

Using Bayesian Meta-Analysis to Explore the Components of Early Literacy Interventions

September 2023
WWC 2023008
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation (NCEE) at IES



CONTENTS

USING BAYESIAN META-ANALYSIS TO EXPLORE THE COMPONENTS OF EARLY LITERACY INTERVENTIONS.....	1
WHY THIS SYNTHESIS?.....	2
WHAT DID THIS SYNTHESIS DO?.....	4
Research questions.....	4
Coding components of early literacy interventions.....	5
About the early literacy interventions.....	8
WWC study data.....	9
The meta-analytic model.....	9
WHAT DID THIS SYNTHESIS FIND?.....	13
What are common components of early literacy interventions?.....	13
What are the component domains of early literacy interventions found to have positive or negative effects?.....	17
Which component domains are positively associated with the effects of early literacy interventions?.....	20
To what extent can the component domains explain variation in intervention effects?.....	23
TAKE-AWAYS AND IMPLICATIONS.....	24
STUDIES INCLUDED IN THE META-ANALYSIS.....	27
OTHER REFERENCES.....	32

TABLES

1.	Implementation of selected components in the instructional practices, structures, and academic supports component type.....	14
2.	Implementation of selected components in the assessment and placement component type	15
3.	Implementation of selected components in the educator supports component type.....	15
4.	Implementation of selected components in the organizational structures and supports component type	16
5.	Implementation of selected components in the non-academic student supports component type	16

FIGURES

1.	The four levels of the component taxonomy.....	5
2.	The five types of early literacy components and 15 component domains	7
3.	Who delivered the interventions?	8
4.	Implementation of component types in early literacy interventions.....	13
5.	Effects of early literacy interventions on independent alphabetic outcomes	17
6.	Component domains included in early literacy interventions that had positive or negative effects on independent alphabetic outcomes	19
7.	Magnitudes of relationships between each component domain and intervention impacts on independent alphabetic outcomes.....	21

Using Bayesian Meta-Analysis to Explore the Components of Early Literacy Interventions

Elias Walsh, John Deke, Silvia Robles, Andrei Streke, and Dan Thal

September 2023

Improving literacy instruction so that all students achieve grade-level proficiency in reading and writing remains a critical challenge in the United States. In this report, which is primarily aimed at researchers rather than education practitioners, the What Works Clearinghouse (WWC) applies two methodological approaches new to the WWC that together aim to improve our understanding of how early literacy interventions may work to improve student outcomes. First, this report pilots a new taxonomy developed by early literacy experts and intervention developers as part of a larger IES effort to develop standard nomenclature for the components of literacy interventions. Most, if not all, education interventions include multiple components that, when implemented together, are meant to improve student outcomes. These components might also be thought of as “active ingredients” or “key features.” Then, the WWC uses Bayesian meta-analysis—a statistical method to systematically summarize evidence across multiple studies that is new to the WWC—to estimate the associations between intervention components and intervention impacts. If some components are associated with positive intervention effects, researchers and innovators might be able to use this information to refine interventions, or develop new ones, and then design studies to test whether the interventions provide even greater benefits to students. The purpose of this methodological report is to explore to what extent observed components of interventions can explain which interventions have positive effects and develop take-aways for future applications of similar methods. The report uses studies of early literacy interventions to conduct this trial of the new methods.

The WWC reviews and summarizes existing research in education, and this report uses WWC study data on literacy interventions—defined as any educational practice, program, product, or curriculum implemented with small groups or broadly at the classroom or school level. The sample includes 29 studies of 25 early literacy interventions in grades K-3 that examined impacts on measures of alphabetic outcomes. Alphabetic skills are important for students to develop early literacy skills and are commonly reported in early literacy studies. They include phonics, phonemic awareness, phonological awareness, and letter identification. Studies were included in the analysis if they were previously reviewed by the WWC and met the WWC’s rigorous research standards. Therefore, the findings of this synthesis may not be representative of all literacy research or literacy interventions. Moreover, the studies examined a wide range of student samples. As such, this report’s findings reflect those for K-3 students generally, rather than only those who are below grade level in reading.

The synthesis team modified and piloted a new taxonomy developed for IES (Scher & Martinez, 2022) to document which intervention components were implemented in each study, and then examined which components have positive associations with the effects of interventions on alphabetic outcomes using Bayesian meta-analysis. This method cannot conclude that specific components caused improvements in alphabetic outcomes. However, it does seek to identify which components might be positively associated with the effects of interventions on measures of alphabetic outcomes.

Because the take-aways from this synthesis are exploratory, they do not inform specific recommendations to support educators or innovators. Instead, this report aims to assess the promise of applying both the new component taxonomy and synthesis methodology to potentially provide these recommendations in the future. The report also suggests directions for future research.

Key take-aways

- *Findings from this synthesis suggest that disentangling which components of interventions are positively associated with impacts is challenging, and the components examined in this synthesis appear to have a limited role in explaining variation in intervention effects on alphabetic outcomes. Although some components are related to intervention effects, other factors also influence the success of an intervention. These factors could include teacher quality; school climate; school resources; or implementation frequency, duration, and quality.*
- *This method identified positive associations between intervention impacts on alphabetic outcomes and components related to using student assessment data to drive decisions, including about how to group students for instruction, and components related to non-academic student supports, including efforts to teach social-emotional learning strategies and outreach to parents and families.*
- *The method found that most early literacy interventions in this synthesis had positive effects on measures of alphabetic outcomes for students. Most interventions were multifaceted, delivered to small groups of six or fewer students or individual students, and required personnel beyond classroom teachers.*

WHY THIS SYNTHESIS?

English language and literacy skills—listening, speaking, reading, and writing—are foundational to students’ academic achievement and setting them on a path to high school graduation and readiness for college and careers (Kuncel & Hezlett, 2010). Mastering foundational reading skills, including phonics, phonemic awareness, phonological awareness, and letter identification—collectively referred to as alphabetic—in the early grades is also vital for setting students up for later college and career success (Adams, 1990; National Reading Panel, 2000; Slavin et al., 2009; Snow et al., 1998; Wrulich et al., 2014). Yet in 2022, only one-third of grade 4 students scored at or above a proficient reading level on the National Assessment of Educational Progress—down from 37 percent in 2017 (The Nation’s Report Card, 2023).

The What Works Clearinghouse (WWC) at the Institute of Education Sciences (IES) routinely reviews and summarizes research on educational interventions, defined as any practice, program, product, or curriculum implemented with small groups or broadly at the classroom or school level. The WWC has developed dozens of products that summarize the research on literacy interventions, including Practice Guides with evidence-based recommendations on [preparing young children for school](#) (Burchinal et al., 2022), [providing reading interventions for students in grades 4–9](#) (Vaughn et al., 2022), [teaching secondary students to write effectively](#) (Graham et al., 2016), [developing foundational reading skills in K–3](#) (Foorman et al., 2016), and [teaching elementary school students to be effective writers](#) (Graham et al., 2012).

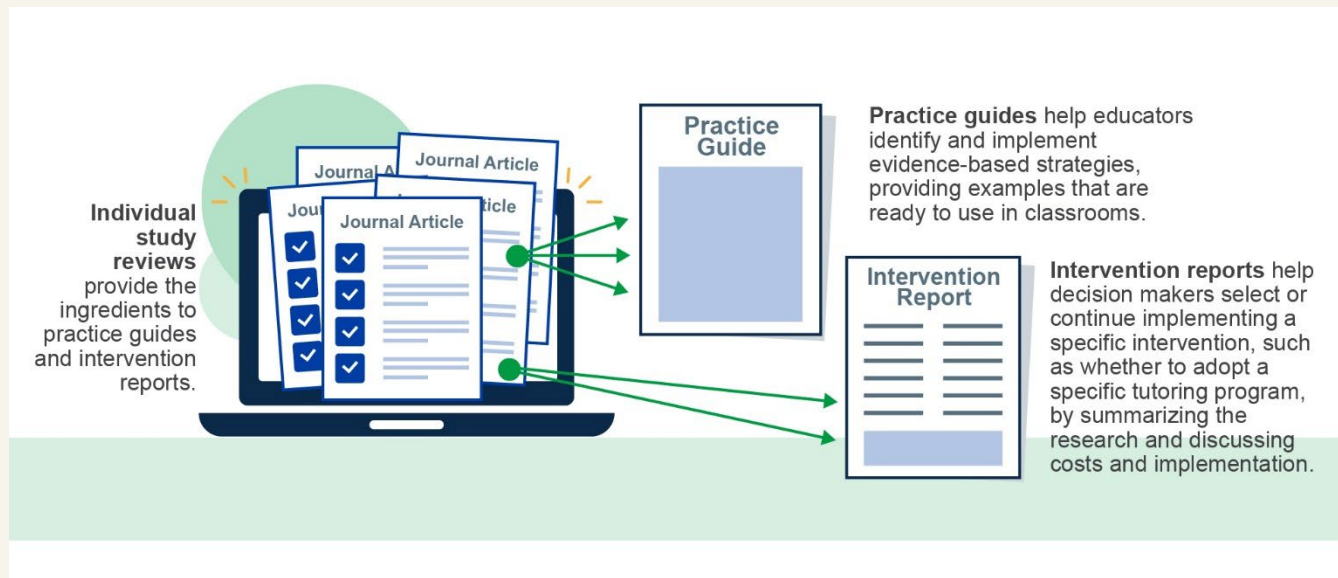
Yet a challenge in understanding ‘what works’ in education is knowing which components of interventions matter the most for student learning. Frequently, educational interventions contain multiple bundled components, and it is often not known why some interventions improved student outcomes more than others. Recognizing this challenge, IES has recently emphasized the potential value of better understanding the components of interventions in its [Standards for Excellence in Education Research](#) (SEER). Knowing which components may lead to increased learning can inform new interventions that may be more cost-effective or provide even greater educational benefits to students. For example, this knowledge might provide more flexibility by allowing:

- **Professional development programs** to emphasize the strategies that may matter most.
- **Education decision makers** at all levels to make evidence-informed changes to the models and programs already in use, rather than swapping entire programs.
- **Intervention innovators**, including the developers of branded models, to revise and refine their products.
- **Funders and researchers** to focus resources on testing the effectiveness of interventions that include components that may be more likely to improve outcomes, and on testing individual components to rigorously verify their effectiveness.

While a range of people working in the field of education might benefit from better understanding which components of interventions matter, the intended audience for this methodological report is narrower and primarily aimed at researchers.

About the WWC and its products that support evidence-based decision making

The WWC reviews existing research in education, determines the quality of the research, and summarizes findings from the high-quality research to help educators and decision makers identify and implement evidence-based programs and practices.



WHAT DID THIS SYNTHESIS DO?

This report applies both a new taxonomy for classifying components of early literacy interventions and a new synthesis methodology to explore the components included in early literacy interventions in high-quality effectiveness studies reviewed by the WWC. The report uses studies of early literacy interventions to conduct a trial of the new methods. The purpose is to assess to what extent components explain why some interventions are successful. Because the findings in this report are exploratory, they are not meant to inform specific recommendations related to evidence-based decision making or the design of literacy interventions. Instead, this report aims to assess the promise of the new approaches to potentially provide these recommendations in the future.

To explore which intervention components are positively associated with the effects of interventions on alphabetic outcomes, the team examined high-quality effectiveness studies of early literacy interventions in grades K-3 to understand what was implemented. The team first piloted a new taxonomy to code individual components of early literacy interventions in each study, placing them into groups of components called “component domains,” which are themselves placed within groups of domains called “component types.” Then, the team conducted a meta-analysis, which is a statistical method to synthesize evidence across multiple studies. Meta-analyses have the potential to provide more comprehensive and balanced information about the effectiveness of interventions than individual studies because they detect patterns across multiple studies, which guards against overreliance on the results of a single study. Meta-analyses summarize existing research and do not directly test the effectiveness of interventions. Therefore, meta-analyses can provide information about which intervention components are positively associated with the effects of interventions, but they cannot conclude that those components caused the improved outcomes. The primary purpose of the meta-analysis is to measure the associations between each component domain and the impacts of early literacy interventions on measures of alphabetic outcomes.

