioners and reading researchers. One of the things that is important about science is that it is constantly evolving; this is true in education science as much as it is in the health sciences. Because this journal is committed to translating empirical findings from reading research in order to make education science accessible to practitioners, the intent of this commentary is to provide a clear description of the findings reported in this recent meta-analysis, addressing the degree to which they align with those reported in similar reviews of OG interventions. We discuss the degree to which the findings represent an evolution of reading science and their implications for instructional practice, policy, and future research.

What is a Meta-Analysis?

The purpose of a meta-analysis is to systematically combine and analyze data from published research studies to better understand what that body of research says about a particular question. Meta-analyses of intervention research studies combine results from multiple, individual research studies in order to provide a sense of the overall strength of intervention effects. When multiple studies are combined, there is potential to better inform the field—because individual research studies have flaws. Meta-analyses can enable us to find a “signal” (in other words, true information about intervention effects) amid the “noise” (in other words, the random, unwanted fluctuation in study results that reflects study flaws or idiosyncrasies rather than truly reflecting intervention effects). We can place more trust in findings when a larger body of research is studied together, especially when rigorous statistical designs and methods are used to calculate average intervention effects across studies.

That said, as with all scientific approaches and methodologies, there are some limitations to meta-analysis. When conducting

a meta-analysis, research teams must make decisions about which studies to include and which to exclude based on a predetermined set of rules or criteria. For example, they may require studies to use certain types of rigorous research designs (such as experimental or quasi-experimental designs and not case study designs), or to include measures of particular types of reading outcomes (such as measures of word reading outcomes). This means that not all studies that have researched a particular intervention will be included in a meta-analysis. Researchers often begin a meta-analysis intending to answer a new set of questions, ones that the individual studies could not answer. When individual studies are combined, the nuances of the studies are lost in favor of answering these new questions. Meta-analysis also does not answer all the possible questions about how well a specific intervention or instructional approach works.

What Did Stevens et al. Do in Their Meta-Analysis?

The research team examined the effects of interventions tested in 16 studies that met their

pre-established inclusion criteria: meta-analyzed studies were studies of small-group, OG interventions that only targeted foundational reading skills. The inclusion criteria also required that the population of students in each study had word reading difficulties. The majority of the interventions studied by the authors were branded as OG (that is, Alphabetic Phonics, Barton Reading and Spelling System, Dyslexia Training Program, Foundations, Herman Method, Language!, Lindamood Bell, Project Assist, Project Read, Recipe for Reading, Slingerland Approach, the Spalding Method, S.P.I.R.E., Starting Over, Take Flight, Wilson Reading System, The Writing Road to Reading); a smaller number were described as being based on OG principles. The authors used meta-analysis to estimate the average effect of OG interventions across these 16 studies and the extent to which the quality of the research design used in the studies in the year when the studies were published was related to the strength of the effect.

What Did the Authors Report?

The authors found that on average across all the studies included in the meta-analysis, students with word-level reading difficulties who received OG interventions did not make statistically significant improvements in foundational skills, vocabulary, or comprehension outcomes when compared to groups who did not receive OG interventions. Descriptively speaking, students across all studies who received OG interventions did have higher mean scores at posttest than their peers who did not receive the OG interventions. Authors calculated the average size of the difference in effects (between students who received OG and those who did not) and reported that students who received OG interventions had scores that were higher by 0.32 of a standard deviation in foundational reading skills and by 0.14 of a standard deviation for vocabulary and comprehension outcomes.

What Do the Results Mean? Statistical Significance vs. Practical Importance

Although the authors found no statistically significant effect of OG interventions, they did report an effect size of 0.22, a value that many scientists would classify as “small but meaningful.” Because of this seemingly conflicting information, it is important to understand the difference between statistical significance and practical importance. Very simply, statistical significance tells us whether a result from a statistical analysis is due to chance, and practical importance tells us whether that result is meaningful enough to act on in some way (for

Very simply, statistical significance tells us whether a result from a statistical analysis is due to chance, and practical importance tells us whether that result is meaningful enough to act on in some way (for example, recommend an intervention, do more research, etc.).

example, recommend an intervention, do more research, etc.). Statistical significance is influenced by many factors, including the number of subjects who participated in the study, the number of variables in the statistical analysis, and what kind of statistical analysis is being conducted. Practical importance is a way to go beyond statistical significance to say something more about the result.

An average effect size of 0.22 is, practically speaking, a meaningful one. To put this effect size in perspective, Russell Gersten et al. (2020) recently meta-analyzed 33 rigorous studies of reading interventions with students with or at risk for reading difficulties in Grades 1-3 and found a significant positive effect on reading outcomes with a mean effect size of 0.39. Jeanne Wanzek et al. (2018) also determined that extensive reading interventions for students with or at risk for reading difficulties in Grades K-3 produced significant positive effects on reading outcomes, with an average effect size of 0.37. Both studies interpreted effects of this size as representing meaningful improvement in reading outcomes as a result of reading intervention.

Practical importance is a way to go beyond statistical significance to say something more about the result.

The lack of statistically significant findings from the Stevens et al. meta-analysis is generally consistent with a previous systematic review of 12 OG intervention studies conducted by Ritchey and Goeke (2006). The authors descriptively summarized results and reported that “the find-

ings were not, however, all positive in favor of OG instructional programs. Nor were the findings [all] statistically significant” (p. 181) in favor of either OG or the alternative instructional program to which OG was compared. These findings are also consistent with published reports from the What Works Clearinghouse (WWC), an agency within the U.S. Department of Education that independently reviewed several studies of both branded and unbranded OG programs. The WWC found that the evidence in favor of OG programs is limited, either due to a mix of positive and negative effects or, more frequently, because available studies of such programs do not meet WWC quality standards (for example, WWC, 2010a; 2010b; 2010c; 2010d).

What Cannot Be Concluded From These Results?

Stevens et al. (2021) were very cautious about interpreting their findings. They concluded by observing that:

the findings from this meta-analysis do not provide definitive evidence that OG interventions significantly improve the reading outcomes of students with or at risk for WLRD [word-level reading disabilities], such as dyslexia. However, the mean ES of 0.32 indicates OG interventions may hold promise for positively impacting the reading outcomes of this population of students. Additional high-quality research is needed to identify whether OG interventions are or are not effective for students with and at risk for WLRD. (p. 16)

We agree that the scientific evidence presented in this meta-analysis supports this cautious conclusion. Research does not suggest that OG interventions do *not* work. Instead, research findings do not provide firm evidence of effectiveness for OG interventions, although the mean effect size of 0.32 in favor of OG interventions constitutes evidence of promise and suggests the need for future research.

Importantly, it should be noted that there is a strong evidence base for early word reading instruction. In particular, explicit and systematic instruction for students with reading difficulties has a robust research evidence base (for example, Gersten et al., 2008; Swanson, 1999; Vaughn et al., 2012). There are mountains of evidence that reading instruction for students with word-level reading difficulties should include systematic instruction in phonological awareness, grapheme-phoneme correspondences, decoding, and word reading (Foorman et al., 2016; Gersten et al., 2020; Wanzek et al., 2016). Because the OG philosophy is inclusive of

The one component of OG approaches for which there is less research evidence is the kinesthetic/tactile instructional component, often called the “multisensory” component. There is little research to suggest this component adds value to explicit and systematic phonics-based instruction.

explicit and systematic phonics-based instruction, it is plausible that the mean effect size of 0.32 is rooted in the delivery of this kind of instruction that has been shown to work for students with reading difficulties. The one component of OG approaches for which there is less research evidence is the kinesthetic/tactile instructional component, often called the “multisensory” component (Al Otaiba et al., 2018). There is little research to suggest this component adds value to explicit and systematic phonics-based instruction.

What Do We Do Next?

There are important implications and limitations of the Stevens et al. (2021) findings for research. First, we echo the calls for more rigorous, experimental research on OG interventions. The inconclusive evidence reported in the Stevens et al. meta-analysis and in other research reviews is due to a confluence of factors, including (a) relatively few studies that meet inclusion criteria due to inadequate rigor in their experimental designs, (b) smaller sample sizes that can either inflate or underestimate program effects, and (c) a lack of designs that specifically test the value of exposure to multisensory instruction. An important limitation of the Stevens et al. meta-analysis is that it did not analyze differences between studies that used branded or unbranded interventions. Further, a number of the studies that used branded interventions included instructional add-ons that made the programs being evaluated slightly different from the published version of the programs. These limitations could serve to encourage reanalysis of data or proposals of new intervention studies.

The implications for policymaking are also clear. The research evidence does not currently support legislative mandates that schools train teachers in the delivery of or provide students

with branded or unbranded OG instruction (or instruction that employs a multisensory component). There is a need to provide explicit, systematic, evidence-based instruction to students with word-level reading difficulties, as well as to mandate teacher training in the delivery of this type of instruction. Based on the current evidence (including the other meta-analyses cited in this paper that show positive effects of explicit, systematic, and intensive instruction in early word reading), educators can advocate for their schools' adoption of reading intervention programs that have these features. They can refer to the Stevens et al. (2021) meta-analysis when making the point that the intervention programs their schools adopt/use need not be multisensory, or branded/unbranded OG programs.