RESEARCH QUESTIONS

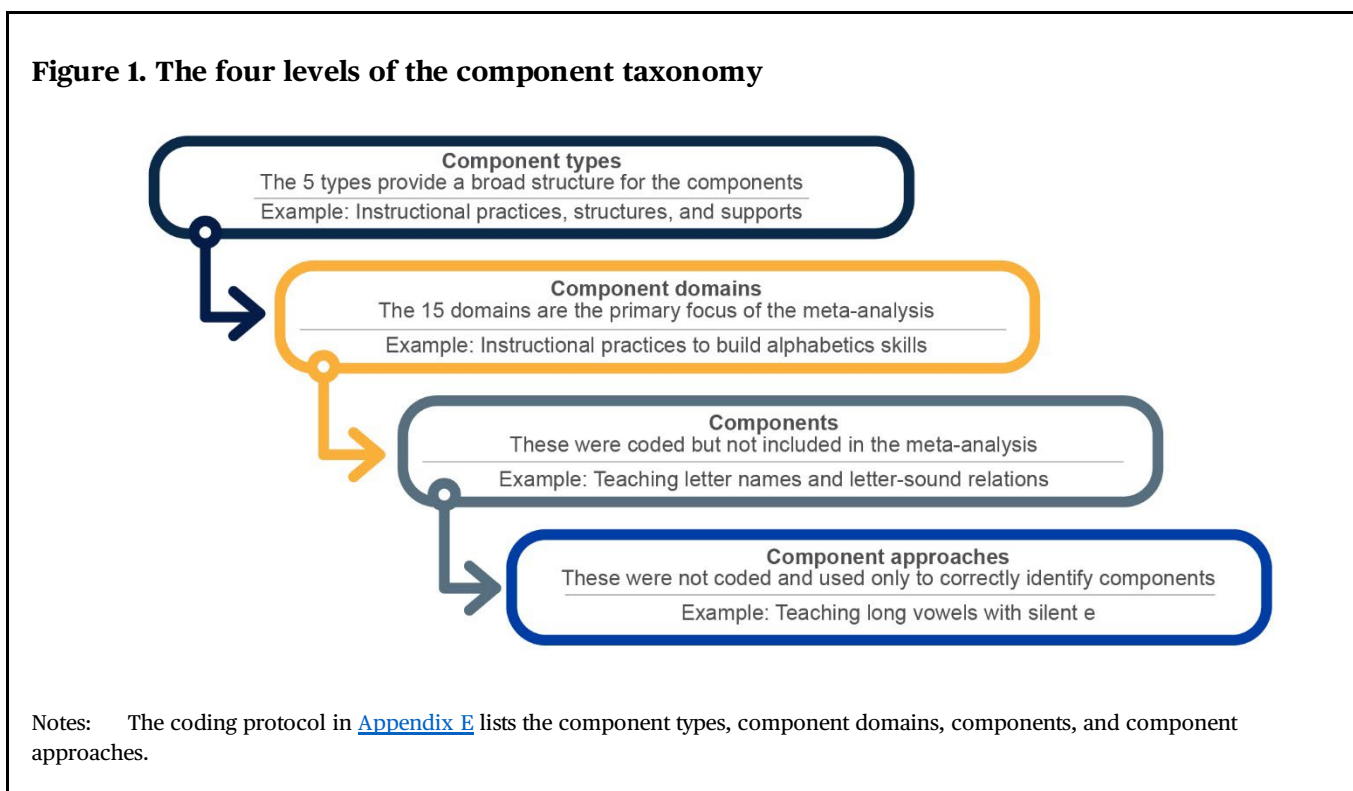
This synthesis addresses four research questions about the components of early literacy interventions examined in high-quality effectiveness studies and their relationships with the impacts of the intervention on alphabetic outcomes. The first two questions are intended to answer questions about the intervention components that were implemented in each study, which the synthesis team documented using a new taxonomy. The second two questions are intended to answer questions related to the results of the Bayesian meta-analysis. Specifically:

1. What are the common components included in early literacy interventions?
2. What are the component domains included in early literacy interventions that had positive or negative effects on alphabetic outcomes for students?
3. Which component domains are positively associated with the effects of early literacy interventions on alphabetic outcomes for students?
4. To what extent do the component domains explain variation in the effects of early literacy interventions on alphabetic outcomes?

CODING COMPONENTS OF EARLY LITERACY INTERVENTIONS

To understand the components of interventions in each study, the team first trained research staff to systematically code intervention components in 29 studies of 25 literacy interventions in grades K-3 that examined impacts on measures of alphabetic outcomes. To accurately record the components of the interventions examined in each of the 29 studies, the synthesis team piloted a new taxonomy with standard nomenclature for the components of literacy interventions. The synthesis team used and built upon a taxonomy previously developed for another IES project (Technical Assistance Supporting Evidence-Building and Use under contract number 91990020F0052; Scher & Martinez, 2022) with input from early literacy experts and intervention developers. As part of the structured coding process based on the taxonomy, the synthesis team also requested additional information from study authors. [Appendix A](#) describes the approach used to identify intervention components in more detail.

The component taxonomy defines four levels of intervention components. At the highest level is the component type, which identifies the broad strategy used by the intervention. The five component types are instructional practices, non-academic supports, organizational structures, educator supports, and assessment and placement. Component domains form the second level of the taxonomy, and the 15 component domains identify more specific practices within a component type, such as “building writing skills” or “incorporating social-emotional learning strategies.” The taxonomy further breaks down the component domains into 81 components and additional component approaches, which are the specific approaches to implementing each component. The four levels of the taxonomy are represented in Figure 1.

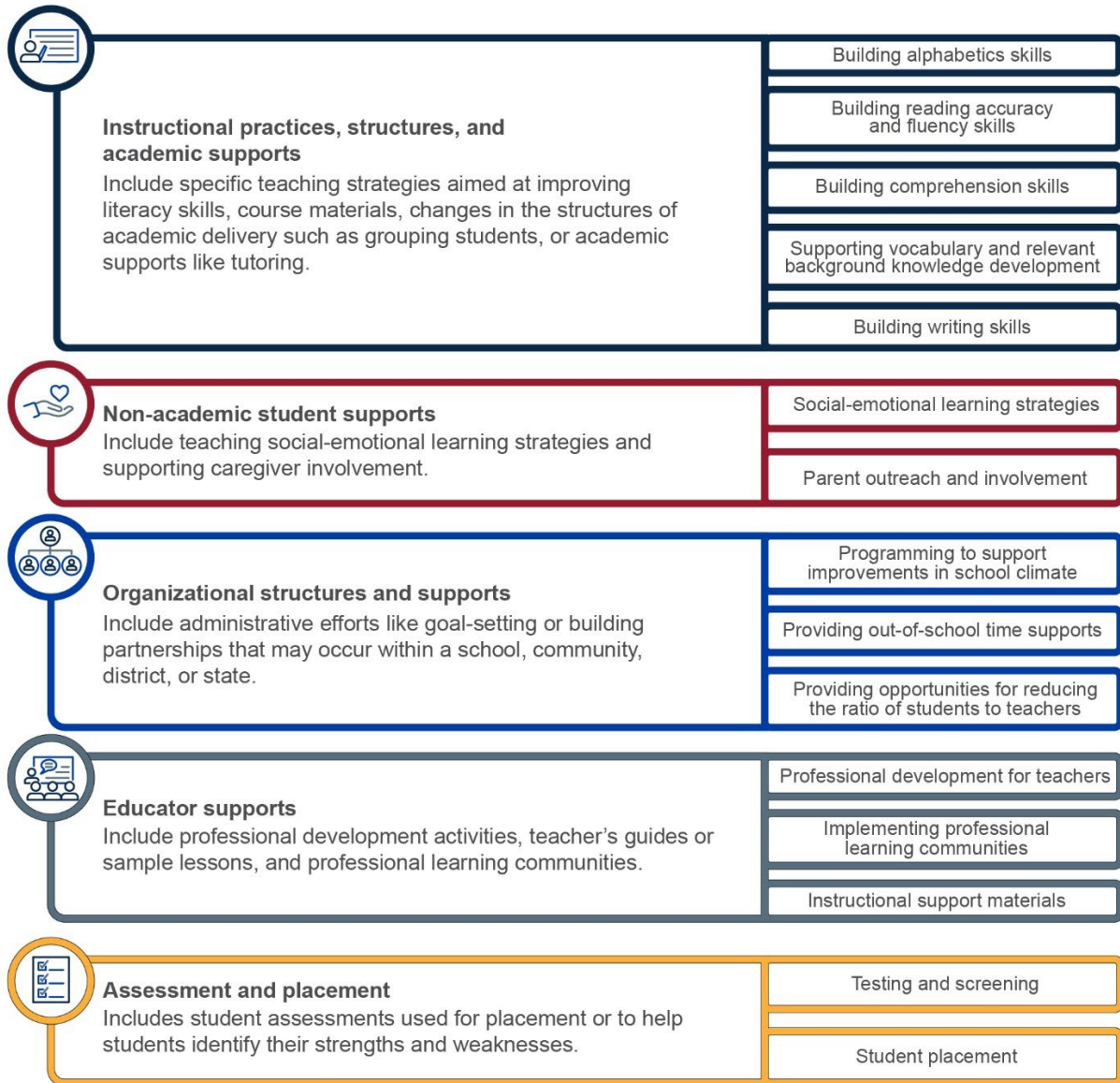


For this report, the team coded the components, and aggregated the components to the level of component domain for the analysis. The meta-analysis focuses on measuring the relationships between the 15 component domains and alphabetic outcomes, rather than between the components or component approaches that are nested within the component domains and alphabetic outcomes. Because the meta-analysis includes just 29 studies that measured impacts on alphabetic outcomes, it is not possible to measure relationships between intervention effects and each of the 81 components with precision. The team coded the components of interventions to provide insight about how the component domains were implemented but did not code which component approaches were used to implement the components. Most studies did not provide enough detail to support reliably coding component approaches. However, the coders did refer to the component approaches to help correctly identify which components an intervention implemented.

The five component types and 15 component domains included in the early literacy component taxonomy are shown in Figure 2. For example, the first component type in Figure 2 is instructional practices, structures, and academic supports. The first component domain within this type encompasses a set of components related to building alphabetic skills. Alphabetic skills include phonics, phonemic awareness, phonological awareness, and letter identification. Phonics instruction—which helps students develop the ability to associate letters and letter combinations with sounds and to use this information to blend the sounds into syllables and read or decode words—would fall within the domain for building alphabetic skills, for example. Within the component domain of building alphabetic skills, there are 13 components (not shown in Figure 2): working with rhyming words, engaging in explicit or direct instruction, teaching students to manipulate segments of sounds in speech, teaching letter names and letter-sound relations, linking letter-sound relationship knowledge to phonemic awareness, implementing handwriting programs to call attention to letter shape, teaching blended and sound-spelling patterns in words, introducing common sound-spelling patterns, teaching students to recognize common word parts, teaching students to read regular and irregular high-frequency words, introducing phonetically irregular words, reading decodable words in isolation and in text, and using other instructional practices to build alphabetic skills. For each component, there are different component approaches. For example, to link letter-sound relationship knowledge to phonemic awareness, teachers might use letter tiles or magnetic letters to build or change words based on sounds, or gradually add more advanced words to support understanding of phonemic patterns (for example, by changing “can” to “cane” or “fat” to “flat”). The components within each domain are listed in [Appendix A](#). The coding protocol the team used to code the components of interventions is included in [Appendix E](#) and lists the components along with the component approaches.

The coding focused on whether the components of early literacy interventions were present or absent in the studies. The synthesis team did not systematically code contextual factors—such as teacher quality, school climate, or school resources—or the quality of implementation for each component. The team attempted to code whether studies reported any specific implementation challenges, but most studies did not provide enough information to confidently assess the quality of implementation. The team recorded who delivered the intervention and some details about the method of delivery, such as whether the intervention was delivered one on one to students or to small groups (defined in the taxonomy as including six or fewer students) or a whole class. These implementation factors provide descriptive information about the interventions, but they were not included in the meta-analysis. Future meta-analyses could identify implementation features to systematically code and include to learn to what extent these features are positively associated with intervention effects.

Figure 2. The five types of early literacy components and 15 component domains



Notes: The coding protocol in [Appendix E](#) includes complete definitions of the component types and domains.