Finally, it is important for all of us, as researchers, educators, and stakeholders in education, to embrace the way in which science is constantly evolving. It remains to be seen what large-scale rigorous research will determine related to the effects of some of these branded or unbranded OG programs. It remains to be seen whether multisensory structured language/phonics instructional approaches add value to approaches that use explicit, systematic structured language/phonics instruction without a multi-sensory component. When rigorous reading research furnishes conclusive answers to these questions, it is our hope that all will be receptive to any potential evolution in our scientific understandings. ■

Acknowledgements. The authors would like to thank Dr. Donald Compton and Dr. Nathan Clemens for reading and providing feedback on earlier versions of this commentary.

References

- Al Otaiba, S., Rouse, A., & Baker, K. (2018). Elementary grade intervention approaches to treat specific learning disabilities, including dyslexia. *Language, Speech, and Hearing Services in Schools (Online)*, 49(4), 829–842.
- Campbell, M. L., Helf, S., & Cooke, N. L. (2008). Effects of adding multisensory components to a supplemental reading program on the decoding skills of treatment resisters. *Education and Treatment of Children*, 31(3), 267–295.
- Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities: From identification to intervention*. Guilford Publications.
- Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., and Tilly, W.D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide*. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/>
- Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness*, 13(2), 401–427.
- Petscher, Y., Cabell, S. Q., Catts, H. W., Compton, D. L., Foorman, B. R., Hart, S. A., ... & Wagner, R. K. (2020). How the Science of Reading Informs 21st-Century Education. *Reading Research Quarterly*, 55, S267–S282.
- Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham—based reading instruction: A review of the literature. *The Journal of Special Education*, 40(3), 171–183.
- Scammacca, N., Vaughn, S., Roberts, G., Wanzek, J., & Torgesen, J. K. (2007). *Extensive reading interventions in grades K–3: From research to practice*. Portsmouth: RMC Research Corporation, Center on Instruction.
- Schlesinger, N., & Gray, S. (2017). The impact of multisensory instruction on learning letter names and sounds, word reading, and spelling. *Annals of Dyslexia*, 67(3), 219–258.
- Stevens, E. A., Austin, C., Moore, C., Scammacca, N., Boucher, A. N., & Vaughn, S. (2021). Current State of the Evidence: Examining the Effects of Orton-Gillingham Reading Interventions for Students With or at Risk for Word-Level Reading Disabilities. *Exceptional Children*. Advance online publication.
- Swanson, H. L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of learning disabilities*, 32(6), 504–532.
- Thorpe, H. W., & Borden, K. S. (1985). The effect of multisensory instruction upon the on-task behaviors and word reading accuracy of learning disabled children. *Journal of Learning Disabilities*, 18(5), 279–286. <https://doi.org/10.1177/002221948501800507>
- Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics: A practice guide*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. <https://www.centeroninstruction.org/files/Intensive%20Interventions%20for%20Students%20Struggling%20in%20Reading%20&%20Math.pdf>
- Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, 51(6), 612–624.
- What Works Clearinghouse. (2010a). Alphabetic Phonics. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_alpha_phonics_070110.pdf
- What Works Clearinghouse. (2010b). Barton Reading & Spelling System. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_barton_070110.pdf
- What Works Clearinghouse. (2010c). Orton-Gillingham-based strategies (unbranded). U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_ortongill_070110.pdf
- What Works Clearinghouse. (2010d). Wilson Reading System. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_wilson_070110.pdf



Emily Solari

Emily Solari, Ph.D., is the coordinator and professor in the Reading Education program in the Department of Curriculum Instruction and Special Education at University of Virginia. Her scholarship focuses on the prevalence, predictors, and underlying mechanisms that drive reading development with the goal of developing and testing the efficacy of targeted interventions to prevent and ameliorate reading difficulties. She is particularly focused on efforts to translate the scientific evidence base in reading by engaging with practitioners and policy makers to leverage scientific evidence to improve practice in school settings.



Yaacov Petscher

Yaacov Petscher, Ph.D., is an Associate Professor at Florida State University, an Associate Director at the Florida Center for Reading Research, and the Deputy Director of the National Center on Improving Literacy. His interests include data analysis, baking bread with his daughters, and forgetting to wash the dishes.



Colby Hall

Colby Hall, Ph.D., is an assistant professor of Reading Education in the Department of Curriculum, Instruction and Special Education at the UVA School of Education and Human Development. She earned her Ph.D. in Special Education from the University of Texas at Austin. Professor Hall's research focuses on reading development, assessment, and instruction. She has played a primary role in the development and testing of various reading instructional interventions for elementary and middle school students with or at risk for reading difficulties.