ABOUT THE EARLY LITERACY INTERVENTIONS

The 25 early literacy interventions include the following types of interventions as classified by the WWC:

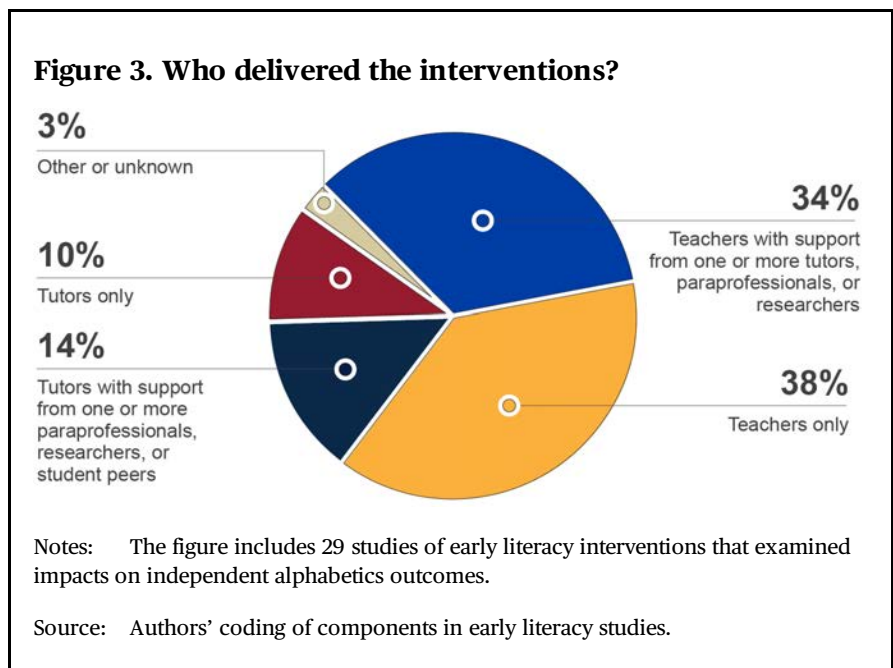
- **Comprehensive reading or literacy curricula** intended to serve as a school’s primary literacy instruction program (for example, *Enhanced Core Reading Instruction*)
- **Software applications** designed to improve literacy skills and used with an entire class or individual students (for example, *Abracadabra* and *MindPlay*)
- **Programs**, including supplemental programs that are intended to enhance whole-school or whole-classroom literacy instruction (for example, *Early Reading Intervention*), as well as programs for students experiencing difficulty reading, such as tutoring programs (for example, *Reading Recovery*)
- **Practices and strategies** intended to address a specific literacy-related skill (for example, phonics or extended vocabulary instruction)
- **Whole-school reform models** that integrate curriculum, school culture, and family and community supports (for example, *Success for All*®)

See Table B1 in [Appendix B](#) for the full list of interventions included in the meta-analysis.

The approach to delivering interventions differed considerably across studies but frequently involved additional staff to support implementation or work with small groups of students. Teachers delivered the intervention alone in 38 percent of the studies (Figure 3). Teachers along with a tutor, a paraprofessional, or a researcher (or multiple staff from these groups together) delivered the intervention in another 34 percent of studies. Tutors delivered the intervention alone or with support in 24 percent of studies.

Across the studies, 72 percent of students participated in at least part of the intervention in small groups, 34 percent of students participated one on one with a teacher or other instructor, and 21 percent of students participated with their whole classrooms. These percentages sum to more than 100 percent because interventions could include multiple student configurations. Including additional staff and working with small groups or individual students, as many of these interventions do, might be expected to improve student outcomes based on prior research (for example, Chetty et al., 2011, and Schanzenbach, 2006).

Lastly, most interventions (90 percent) were delivered during school hours, and 10 percent were delivered after school hours. At least part of the intervention was delivered using technology in 20 percent of studies.



WWC STUDY DATA

The study data were collected from the WWC’s publicly available database of [reviewed studies](#). The WWC reviews studies both as part of systematic reviews—which are summarized in intervention reports or practice guides—and individually, outside of systematic reviews. The WWC database includes information about the quality of the study, findings from the study, and other details about each study, such as the setting, demographics of study participants, and sample size. [Appendix B](#) provides more details about the WWC’s process for reviewing studies.

The team identified 29 high-quality studies of literacy interventions for students in grades K-3 that met specific inclusion criteria, including meeting the WWC’s rigorous research standards and examining impacts on measures of alphabetic outcomes for students. Citations for each study are included in the references, with links to each WWC study page for details about the study, the intervention, and the results of the WWC’s review. [Appendix B](#) outlines all criteria for inclusion of studies in this analysis. The studies include only those that have been reviewed by the WWC, which means that findings from this synthesis may not generalize to all literacy interventions. For example, some literacy interventions have not been studied using designs that can meet WWC standards. Also, the WWC has not exhaustively reviewed all literacy research, and the WWC might be more likely to review certain literacy interventions.

The studies were conducted in the United States, Australia, Canada, New Zealand, and the United Kingdom and collectively included 22,596 students. Thirty-eight percent of the studies included students in kindergarten, 66 percent included grade 1 students, 45 percent included grade 2 students, and 24 percent included grade 3 students. The studies examined a wide range of student samples, including English learners, students experiencing difficulty with reading, and whole-school populations. Therefore, this report’s findings do not necessarily reflect findings only for students who are below grade level in reading, and instead reflect findings for K-3 students more generally.

The 29 studies examined the effectiveness of 25 distinct interventions and include 108 findings that measure impacts of the interventions on alphabetic outcomes. A finding is an impact estimate for a particular outcome measure and a particular sample within a study. For example, a study might include findings for measures of phonemic awareness and letter identification, and for students in each of grades K-3. The impact estimate measures whether and how much the intervention appeared to improve each outcome.

The synthesis includes 100 of the 108 impact estimates for alphabetic outcomes that were measured independently of study authors or intervention developers (Walsh et al., 2023). Previous research indicates the impacts of the literacy interventions tend to be smaller for measures developed independently of study authors and developers and independent outcomes may have more practical significance than nonindependent outcomes. (Wolf & Harbatkin, 2022). As such, this report focuses on the findings that used independent measures.

THE META-ANALYTIC MODEL

The meta-analysis has three objectives: (1) provide information about which early literacy interventions in high-quality studies had positive effects on alphabetic outcomes, (2) explore which intervention component domains are positively associated with the effects of interventions on measures of alphabetic outcomes, and (3) measure to what extent the component domains examined in this synthesis can explain variation in intervention effects on alphabetic outcomes. A positive association between a component domain and the effects of interventions

means that, for the studies and interventions included in the synthesis, the impacts of interventions that include the component domain tended to be more positive than those without the component domain after accounting for the presence of any other measured component domains also implemented in the interventions.

The meta-analysis removes variation in intervention effectiveness across grade levels, outcome domains, and outcome measure type. These factors were held constant to focus attention on differences in intervention effectiveness related to component domains, rather than on differences due to how the components were studied. The meta-analysis also applies other adjustments to ensure that the impacts of interventions are comparable across studies. These adjustments account for sampling error due to small sample sizes, statistical bias in impact estimates that arises from the quality of the research design, and publication bias. These adjustments are described further in [Appendix C.2](#). These adjustments are intended to produce more accurate information about the true impacts of interventions and, as a result, tend to produce smaller effect sizes than impacts without these adjustments.

The meta-analysis uses a Bayesian approach. This type of synthesis differs from the way the WWC typically synthesizes evidence across studies. The model accounts for the issues described above that can sometimes lead to misleading results in other meta-analyses, but the Bayesian approach is not necessary to make most of these adjustments. Instead, the primary benefit of the Bayesian approach is in how it provides a more useful assessment of uncertainty. Statistical significance has often been used to communicate confidence in research findings, but the American Statistical Association warns that statistical significance does not mean what most people think it means and can lead to overconfidence in research findings (Greenland et al., 2016; Wasserstein & Lazar, 2016). By itself, statistical significance provides no information about the probability that an estimated relationship is due to random chance as opposed to a genuine effect. But with Bayesian methods, researchers can calculate the probability that a relationship is positive. Consequently, the Bayesian approach provides new opportunities to report research findings in ways that may ultimately be easier for decision makers to understand. Although the Bayesian approach may be unfamiliar to some audiences, it has the potential to convey the meaning of research findings more accurately to a broad audience. The Bayesian approach is a new way to characterize and report on the evidence for the WWC, and the findings from this meta-analysis are not aligned with definitions of evidence from the U.S. Department of Education. As such, the findings in this report are exploratory. [Appendix C](#) provides more information about the Bayesian method, its advantages compared to other widely used meta-analytic methods, and the technical details of estimating the Bayesian meta-analytic model.

The Bayesian meta-analysis draws on data about intervention effects from WWC-reviewed studies across multiple topic areas in addition to early literacy and other sources to form priors for parameters in the model. In addition, while the meta-analysis focuses on findings on alphabetic outcomes, the meta-analysis model also draws from 25 additional studies of early literacy interventions with findings on other literacy outcomes. Relationships between component domains and intervention effects on a range of literacy outcomes estimated from these additional studies inform the estimated relationships with measures of alphabetic outcomes. In a frequentist analysis, the findings from these 25 additional studies might be excluded from the analysis. However, in a Bayesian analysis, including these additional findings improves the precision of the relationships estimated from the meta-analysis (Gelman et al., 2013). The synthesis team used the same approach to code intervention components in these additional studies that the team used to code components in the 29 studies that examined impacts on measures of alphabetic outcomes. The literacy outcomes in these additional studies are described in [Appendix B](#). [Appendix D](#) includes synthesis findings based on all literacy outcomes. [Appendix C](#) provides more details about the Bayesian model, including all priors used in the model.

Readers should use caution when interpreting the findings from the exploratory meta-analysis. In particular, the results from the Bayesian analysis have the following limitations:

Associations between component domains and intervention impacts should not be interpreted as causal. Readers should not interpret the results from this synthesis as causal because the synthesis cannot separate the effect of each component domain from all other factors related to intervention effectiveness, such as who delivered the intervention; frequency, duration, and quality of implementation; teacher quality; available resources; or measurement properties of outcome measures besides their domain and independence. For example, the meta-analysis cannot conclude that a component domain, by itself, led to larger impacts for an intervention. It is possible that a component domain that is positively associated with the effects of interventions has no effect on student outcomes because the interventions that include the component domain also include other effective but unmeasured components. The meta-analysis can disentangle the 15 component domains that were coded for the synthesis, but it is not possible to separate the effect of these components from factors (or components) that are unmeasured. Therefore, a component domain that is found to be positively associated with the effects of interventions on measures of alphabetic outcomes does not mean that the component domain *caused* the intervention to be more effective. Nor would this finding necessarily indicate that adding that component domain to another intervention will improve literacy outcomes.

To verify whether a component domain causes improved literacy outcomes, researchers would need to conduct a new evaluation focused specifically on measuring the causal impact of the particular component domain. For example, to assess the value of a professional development component, an evaluation could compare the outcomes for students who either received an intervention with the professional development component or the same intervention without that component. Because most studies in the meta-analysis included bundles of components, some of which may not be captured in the component taxonomy, the meta-analysis cannot substitute for focused evaluations of particular component domains.

Generalizability. The analysis is limited to interventions and studies contained in the WWC database, which are those the WWC has rigorously studied and reviewed, as well as to the samples and contexts of those studies. This has several consequences for the generalizability of the findings. First, only early literacy interventions that have been rigorously studied are represented in the analysis, but these interventions may not be representative of all early literacy interventions. Second, the studies in the database examine a wide range of student samples. Therefore, this report's findings do not necessarily reflect findings for students who are below grade level in reading and instead reflect findings for all students more generally. It is not possible to make precise statements about how effective the interventions or their components are for different groups of students because studies do not always provide subgroup findings for these student groups. Third, among the rigorous studies of early literacy interventions, some have not been reviewed by the WWC and would therefore not be represented in the findings. For example, the WWC may have been more likely to select studies with more favorable findings to review. And finally, some studies of early literacy interventions that have been conducted are never published. This might occur when study authors choose to publish only studies with positive findings. The meta-analysis does attempt to account for these last two issues, as described in [Appendix C](#).

Small sample sizes. The meta-regression analysis uses findings about intervention effectiveness from 29 studies and 25 interventions to distinguish relationships with impacts for the 15 component domains. A larger number of studies and interventions would produce more precise findings, and the reporting in this synthesis appropriately reflects the uncertainty of the findings. Additionally, although the meta-analysis includes findings for additional literacy outcomes across a total of 54 studies, there are smaller numbers of findings for other groups of literacy

outcomes besides alphabets, like reading comprehension outcomes. This means that this analysis was limited to examining relationships between component domains and intervention effects on alphabets outcomes, although we report results for groups of other outcome measures in [Appendix D](#). Furthermore, the relationships for alphabets might change when using a larger study dataset with more alphabets findings that allows for more precise estimates. Because this study employs a Bayesian meta-analysis, the results for any subgroup of studies will tend to be adjusted (or “shrunk”) towards the results for the full sample. Therefore, the relationships estimated for alphabets outcomes are influenced by relationships with other literacy outcomes. Finally, the small sample size does not allow for measuring precise relationships between each of the 81 components within the component domains and impacts of literacy interventions, although we discuss the likely magnitude of these impacts in [Appendix D](#).

Component uncertainty. The synthesis team documented which components appear to be present in the interventions, but there was some uncertainty without direct knowledge of the interventions. The studies do not always provide the level of detail needed to assess whether a component was present. To address this limitation, the synthesis team supplemented the publicly available documentation with information collected by querying study authors. In cases where authors did not respond to queries, some components coded as not present might actually be present. Among components coded as present, it is not possible to reliably determine if they were implemented as intended.

Actionable variation in intervention impacts. The extent to which component domains can explain variation in intervention effects on alphabets outcomes may be limited by practical and measurement issues. If, for example, the amount of actionable variation in intervention impacts were only 20 percent, then explaining only a portion of that remaining variation would appear small relative to the total amount of variation. The amount of actionable variation may be smaller than the total amount of variation in intervention impacts for two reasons. First, if any intervention—no matter how well designed—is limited in its capacity to improve student achievement, some variation in impacts could not be explained by the component domains. Second, it may be case that the quality of measures obscures the importance of component domains. The psychometric properties of measures—including their reliability and validity—vary widely in education research. Although the WWC (2022) has requirements for outcome measures, differences in quality remain. If outcome measures are poor indicators of the underlying constructs components are meant to affect, our interpretation of components’ efficacy is likely skewed.

Despite these caveats, the results of this synthesis offer insights into both literacy interventions and the methodological approach. These include descriptive information about literacy interventions and their components, exploratory findings about which component domains are positively associated with the effects of interventions, and lessons for researchers interested in applying similar methods.

WHAT DID THIS SYNTHESIS FIND?

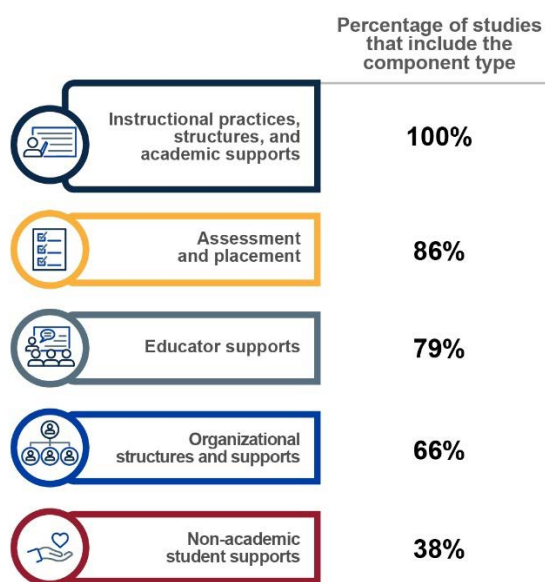
WHAT ARE COMMON COMPONENTS OF EARLY LITERACY INTERVENTIONS?

This section provides descriptive findings based on the application of the taxonomy to the studies of early literacy interventions about which component types, domains, and individual components were implemented in the studies.

Implementation of component types. According to the new taxonomy, the early literacy interventions examined in this synthesis generally include multiple components, with most studies including at least three of the five component types. Of the 29 studies that examined impacts on independent alphabetic outcomes, 21 percent included all five types and another 41 percent included four of the five types.

All studies—100 percent, as shown in the top row of Figure 4—included a component to provide instructional practices, structures, or academic supports. Eighty-six percent of studies included an assessment and placement component. Assessment and placement components include screening of students for eligibility, monitoring student progress, multitiered systems of support, and homogenous or heterogeneous grouping of students (in any group size) based on assessment and screening results. Interventions that do not include assessment and placement components include those delivered to whole classrooms or schools. They also include interventions with one-on-one or small group instruction, but the groups would not be formed specifically based on student assessment data to identify specific groups of students, such as students below grade level in reading.

Figure 4. Implementation of component types in early literacy interventions



Notes: The figure includes 29 studies of early literacy interventions that examined impacts on independent alphabetic outcomes.

Source: Authors' coding of components in early literacy studies.

Seventy-nine percent of studies included educator supports, such as professional development, professional learning communities, or instructional support materials such as sample lesson plans. Sixty-six percent of studies included organizational structures and supports, such as small-group literacy blocks, reduced class sizes, programming to support improvements in school climate, or summer or after-school programming. Less frequently used were non-academic student supports (38 percent), such as social-emotional learning strategies and involving parents and families.

Implementation of component domains. Some of the 15 intervention component domains nested within the five component types defined in the new taxonomy were more widely used than others. For example, 100 percent of the 29 early literacy interventions included a component to build alphabetic skills, which is expected given the focus on early literacy skills. Seventy-nine percent of studies included a component domain to build

reading accuracy and fluency skills and 34 percent included a component domain to build writing skills. The number and percentages of the 29 studies that implemented each component domain are reported in the highlighted rows in Tables 1-5, with one table for each component type. The subsequent rows report the most commonly implemented components within each of the component domains. For example, in the 24 studies that implemented a testing and screening component, 75 percent implemented universal screening for students and 71 percent implemented formative assessments. In the 19 studies that implemented a student placement component, 95 percent used student assessment data to form homogenous skill groups and 53 percent implemented a multitiered system of support or response-to-intervention framework.

Tables 1-5 include all components implemented in at least 80 percent of studies that include the component domain and the two most frequently implemented components for each component domain. [Appendix A](#) provides this descriptive information for all 81 components.

Table 1. Implementation of selected components in the instructional practices, structures, and academic supports component type

Component	Percentage of studies
29 studies (100 percent) implemented at least one component to build alphabetic skills	
Engaging in explicit/direct instruction	90
Teaching letter names and letter sound relations	90
Teaching students to recognize and manipulate segments of sound in speech	83
23 studies (79 percent) implemented at least one component to build reading accuracy and fluency skills	
Providing opportunities for oral reading practice with feedback	91
Supporting oral reading through modeling, scaffolding, and feedback	91
Teaching self-monitoring and self-correcting reading skills	91
22 studies (76 percent) implemented at least one component to build comprehension skills	
Teaching how to use a specific reading comprehension strategy	64
Guiding students through focused discussion on the meaning of the text	64
19 studies (66 percent) implemented at least one component to support vocabulary development	
Providing instruction to build relevant vocabulary and background knowledge	89
Teaching academic vocabulary in context	74
10 studies (34 percent) implemented at least one component to build writing skills	
Providing instruction to develop writing quality	60
Teaching writing conventions	40

Table 2. Implementation of selected components in the assessment and placement component type

Component	Percentage of studies
24 studies (83 percent) implemented at least one testing and screening component	
Implementing universal screening for students in grades K-3	75
Formative assessments through curriculum-based measurement	71
19 studies (66 percent) implemented at least one student placement component	
Using student assessment data to form homogenous skill groups for literacy instruction	95
Implementing a multitiered system of support or response-to-intervention framework to identify students in need of different levels of support	53

Table 3. Implementation of selected components in the educator supports component type

Component	Percentage of studies
22 studies (76 percent) implemented at least one professional development for teachers component	
Supporting instructional practices	91
Supporting the link between student assessment and practice	55
20 studies (69 percent) implemented at least one component to provide instructional support materials	
Accessing and using sample lesson plans provided by developer	85
Accessing and using curricular guides and developer handbooks	70
6 studies (21 percent) implemented at least one professional learning communities component	
Building grade-level professional learning communities	50
Providing common planning/prep time	50

Table 4. Implementation of selected components in the organizational structures and supports component type

Component	Percentage of studies
15 studies (52 percent) implemented at least one component to reduce the ratio of students to teachers	
Scheduling small-group literacy blocks (including groups of any size smaller than a classroom)	93
Using trained assistants or paraprofessionals in literacy blocks	67
5 studies (17 percent) implemented at least one component to provide out-of-school time supports	
Tutoring outside of school time	60
Summer programming	40
After-school programming	40
Partnering with community organizations	40
4 studies (14 percent) implemented at least one component to support improvements in school climate	
Schoolwide efforts to promote positive behaviors	100
Implementing multitiered systems of support around school climate and behaviors	25

Table 5. Implementation of selected components in the non-academic student supports component type

Component	Percentage of studies
9 studies (31 percent) implemented at least one social-emotional learning strategies component	
Teaching self-management skills	89
Support development of a growth mindset	56
6 studies (21 percent) implemented at least one parent outreach and involvement component	
Parent meetings and conferences to discuss learning and growth	50
Programming to encourage parent involvement in reading activities	50
Providing books to families	50

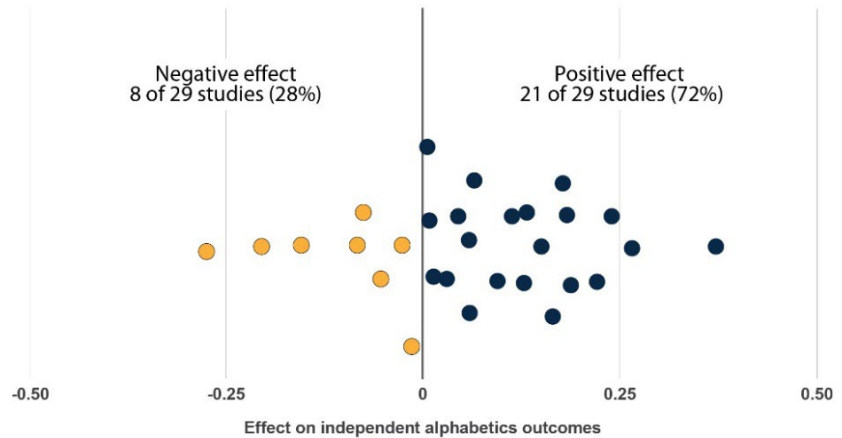
WHAT ARE THE COMPONENT DOMAINS OF EARLY LITERACY INTERVENTIONS FOUND TO HAVE POSITIVE OR NEGATIVE EFFECTS?

The new methodology provides descriptive findings about which component domains were present in studies of early literacy interventions that had positive or negative effects on alphabetic outcomes, which this section describes. As a first step, the synthesis team used estimated effect sizes from the meta-analysis model to determine which interventions had positive or negative effects on independent alphabetic outcomes. The team aggregated across findings within each study by averaging the effect sizes for independent alphabetic outcomes.

The new methodology found that most interventions included in this synthesis had positive effects on literacy outcomes for the students who participated in the studies. In fact, about 72 percent

of the impacts of early literacy interventions on independent alphabetic outcomes were positive, as shown in Figure 5. The median effect size was 0.07 standard deviations. An impact of 0.07 standard deviations is equivalent to lifting a student at the 50th percentile of test scores in a study's sample to the 53rd percentile of test scores—an increase of 3 percentile points.

Figure 5. Effects of early literacy interventions on independent alphabetic outcomes



Notes: Results are based on 100 findings from 29 studies of early literacy interventions that examined impacts on independent alphabetic outcomes. The range of impact estimates in the table reflects variation in impacts across grade levels, outcomes, and interventions. Effect sizes in standard deviations of student achievement are based on adjustments applied in the meta-analysis to address statistical error in impact estimates, file-drawer bias, and the quality of the research design, as described in [Appendix C.2](#). Each dot in the figure represents a study.

Source: Authors' calculations based on data from the WWC's database of reviewed studies.

How to interpret the effect sizes in Figure 5

Positive and negative intervention effects. A positive effect is an effect that is larger than 0 after adjustments applied in the meta-analysis to ensure that the effect sizes are comparable across studies, as described in Appendix C.2, including to address sampling error due to small sample sizes, statistical bias in impact estimates that arises from the quality of the research design, and publication bias. A negative effect is an effect that is smaller than 0 after these adjustments. According to these results from the meta-analysis, interventions with positive effects improved alphabetic outcomes for students relative to a comparison condition whereas interventions with negative effects led to worse outcomes for students.

Independent versus non-independent measures. The effect sizes are measured for independently developed alphabetic measures, rather than those developed by study authors or the developers of the intervention. The impacts of the literacy interventions in this synthesis tend to be smaller for measures developed independently of study authors and developers, as shown in Table D1 in Appendix D.

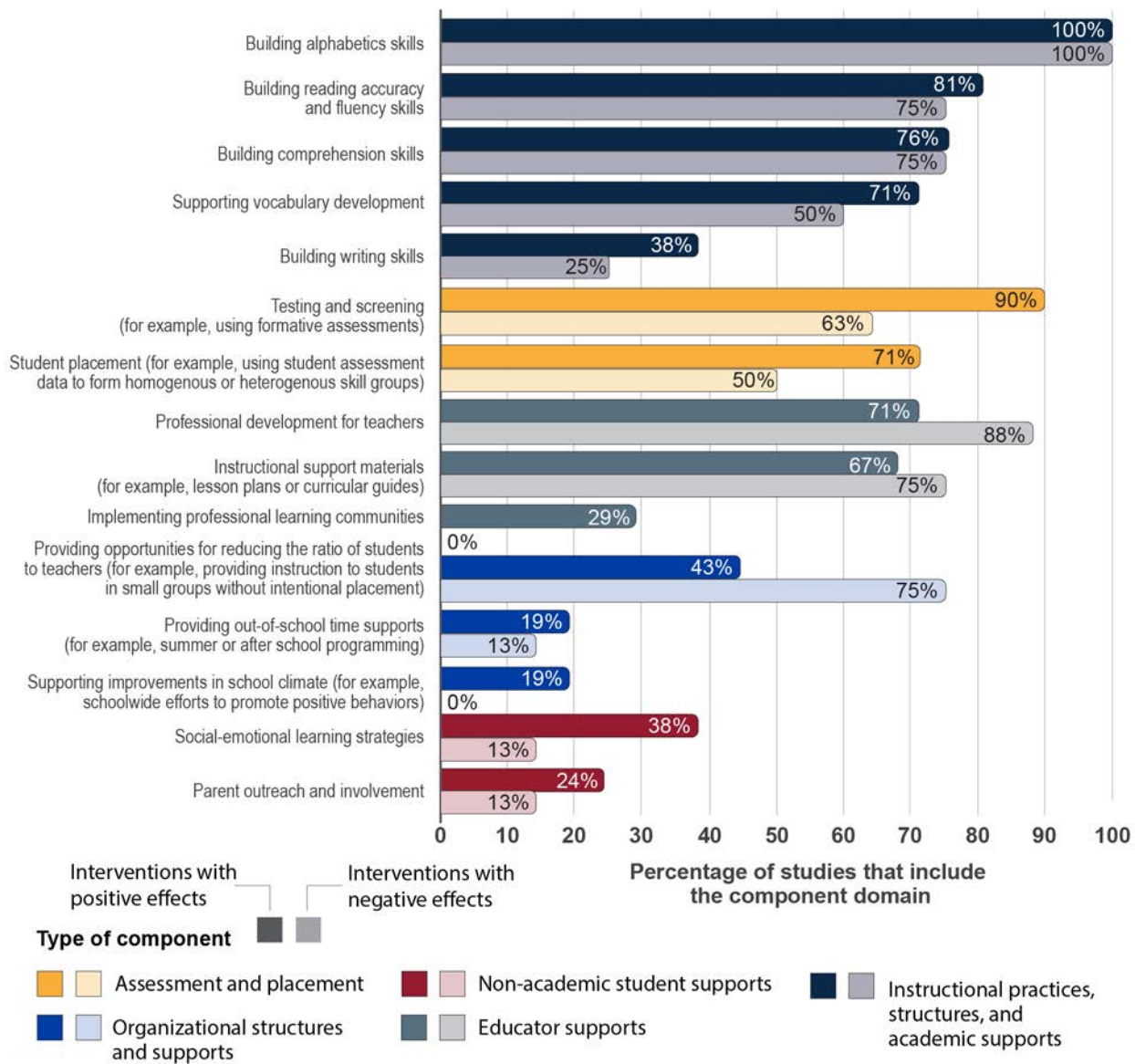
Comparison condition. The impacts of interventions are always measured against a comparison condition in which students receive some services intended to develop student literacy skills. Some interventions provide students in the intervention condition with different literacy instruction from those in the comparison condition and others provide supplemental literacy instruction to the intervention group while providing the standard instructional services to the students in both the intervention and comparison conditions.

In a second step, the synthesis team used the data collected from applying the new taxonomy to measure how often each component domain was implemented in studies of early literacy interventions that had positive or negative effects on alphabetic outcomes, reported in Figure 6. The top, darker bars in each pair report how frequently the component domain was implemented in the 21 studies that examined interventions that had positive effects. The bottom, lighter bars report how frequently the component domain was implemented in the eight studies that examined interventions that had negative effects.

This application of the new methodology found that some component domains were more likely to be present in interventions that had positive effects. For example, the testing and screening component domain is present in 90 percent of interventions with positive effects but only in 63 percent of interventions with negative effects, and student placement is present in 71 percent of interventions with positive effects but in only 50 percent of interventions with negative effects. Other component domains were more likely to be present in interventions with negative effects. For example, professional development was implemented in 71 percent of interventions with positive effects and in 88 percent of interventions with negative effects, and providing opportunities for reducing the ratio of students to teachers was implemented in 43 percent of interventions with positive effects and in 75 percent of interventions with negative effects.

These descriptive differences in which component domains were included in interventions that had positive versus negative effects are suggestive of which component domains might have positive associations with the effects of interventions. However, multiple component domains were implemented together in interventions, so Figure 6 does not answer which components might have positive associations with the effects of intervention, after controlling for the presence of other component domains.

Figure 6. Component domains included in early literacy interventions that had positive or negative effects on independent alphabetic outcomes



Notes: The figure includes 29 studies of early literacy interventions that examined impacts on independent alphabetic outcomes. Interventions with positive effects are those in the 21 studies in Figure 5 with positive impacts on independent alphabetic outcomes. Interventions with negative effects are those in the eight studies in Figure 5 with negative impacts on independent alphabetic outcomes.

Source: Authors' coding of components in early literacy studies.

WHICH COMPONENT DOMAINS ARE POSITIVELY ASSOCIATED WITH THE EFFECTS OF EARLY LITERACY INTERVENTIONS?

This section provides results from the meta-analysis that seeks to disentangle which component domains are positively associated with the effects of interventions. The results in this section describe relationships between component domains and impacts of interventions on independent alphabetic outcomes based on applying the new methodology to this topic area. Because multiple component domains are typically implemented together, the meta-analysis estimates relationships for each component domain while controlling for the presence of the other component domains. For example, two related component domains are (1) reducing the ratio of students to teachers and (2) student placement. Both domains can involve providing instruction to small groups of students, but the way each is defined in the component taxonomy is different. Specifically, only the student placement domain requires grouping of students based on student assessment data. The meta-analysis disentangles the efficacy for each component domain by controlling for the other component domains. Therefore, the relationship for the reducing the ratio of students to teachers component domain reflects how much the effects of an intervention might change when adding a component delivered to small groups (of any size smaller than a classroom), but only small groups that are formed without intentional placement or consideration of student needs. As previously discussed, component domains are also implemented together with unmeasured component domains; the meta-analysis does not account for the presence of these unmeasured component domains or other features of interventions that are not included in the model, which could also influence the relationships.

Even though the meta-analysis model found that most early literacy interventions in this synthesis had positive effects, some component domains are more likely to be positively associated than others with intervention effects when controlling for the presence of other component domains. Based on the results of the meta-analysis, Figure 7 reports the size and direction of associations between component domains and the effects of interventions. For each component domain, the figure shows an estimated effect size that measures the size of the association between the component domain and intervention effects, 90-percent credible interval for the effect size, and probability that the effect size is positive. The effect sizes, shown as dots in Figure 7, measure the strength of the component domain's association with the intervention effects. Specifically, the effect size gives the average estimate of how much larger the effect of an intervention might be when adding the component domain. This estimate is only suggestive, however, because other, unmeasured features of interventions beside the component domain could be contributing to the effect size magnitudes. A negative effect size means that the component domain is negatively associated with intervention effects, not that the interventions with the component domain have negative impacts on student outcomes.

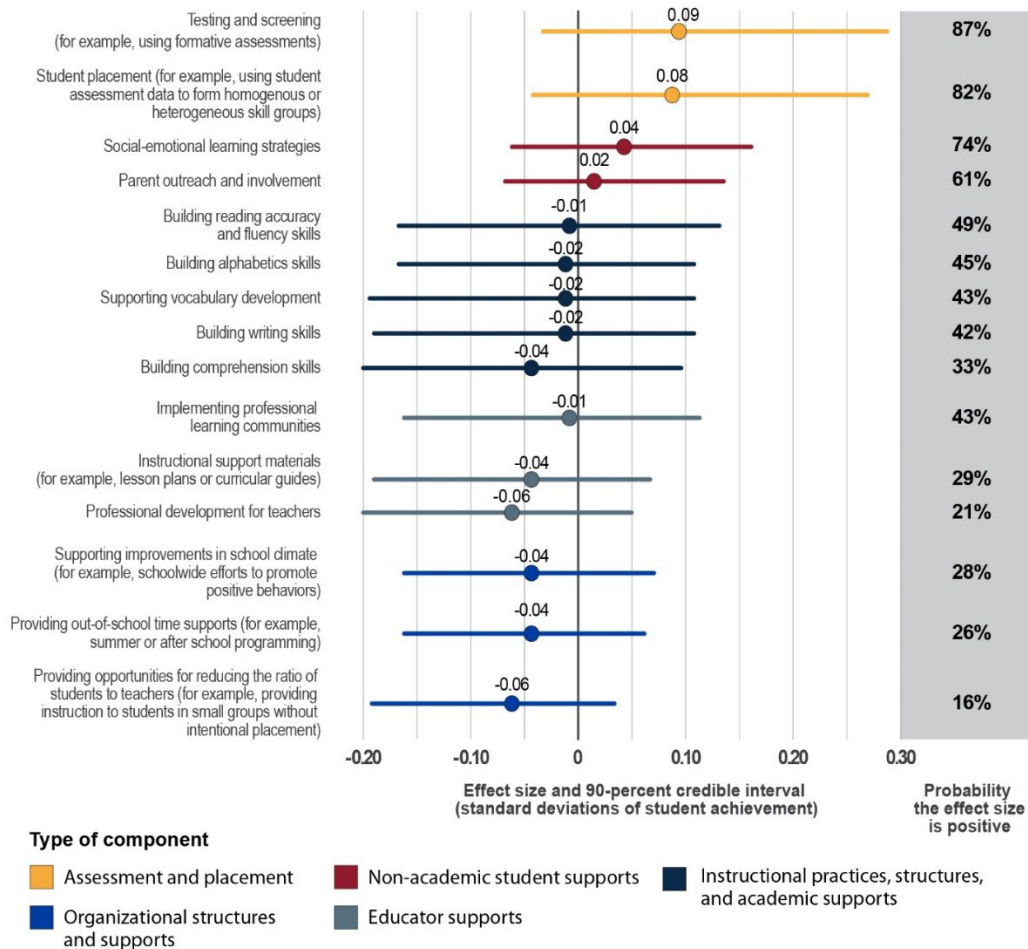
The 90-percent credible intervals, shown by the lines that extend on either side of each effect size, give the range of most likely effect sizes so that there is only a 10 percent chance that the effect size is either above or below the indicated range. The probability, shown in the column on the right, indicates the chance that the effect size is positive (see box on this page). Component domains with positive effect sizes are those with positive associations with intervention effects.

How to interpret estimated probabilities

The synthesis team selected the Bayesian methods in this synthesis with an aim to provide interpretable findings and conclusions about the relationships between component domains and improved literacy outcomes. In addition to effect sizes and credible intervals, the findings are reported in terms of probabilities.

A probability indicates what percentage of the time a statement will be true. For example, if the synthesis concluded that each of 10 components had a 70 percent chance of being related to improved outcomes, one should expect that seven of the 10 components are in fact related to improved outcomes. All research findings have some uncertainty about their conclusions. The intention with reporting these probabilities is to be transparent about the extent of uncertainty.

Figure 7. Magnitudes of relationships between each component domain and intervention impacts on independent alphabetic outcomes



Notes: Results are based on 100 findings from 29 studies of early literacy interventions that examined impacts on independent alphabetic outcomes. The purpose of the credible interval is to communicate uncertainty regarding the true effect size. The choice of a 90-percent interval (as opposed to a 95-percent or 80-percent interval, for example) is arbitrary. A negative effect size means that the component domain is negatively associated with intervention impacts, not that the interventions with the component domain have negative effects on student outcomes.

Source: Authors' calculations based on data from the WWC database of reviewed studies.

Assessment and placement domains. This application of the new methods found that testing and screening and student placement domains are each positively associated with intervention impacts on independent alphabetic outcomes. The association between the testing and screening domain and intervention impacts is 87 percent likely to be positive. The estimated effect size is 0.09 standard deviations, although there is a 10 percent chance that the actual relationship is below -0.03 or above 0.28 standard deviations. The association between the student placement domain and intervention impacts is 82 percent likely to be positive. The estimated effect size is 0.08 standard deviations, although there is a 10 percent chance that the actual relationship is below -0.04 or above 0.27 standard deviations.

Non-academic student supports domains. This application of the new methods also found that non-academic student supports are positively associated with intervention impacts. The association between the social-emotional learning strategies domain and intervention impacts is 74 percent likely to be positive, and the association for the parent outreach and involvement domain is 61 percent likely to be positive. The estimated effect size for the social-emotional learning strategies domain is 0.04 standard deviations, and the estimated effect size for the parent outreach and involvement domain is 0.02 standard deviations.

Other component domains. Because the remaining 11 component domains each have probabilities below 50 percent, this application of the new methods found that these component domains are negatively associated with intervention effects. Although some component domains are negatively associated with intervention effects, many interventions that included these components had positive effects on alphabetic outcomes (Figure 6). Therefore, this synthesis is not sufficient to support purposively excluding any of these component domains from interventions. The two component domains with the most negative associations with intervention effects according to Figure 7 are the professional development and reducing ratio of students to teachers domains:

- *Professional development domain.* This application of the new methods found that the association between the professional development for teachers domain and intervention impacts is only 21 percent likely to be positive with an estimated effect size of -0.06 standard deviations. This result is consistent with this component having been implemented more often in interventions that had negative effects than those with positive effects (Figure 6). However, this finding may be due to other features of the interventions not accounted for in the meta-analytic model rather than the professional development itself. For example, professional development might be offered in interventions that more often include other unmeasured components that are difficult to implement well, or in interventions where the school did not provide additional staff to implement the intervention. Alternatively, logistical challenges of implementing professional development could have placed additional demands on educator time resulting in less time for instruction. This synthesis cannot distinguish between these or other explanations. Even though professional development by itself may be insufficient to improve literacy outcomes, early literacy interventions included in this synthesis often included professional development (76 percent of studies as shown in Table 3). In some cases, professional development may have been an essential component of interventions.
- *Reducing ratio of students to teachers domain.* Similarly, this application of the new methods also found that the association between the reducing ratio of students to teachers domain and intervention impacts is only 16 percent likely to be positive with an estimated effect size of -0.06 standard deviations. This component domain is present in 52 percent of studies (Table 4). One possible explanation for the low ranking of this component domain is that the studies that include it also tend to include a component in the closely related student placement domain, which was highly ranked. This means that the association for reducing the ratio of students to teachers reflects the association between intervention impacts and providing small-group

instruction, but only when small groups are formed without intentional placement or consideration of student data. In other words, the student placement domain may be statistically cancelling out any positive effects associated with the reducing the ratio of students to teachers domain, but the results suggest that reducing the ratio of students to teachers is more likely to be associated with positive effects when student assessment data are used to place students into groups.

Finally, given that that the focus of the meta-analysis is on better understanding the effects of interventions on alphabetic outcomes, it might seem counter-intuitive that this application of the new methods found that building alphabetic skills domain is not positively associated with effects on these outcomes. The meta-analysis found that building alphabetic skills is only 45 percent likely to be associated with positive intervention effects on alphabetic outcomes. However, as reported in Table 1, all interventions that examined impacts on independent alphabetic outcomes included this component domain. Therefore, it is not possible for the meta-analysis to compare impacts between studies with and without the component domain because there are no studies with independent alphabetic outcomes that did not implement the building alphabetic skills domain. In this case, the Bayesian meta-analysis draws on associations estimated for studies that did not examine impacts on alphabetic outcomes, which were reported in additional 25 studies that examined impacts on other literacy outcomes. In other words, by extrapolating from evidence of intervention effects on outcomes in the other literacy domains, the model estimated that interventions that did not focus on building alphabetic skills were about 45 percent likely to improve independent alphabetic outcomes.

TO WHAT EXTENT CAN THE COMPONENT DOMAINS EXPLAIN VARIATION IN INTERVENTION EFFECTS?

This synthesis investigated to what extent the component domains explained the variation in the intervention impacts. This synthesis found that coded component domains explain only 9 percent of the variation in intervention impacts on alphabetic outcomes. The 9 percent of variation reflects variation explained in intervention impacts after accounting for sampling error due to small sample sizes, statistical bias in impact estimates that arises from the quality of the research design, and publication bias. Therefore, the component domains explain 9 percent of the variation in the true impacts of interventions, rather than the estimated impacts of interventions as reported in the studies. Figure D2 in [Appendix D](#) shows the degree to which the intervention impacts vary in ways the component domains in the meta-analytic model cannot explain.

However, the estimated 9 percent of explained variation may understate the importance of component domains, as noted in the limitations section above, because interventions are limited in their capacity to improve outcomes or because the poor measure quality obscures the importance of component domains. For example, if the amount of actionable variation in intervention impacts were only 20 percent, then the component domains explain a meaningful portion of the actionable variation.

Nevertheless, this finding from the application of the new methods means that 91 percent of the variation in intervention effects on alphabetic outcomes appears to be due to factors not examined in this synthesis. These factors might include (1) component domains that this synthesis did not measure, such as communicating goals to students or using specific types of instructional modeling; (2) contextual factors, such as who delivered the intervention, teacher quality, school climate, or school resources; (3) frequency, duration, or quality of implementation; (4) characteristics of measurement, such as the validity and reliability of outcome measures; and (5) unmeasured research design quality factors that could lead to statistical bias in some studies. Therefore, future research might examine to what extent these other factors contribute to explaining the variation in intervention impacts.

Finally, although the meta-analysis could not reliably distinguish among the 81 individual components within the 15 component domains, these individual components are potentially important for understanding why some interventions are more effective than others. As previously discussed, the meta-analysis is estimated at the level of component domain because the study sample size is too small to reliably differentiate such a large number of components. However, when including the 81 individual components in the meta-analytic model along with the 15 component domains, the amount of variation in intervention effectiveness that is explained by the model increases from 9 to 19 percent. It is possible that including additional levels of implementation detail, such as the component approaches listed in the component taxonomy, would lead to additional increases in variation explained. However, as discussed in the next section, it can be costly to collect this detailed information or it may not be reported in effectiveness studies. Furthermore, a larger sample of studies would be needed to measure precise associations between any additional factors and intervention effectiveness.

TAKE-AWAYS AND IMPLICATIONS

This report provides insights from conducting a Bayesian meta-analysis to explore the components of early literacy interventions and which components have the strongest associations with intervention impacts.

First, this synthesis identified challenges for meta-analysts attempting to use component taxonomies to identify *why* an intervention is effective and how effectiveness might be improved. The challenges included (1) developing a systematic and thorough approach to coding components of early literacy interventions that reflects how interventions might work; (2) not knowing exactly how, or how well, interventions and their components were implemented; and (3) disentangling in the meta-analysis the contributions of many individual components. Despite these challenges, identifying and measuring factors that contribute to intervention effectiveness merits continued research because understanding *why* interventions work is critically important to improving student outcomes.

The component taxonomy and associated coding protocol used in this synthesis are new, and this synthesis is the first to use the taxonomy in a meta-analysis. The component taxonomy should not be thought of as a definitive classification of the components of early literacy interventions and likely did not capture all relevant components and component domains. In future efforts to develop a component taxonomy and code components of interventions, researchers and content experts should think carefully about how interventions might work. These efforts will undoubtedly identify areas of disagreement that are valuable to explore and reconcile with further investigation. As such, the component taxonomy used in this study contributes to the ongoing conversation about what are key components of effective literacy interventions.

The results of this synthesis also suggest there are important unmeasured factors, beyond components and component domains, that influence intervention effectiveness. This synthesis did not address who delivered the intervention; teacher quality; school climate; school resources; or implementation frequency, duration, and quality. Past research suggests that getting implementation “right” is likely a crucial determinant for whether an intervention can be successful. For example, one meta-analysis provided evidence that implementation quality and organizational capacity are important factors associated with success of youth programs intended to address externalizing behavior issues (Wilson et al., 2020). Such findings could add important insights about best practices for intervention design and implementation (List, 2022).

Unfortunately, the effort to code information provided in original studies for this synthesis identified gaps that present a practical barrier to collecting more and better data about the implementation of interventions. It was not possible in this synthesis to code the presence or absence of components with complete accuracy, nor was it

possible to consistently code information about implementation challenges in the research. Most studies do not report this information. Coding this information is also time-intensive, especially when it is necessary to request additional information beyond that provided in the studies.

There is also a trade-off between coding more detailed information about interventions and their implementation and the number of studies needed to measure precise associations between each feature and intervention impacts. For example, contextual details, such as who delivered the intervention, could be collected and analyzed in the meta-analytic model. However, when including more variables in the model, it may be necessary to include more studies and findings in the analysis to obtain precise information about each feature's relationship with intervention effectiveness, especially when interventions are complex and bundle many components together. Given the small number of early literacy studies included in the analysis, this synthesis was unable to measure precise relationships for the 81 individual components within the 15 component domains. Also, the findings in this report provide relatively limited precision even when aiming to measure relationships between intervention impacts and only the 15 component domains. However, despite the limitations, the application of the taxonomy revealed some patterns in early literacy interventions and the Bayesian meta-analysis method identified some components that had positive associations with intervention impacts.

Second, this synthesis found that most early literacy interventions examined in this synthesis were multifaceted and had positive effects on alphabetic outcomes. Most interventions included in this synthesis contained multiple component types and component domains. Most required personnel beyond classroom teachers. And most were implemented with small groups or individual students. Although information about the costs of implementing the interventions is not typically reported in the studies, these costs—including resource costs, time, and staffing—likely vary widely. Seventy-two percent of the interventions had positive effects on alphabetic outcomes after accounting for sampling error due to small sample sizes, statistical bias in impact estimates that arises from the quality of the research design, and publication bias. However, this synthesis examined only early literacy interventions that have been rigorously studied by researchers and reviewed by the WWC. Therefore, findings from this synthesis may not represent all literacy interventions, nor generalize to all students or to those who are in the most need of reading interventions.

Finally, this synthesis highlights component domains that are positively associated with intervention effects on alphabetic outcomes, but these results come with caveats. The Bayesian meta-analysis method found that interventions that emphasize student assessment and placement and non-academic student supports were positively associated with intervention effects on alphabetic outcomes, but there is much more to be learned. The meta-analytic model accounts for some issues that can sometimes lead to misleading results in other meta-analyses, such as poor study quality and publication bias, and the Bayesian approach provides an opportunity to explore ways to report research findings in ways that may be more accessible to broad audiences. However, the approach is exploratory and caution in interpreting these findings is warranted because this synthesis cannot conclude that specific component domains *caused* improved alphabetic outcomes. These correlational findings based on the new method are still useful for researchers, and highlight which component domains may be more promising than others in improving alphabetic outcomes.

More research is needed to rigorously test the effectiveness of individual components and component domains in improving literacy outcomes to provide actionable recommendations to educators and intervention innovators before they can potentially develop and implement more effective early literacy interventions. It is not enough to report which component domains have the strongest associations with intervention impacts.

Decision makers must also carefully weigh the potential benefits of implementing these components with the costs of implementation. For example, high-cost interventions may not be feasible in some districts and schools. Therefore, detailed information about costs and implementation would be needed. Additionally, guidance from experts in early literacy is critical to design potentially more effective interventions.

It is possible the information available to meta-analysts will improve over time as study authors are more careful to collect and document details on implementation. The principles outlined in IES's [SEER](#) standards, which ask authors of effectiveness studies to provide more details about implementation, may contribute to a better understanding about what makes educational interventions effective. Additionally, guidance from experts, including educators and those with relevant lived experiences, can improve the component taxonomy and help fill in the gaps in the information available in research. These experts can also help to interpret the findings from meta-analyses that incorporate this information to form recommendations and provide important guidance on implementation.

WWC 2023008

September 2023

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-16-R-0011 by Mathematica. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. Any references within the document to specific education products are illustrative and do not imply endorsement of these products to the exclusion of other products that are not referenced.

This report is in the public domain. Although permission to reprint this publication is not necessary, it should be cited as:

Walsh, E., Deke, J., Robles, S., Streke, A., and Thal, D. (2023). *Using Bayesian meta-analysis to explore the components of early literacy interventions* (WWC 2023008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://whatworks.ed.gov>.

This report is available on the What Works Clearinghouse website at <https://whatworks.ed.gov> and on the Institute of Education Sciences website at <http://ies.ed.gov/ncee>.

STUDIES INCLUDED IN THE META-ANALYSIS

[Amendum, S. J., Vernon-Feagans, L., & Ginsberg, M. C. \(2011\)](https://eric.ed.gov/?id=EJ963705). The effectiveness of a technologically facilitated classroom-based early reading intervention. *Elementary School Journal*, 112(1), 107-131.

<https://eric.ed.gov/?id=EJ963705>.

[Baker, D. L., Basaraba, D. L., Smolkowski, K., Conry, J., Hautala, J., Richardson, U., English, S., & Cole, R. \(2017\)](https://eric.ed.gov/?id=ED588912). Exploring the cross-linguistic transfer of reading skills in Spanish to English in the context of a computer adaptive reading intervention [Grantee submission]. <https://eric.ed.gov/?id=ED588912>.

[Baker, D. L., Burns, D., Kame'enui, E. J., Smolkowski, K., & Baker, S. K. \(2015\)](https://eric.ed.gov/?id=ED588218). Does supplemental instruction support the transition from Spanish to English reading instruction for first-grade English Learners at risk of reading difficulties? [Grantee submission]. <https://eric.ed.gov/?id=ED588218>.

[Baker, S. K., Santoro, L. E., Chard, D. J., Fien, H., Park, Y., & Otterstedt, J. \(2013\)](https://eric.ed.gov/?id=EJ1013951). An evaluation of an explicit read aloud intervention taught in whole-classroom formats in first grade. *Elementary School Journal*, 113(3), 331-358. <https://eric.ed.gov/?id=EJ1013951>.

[Borman, T. H., Borman, G. D., Houghton, S., Park, S. J., Zhu, B., Martin, A., & Wilkinson-Flicker, S. \(2019\)](https://eric.ed.gov/?id=EJ1229779). Addressing literacy needs of struggling Spanish-speaking first graders: First-year results from a national randomized controlled trial of Descubriendo La Lectura. *AERA Open*, 5(3). <https://eric.ed.gov/?id=EJ1229779>.

[Buckingham, J., Wheldall, K., & Beaman, R. \(2012\)](https://eric.ed.gov/?id=EJ985895). A randomised control trial of a tier-2 small-group intervention (MiniLit) for young struggling readers. *Australian Journal of Learning Difficulties*, 17(2), 79-99.

<https://eric.ed.gov/?id=EJ985895>.

[Cann, D., Karakaplan, M., Lubke, M., & Rowland, C. \(2015\)](https://ccpi.unm.edu/sites/default/files/publications/EvaluatorReport.pdf). *New Mexico StartSmart K-3 Plus validation study* (Evaluator's report). Utah State University.

<https://ccpi.unm.edu/sites/default/files/publications/EvaluatorReport.pdf>.

[Case, L., Speece, D., Silverman, R., Schatschneider, C., Montanaro, E., & Ritchey, K. \(2014\)](https://eric.ed.gov/?id=EJ1030356). Immediate and long-term effects of tier 2 reading instruction for first-grade students with a high probability of reading failure. *Journal of Research on Educational Effectiveness*, 7(1), 28-53. <https://eric.ed.gov/?id=EJ1030356>.

[Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P. C., Karanzalis, M., & Gifford, R. \(2011\)](https://eric.ed.gov/?id=EJ963698). Small-group computer-assisted tutoring to improve reading outcomes for struggling first and second graders. *Elementary School Journal*, 111(4), 625-640. <https://eric.ed.gov/?id=EJ963698>.

[Cheatham, J. P., Allor, J. H., & Roberts, J. K. \(2014\)](https://eric.ed.gov/?id=EJ1021050). How does independent practice of multiple-criteria text influence the reading performance and development of second graders? *Learning Disability Quarterly*, 37(1), 3-14.

<https://eric.ed.gov/?id=EJ1021050>.

[Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. \(2013\)](https://doi.org/10.1177/0956797613509000). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students reading from first through third grade. *Psychological Science*, 24(8), 1408-1419.

[Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P., Bayraktar, A., Crowe, E. C., & Schatschneider, C. \(2011\).](https://eric.ed.gov/?id=EJ935111) Classroom instruction, child X instruction interactions and the impact of differentiating student instruction on third graders reading comprehension. *Reading Research Quarterly*, 46(3), 189-221. <https://eric.ed.gov/?id=EJ935111>.

[Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. \(2011\).](https://eric.ed.gov/?id=EJ932552) Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders word reading achievement. *Journal of Research on Educational Effectiveness*, 4(3), 173-207. <https://eric.ed.gov/?id=EJ932552>.

[Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., Simmons, L., Hagan-Burke, S., & Civetelli, C. \(2013\).](https://eric.ed.gov/?id=EJ994735) Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, 6(1), 1-23. <https://eric.ed.gov/?id=EJ994735>.

[Coyne, M. D., McCoach, D. B., Ware, S., Austin, C. R., Loftus-Rattan, S. M., & Baker, D. L. \(2018\).](https://eric.ed.gov/?id=EJ1202483) Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children*, 1-16. <https://eric.ed.gov/?id=EJ1202483>.

[Denton, C. A., Tolar, T. D., Fletcher, J. M., Barth, A. E., Vaughn, S., & Francis, D. J. \(2013\).](https://eric.ed.gov/?id=EJ1054506) Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology*, 105(3), 633-648. <https://eric.ed.gov/?id=EJ1054506>.

[Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., Zagar, E., Wood, T., & Petscher, Y. \(2017\).](https://eric.ed.gov/?id=EJ109320) Acquiring science and social studies knowledge in kindergarten through fourth grade: Conceptualization, design, implementation, and efficacy testing of content-area literacy instruction (CALI). *Journal of Educational Psychology*, 109(3), 301-320.

[Duff, F. J., Hayiou-Thomas, M. E., & Hulme, C. \(2012\).](https://eric.ed.gov/?id=EJ956317) Evaluating the effectiveness of a phonologically based reading intervention for struggling readers with varying language profiles. *Reading and Writing*, 25(3), 621-640. <https://eric.ed.gov/?id=EJ956317>.

[Duff, F. J., Hulme, C., Grainger, K., Hardwick, S. J., Miles, J. N. V., & Snowling, M. J. \(2014\).](https://eric.ed.gov/?id=EJ1042631) Reading and language intervention for children at risk of dyslexia: A randomised controlled trial. *Journal of Child Psychology and Psychiatry*, 55(11), 1234-1243. <https://eric.ed.gov/?id=EJ1042631>.

[Fuchs, D., Hendricks, E., Walsh, M. E., Fuchs, L. S., Gilbert, J. K., Tracy, W. Z., Patton, S. III, Davis, N., Kim, W., Elleman, A. M., & Peng, P. \(2018\).](https://eric.ed.gov/?id=EJ1168541) Evaluating a multidimensional reading comprehension program and reconsidering the lowly reputation of tests of near transfer. *Learning Disabilities Research & Practice*, 33(1), 11-23. [Study 1: Reading comprehension. Study 2: Reading comprehension plus working memory training.] <https://eric.ed.gov/?id=EJ1168541>.

[Gilbert, J. K., Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L. A., & Cho, E. \(2013\).](https://eric.ed.gov/?id=EJ997860) Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly*, 48(2), 135-154. <https://eric.ed.gov/?id=EJ997860>.

[Gunn, B., Smolkowski, K., & Vadasy, P. \(2011\).](https://eric.ed.gov/?id=EJ911535) Evaluating the effectiveness of Read Well Kindergarten. *Journal of Research on Educational Effectiveness*, 4(1), 53-86. <https://eric.ed.gov/?id=EJ911535>.

[Hagans, K., & Good, R. \(2013\).](https://eric.ed.gov/?id=EJ1040626) Decreasing reading differences in children from disadvantaged backgrounds: The effects of an early literacy intervention. *Contemporary School Psychology*, 17(1), 103-117. <https://eric.ed.gov/?id=EJ1040626>.

[Hill, D. V., & Lenard, M. A. \(2016\).](https://eric.ed.gov/?id=ED581879) *The impact of Achieve3000 on elementary literacy outcomes: Randomized control trial evidence, 2013-14 to 2014-15* (DRA Report No. 16.02). Wake County Public School System, Data and Accountability Department. <https://eric.ed.gov/?id=ED581879>.

[Hooper, S. R., Costa, L. C., McBee, M., Anderson, K. L., Yerby, D. C., Childress, A., & Knuth, S. B. \(2013\).](https://eric.ed.gov/?id=EJ995670) A written language intervention for at-risk second grade students: A randomized controlled trial of the process assessment of the learner lesson plans in a tier 2 response-to-intervention (RtI) model. *Annals of Dyslexia*, 66(1), 44-64. <https://eric.ed.gov/?id=EJ995670>.

[Jacob, R. T., Armstrong, C., & Willard, J. A. \(2015\).](https://eric.ed.gov/?id=ED562165) *Mobilizing volunteer tutors to improve student literacy: Implementation, impacts, and costs of the Reading Partners program*. MDRC. <https://eric.ed.gov/?id=ED562165>.

[Johnson, L., Terry, N. P., Connor, C. M. D., & Thomas-Tate, S. \(2017\).](https://eric.ed.gov/?id=EJ1156747) The effects of dialect awareness instruction on non-mainstream American English speakers. [Study 1a: DAWS.] *Reading and Writing*, 30(9), 2009-2038. <https://eric.ed.gov/?id=EJ1156747>.

[Jones, C. \(2014\).](https://uwm.edu/education/evaluation-of-the-milwaukee-community-literacy-project-spark-program/) *Evaluation of the Milwaukee Community Literacy Project/SPARK Program: Interim findings from the second cohort*. University of Wisconsin-Milwaukee. <https://uwm.edu/education/evaluation-of-the-milwaukee-community-literacy-project-spark-program/>.

[Jones, C. \(2015\).](https://eric.ed.gov/?id=ED567484) *The results of a randomized control trial evaluation of the SPARK literacy program*. Socially Responsible Evaluation in Education (SREED), University of Wisconsin-Milwaukee. <https://eric.ed.gov/?id=ED567484>.

[Jones, C., & Lander, R. \(2014\).](https://uwm.edu/education/evaluation-of-the-milwaukee-community-literacy-project-spark-program/) *Evaluation of the Milwaukee Community Literacy Project/SPARK Program: Findings from the first cohort*. University of Wisconsin-Milwaukee. <https://uwm.edu/education/evaluation-of-the-milwaukee-community-literacy-project-spark-program/>.

[Kloos, H., Sliemers, S., Cartwright, M. D., Mano, O., & Stage, S. \(2019\).](https://eric.ed.gov/?id=ED596283) MindPlay Virtual Reading Coach: Does it affect reading fluency in elementary school? In *Frontiers in Education* (Vol. 4, p. 67). Frontiers. <https://eric.ed.gov/?id=ED596283>.

[Konstantopoulos, S. \(2018\).](https://eric.ed.gov/?id=ED582923) *HEROES i3 development project: External evaluation report*.

[Kuchle, L., Brown, S., & Coukoulis, N. \(2018\).](https://eric.ed.gov/?id=ED582923) *Annual evaluation report for the Pennsylvania Dyslexia Screening and Early Literacy Intervention Pilot Program: Pilot year 2, 2016-17 school year*. American Institutes for Research. <https://eric.ed.gov/?id=ED582923>.

[Little, M. E., Rawlinson, D., Simmons, D. C., Kim, M., Kwok, O., Hagan-Burke, S., Simmons, L. E., Fogarty, M., Oslund, E., & Coyne, M. D. \(2012\).](https://eric.ed.gov/?id=EJ986226) A comparison of responsive interventions on kindergarteners early reading achievement. *Learning Disabilities Research and Practice*, 27(4), 189-202. <https://eric.ed.gov/?id=EJ986226>.

[Madden, N. A., & Slavin, R. E. \(2017\).](https://eric.ed.gov/?id=EJ1155720) Evaluations of technology-assisted small-group tutoring for struggling readers. *Reading & Writing Quarterly*, 1-8 (Study 1). <https://eric.ed.gov/?id=EJ1155720>.

- [May, H., Gray, A., Gillespie, J. N., Sirinides, P., Sam, C., Goldsworthy, H., Armijo, M., & Tognatta, N. \(2013\).](https://eric.ed.gov/?id=ED547669) *Evaluation of the i3 scale-up of Reading Recovery year one report, 2011-12.* Consortium for Policy Research in Education. <https://eric.ed.gov/?id=ED547669>.
- [May, H., Sirinides, P., Gray, A., & Goldsworthy, H. \(2016\).](https://eric.ed.gov/?id=ED593261) *Reading Recovery: An evaluation of the four-year i3 scale-up.* Consortium for Policy Research in Education. <https://eric.ed.gov/?id=ED593261>.
- [Nakamoto, I., Sobolew-Shubin, S., & Orland, M. \(2015\).](https://eric.ed.gov/?id=ED560945) *The Beaverton School District Arts for Learning (A4L) lessons project: An Investing in Innovation (i3) development grant.* WestEd. <https://eric.ed.gov/?id=ED560945>.
- [Quint, J., Zhu, P., Balu, R., Rappaport, S., & DeLaurentis, M. \(2015\).](https://eric.ed.gov/?id=ED579089) *Scaling up the Success for All model of school reform: Final report from the Investing in Innovation (i3) evaluation.* MDRC. <https://eric.ed.gov/?id=ED579089>.
- [Ransford-Kaldon, C., Ross, C., Lee, C., Sutton Flynt, E., Franceschini, L., & Zoblotsky, T. \(2013\).](https://authorzilla.com/OJrN/efficacy-of-the-leveled-literacy-intervention-system-report-2.html) *Efficacy of the Leveled Literacy Intervention System for K-2 urban students: An empirical evaluation of LLI in Denver Public Schools.* Center for Research in Educational Policy, University of Memphis. <https://authorzilla.com/OJrN/efficacy-of-the-leveled-literacy-intervention-system-report-2.html>.
- [Ross, S. M., Laurenzano, M. J., & Madden, N. A. \(2017\).](http://www.successforall.org/wp-content/uploads/2018/05/Ross-Laurenzano-Madden-2017.pdf) *An evaluation of the Lightning Squad computer-assisted small group tutoring program on the reading achievement of disadvantaged students in grades 1-3.* Center for Research and Reform in Education, Johns Hopkins University. <http://www.successforall.org/wp-content/uploads/2018/05/Ross-Laurenzano-Madden-2017.pdf>.
- [Savage, R., Piquette, N., Deleveaux, G., Abrami, P. C., Wood, E., Sanghera-Sidhu, S., & Burgos, G. \(2013\).](https://eric.ed.gov/?id=EJ1007934) A (pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology*, 105(2), 310-328. <https://eric.ed.gov/?id=EJ1007934>.
- [Schechter, R., Macaruso, P., Kazakoff, E. R., & Brook, E. \(2015\).](https://eric.ed.gov/?id=EJ1082218) Exploration of a blended learning approach to reading instruction for low SES students in early elementary grades. *Computers in the Schools*, 32(3-4), 183-200. <https://eric.ed.gov/?id=EJ1082218>.
- [Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Oi-Man, K., Simmons, L., Johnson, C., Zou, Y., Taylor, A. B., Mcalenney, L., Ruby, M., & Crevecoeur, Y. C. \(2011\).](https://eric.ed.gov/?id=EJ918891) Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children*, 77(2), 207-228. <https://eric.ed.gov/?id=EJ918891>.
- [Slavin, R. E., Madden, N., Calderon, M., Chamberlain, A., & Hennessy, M. \(2011\).](https://eric.ed.gov/?id=EJ918010) Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, 33(1), 47-58. <https://eric.ed.gov/?id=EJ918010>.
- [Smith, J. L. M., Nelson, N. J., Smolkowski, K., Baker, S. K., Fein, H., & Kosty, D. \(2016\).](https://eric.ed.gov/?id=EJ1103958) Examining the efficacy of a multitiered intervention for at-risk readers in grade 1. *The Elementary School Journal*, 116(4). <https://eric.ed.gov/?id=EJ1103958>.
- [Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. \(2017\).](https://eric.ed.gov/?id=ED577026) Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal*, 50(1), 282S-306S. <https://eric.ed.gov/?id=ED577026>.

[Tracey, L., Chambers, B., Slavin, R. E., Madden, N. A., Cheung, A., & Hanley, P. \(2014\).](#) Success for All in England: Results from the third year of a national evaluation. *SAGE Open*, 4(3), 1-10. <https://doi.org/10.1177/2158244014547031>.

[Tse, L., & Nicholson, T. \(2014\).](#) The effect of phonics-enhanced Big Book reading on the language and literacy skills of six-year-old pupils of different reading ability attending lower SES schools. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01222>.

[Vadasy, P. F., & Sanders, E. A. \(2011\).](#) Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, 15(6), 471-497. <https://eric.ed.gov/?id=EJ945761>.

[Vernon-Feagans, L., Kainz, K., Hedrick, A., Ginsberg, M., & Amendum, S. \(2013\).](#) Live webcam coaching to help early elementary classroom teachers provide effective literacy instruction for struggling readers: The Targeted Reading Intervention. *Journal of Educational Psychology*, 105(4), 1175-1187. <https://eric.ed.gov/?id=EJ1054424>.

[Wood, C., Fitton, L., Petscher, Y., Rodriguez, E., Sunderman, G., & Lim, T. \(2018\).](#) The effect of e-book vocabulary instruction on Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, 61(8), 1945-1969. <https://eric.ed.gov/?id=ED588534>.

[Zvoch, K., & Stevens, J. J. \(2012\).](#) Summer school effects in a randomized field trial. *Early Childhood Research Quarterly*, 28(1), 24-32. <https://eric.ed.gov/?id=EJ1007876>.

OTHER REFERENCES

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. The MIT Press.
- Berry, D. A., Hochberg, Y. (1999). Bayesian perspectives on multiple comparisons. *Journal of Statistical Planning and Inference*, 82(1-2), 215-227. [https://doi.org/10.1016/S0378-3758\(99\)00044-0](https://doi.org/10.1016/S0378-3758(99)00044-0).
- Bloom, H. S., Hill, C. J., Black, A. B., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4), 289-328.
- Burchinal, M., Krowka, S., Newman-Gonchar, R., Jayanthi, M., Gersten, R., Wavell, S., Lyskawa, J., Haymond, K., Bierman, K., Gonzalez, J. E., McClelland, M. M., Nelson, K., Pentimonti, J., Purpura, D. J., Sachs, J., Sarama, J., Schlesinger-Devlin, E., Washington, J., & Rosen, E. (2022). *Preparing young children for school* (WWC 2022-009). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, 126(4), 1593-1660. <https://doi.org/10.1093/qje/qjr041>.
- Deke, J., Finucane, M., & Thal, D. (2022). *The BASIE (BAyesian Interpretation of Estimates) framework for interpreting findings from impact evaluations: A practical guide for education researchers*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/pubs/2022005/pdf/2022005.pdf>.
- Efron, B., Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, 236(5), 119-127. <https://www.jstor.org/stable/24954030>.
- Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE 2016-4008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuides>.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin D. B. (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511790942>.
- Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, 5(2), 189-211 <https://doi.org/10.1080/19345747.2011.618213>.
- Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Education and Behavioral Statistics*, 40(5), 530-543. <https://doi.org/10.3102/1076998615606113>.

- Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457-472. <https://doi.org/10.1214/ss/1177011136>.
- Gleason, P., Crissey, S., Chojnacki, G., Zukiewicz, M., Silva, T., Costelloe, S., & O'Reilly, F. (2019). *Evaluation of support for using student data to inform teachers' instruction* (NCEE 2019-4008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NCEE20194008>.
- Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers* (NCEE 2012-4058). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuides>.
- Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., Kim, J., Lyskawa, J., Olson, C. B., & Smither Wulsin, C. (2016). *Teaching secondary students to write effectively* (NCEE 2017-4002). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuides>.
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, P., Goodman, S. N. & Altman, D.G. (2016). Statistical tests, *p*-values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, 31(4), 337-350. <https://doi.org/10.1007/s10654-016-0149-3>.
- Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuides>.
- Juárez, M. A., & Steel, M. F. J. (2010). Model-based clustering of non-Gaussian panel data based on skew-*t* distributions. *Journal of Business & Economic Statistics*, 28(1), 52-66. <https://doi.org/10.1198/jbes.2009.07145>.
- Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, 19(6), 339-345. <https://doi.org/10.1177/0963721410389459>.
- List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Penguin Randomhouse.
- Marks-Anglin, A., & Chen, Y. (2020). A historical review of publication bias. *Research Synthesis Methods*, 11(6), 725-742. <https://doi.org/10.1002/jrsm.1452>.
- Nation, K., & Snowling, M. (1997.) Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67(3), 359-370. <https://doi.org/10.1111/j.2044-8279.1997.tb01250.x>.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development. <https://eric.ed.gov/?id=ED444126>.
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40(2), 184-202. <https://doi.org/10.1598/RRO.40.2.3>.

Schanzenbach, D. W. (2006). What Have Researchers Learned from Project STAR? *Brookings Papers on Education Policy*, 205-228. <https://eric.ed.gov/?id=EJ896638>.

Scher, L. & Martinez, A. (2022). *Developing a core components nomenclature in education: An update on IES-funded work*. <https://ies.ed.gov/seer/pdf/CoreComponentsNomenclature.pdf>.

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, 79(4), 1391-1466. <https://doi.org/10.3102/0034654309341374>.

Snow, C. E., Burns, S. M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academy Press. <https://doi.org/10.1002/pits.10011>.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 197-206. <https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Third-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Inadmissibility-of-the-Usual-Estimator-for-the-Mean-of-a-bsmsp/1200501656>.

The Nation's Report Card. (2023). NAEP Report Card: Reading: National Achievement-Level Results. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. <https://www.nationsreportcard.gov/reading/nation/achievement/?grade=4>.

Vaughn, S., Gersten, R., Dimino, J., Taylor, M. J., Newman-Gonchar, R., Krowka, S., Kieffer, M. J., McKeown, M., Reed, D., Sanchez, M., St. Martin, K., Wexler, J., Morgan, S., Yañez, A., & Jayanthi, M. (2022). *Providing reading interventions for students in grades 4-9* (WWC 2022-007). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuides>.

Walsh, E., Deke, J., Robles, S., Streke, A., and Thal, D. (2023). *Public dataset for "Exploring the components of effective early literacy interventions."* National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://whatworks.ed.gov/>.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician*, 70(2), 129-133.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/Handbooks>.

Wilson, S. J., Lipsey, M. W., Aloe, A., & Sahni, S. (2020). *Developing evidence-based practice guidelines for youth programs: Technical report on the core components of interventions that address externalizing behavior problems*. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. https://youth.gov/sites/default/files/Technical-Report_Developing-EBP-Guidelines.pdf.

Wolf, B., & Harbatkin, E. (2023) Making sense of effect sizes: Systematic differences in intervention effect sizes by outcome measure type. *Journal of Research on Educational Effectiveness*, 16(1), 134-161. <https://doi.org/10.1080/19345747.2022.2071364>.

Wrulich, M., Brunner, M., Stadler, G., Schalke, D., Keller, U., & Martin, R. (2014). Forty years on: Childhood intelligence predicts health in middle adulthood. *Health Psychology, 33*(3), 292-296.

<https://doi.org/10.1037/a0030727>.

Zurovac, J., Cook, T. D., Deke, J., Finucane, M. M., Chaplin, D., Coopersmith, J. S., Barna, M., & Forrow, L. V. (2021). *Absolute and relative bias in eight common observational study designs: Evidence from a meta-analysis*. arXiv.

<https://arxiv.org/abs/2111.06941v2>.