IES Institute of Education Sciences

Effectiveness of Early Literacy Instruction: Summary of 20 Years of Research

Regional Educational Laboratory Southeast

At Florida State University

REL 2021–084
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation and Regional Assistance at IES

Effectiveness of Early Literacy Instruction: Summary of 20 Years of Research

Sarah Herrera, Beth M. Phillips, Yi-Chieh Newton, Jennifer L. Dombek, and James A. Hernandez July 2021

Children entering kindergarten vary greatly in their language and literacy skills. Therefore, up-to-date information about evidence-based practices is essential for early childhood educators and policymakers as they support preschool children's language and literacy development. This study used a process modeled after the What Works Clearinghouse (WWC) methodology to systematically identify effective early childhood curricula, lesson packages, instructional practices, and technology programs in studies conducted from 1997 to 2017. More than 74,000 studies were analyzed to identify interventions that improved students' performance in six language and literacy domains (language, phonological awareness, print knowledge, decoding, early writing, and general literacy). The study team identified 132 interventions evaluated by 109 studies that the study team determined were high-quality experimental or quasi-experimental studies. The WWC's evidence standards are used to assess the quality of an evaluation study and the strength of its claims about whether an intervention caused the observed effect on student achievement. To better understand the effectiveness of the interventions, their implementation characteristics and instructional features were coded for the relevant language and literacy domains. The findings revealed that instruction that teaches a specific domain is likely to increase performance in that domain. Interventions that teach language exclusively might be more beneficial when conducted in small groups or one-on-one than in larger group sizes. In addition, teaching both phonological awareness and print knowledge might benefit performance in print knowledge. Finally, some evidence indicates that instruction that teaches both phonological awareness and print knowledge might also lead to improvements in decoding and early writing performance.

Why this study?

Children entering kindergarten vary greatly in their language and literacy skills (for example, Denton et al., 2009; Reardon & Portilla, 2016). Gaps in school readiness, which appear well before kindergarten entry (Burchinal et al., 2010), are associated with difficulties in achieving grade-level reading proficiency later. Although some children catch up to their peers, others fall even further behind as schooling progresses (Dale et al., 2014; Reynolds & Fish, 2010). Despite the expansion of publicly funded prekindergarten in recent decades, some programs provide poor-quality instruction, particularly in early language development (Neuman & Dwyer, 2009; Phillips et al., 2018). This is due in part to their use of curricula, lesson packages, instructional methods, and technology programs that are not empirically supported (Moiduddin et al., 2012), perhaps because educators and policymakers lack knowledge of the most effective practices (Piasta et al., 2017). Early childhood educators and policymakers could therefore benefit from up-to-date information about evidence-based practices that support language and literacy development in preschool children.

The National Early Literacy Panel (NELP) and the What Works Clearinghouse (WWC) provide quality information about evidence-based practices in early childhood education, but as of the writing of this report in 2020, neither has comprehensive fully up-to-date information that encompasses all the latest research. A 2008 NELP report identified the best available evidence about early predictors and instructional practices that improved language and literacy performance in preschool and kindergarten (National Early Literacy Panel, 2008). However, it included only published peer-reviewed research produced before

For additional information, including background on the study, technical methods, and supporting analyses, access the report appendixes at https://go.usa.gov/x6trG.

2004. The WWC provides standards for evaluating the rigor of research design and produces intervention reports that provide the highest quality evidence for specific (and often commercially available) early childhood curricula, programs, practices, and policies. Yet, as with the NELP, some of the WWC intervention reports that focus on early childhood education are more than 10 years old. There is now over a decade of intervention research that likely expands and sharpens understanding of effective curricula, lesson packages, instructional practices, and technology programs for early language and literacy development.

Language, phonological awareness, print knowledge, decoding, and early writing skills are important early predictors of later language and literacy development (Furnes & Samuelsson, 2009; Melby-Lervåg et al., 2012; National Early Literacy Panel, 2008). And instructional practices that teach certain combinations of these skills might lead to greater improvements in performance on taught and untaught skills (National Early Literacy Panel, 2008). Accordingly, it is worth examining what recent research indicates about whether and how teaching language and literacy skills (individually or in combination) impact taught and untaught skills. For example, teaching both phonological awareness and print knowledge might support children's development in one or both domains more than teaching just one, given that performance in these domains is known to be highly related (Kim et al., 2010; Lerner & Lonigan, 2016).

It is also important to explore the impacts of instruction in these domains when assessed both by researcher-developed outcome measures, which typically are very similar to the specific content being taught, and by standardized outcome measures, which typically capture a broader, less specific representation of the skill area. Researcher-developed and standardized outcome measures provide complementary ways of understanding how children can benefit from instruction in these important skill areas.

The purpose of this review is to update the evidence on early literacy interventions by evaluating the past 20 years of published peer-reviewed research and other available research sources. Its goal is to identify effective commercially available and researcher-developed interventions and to identify the specific instructional domains and features, and implementation characteristics that lead to improvements in language and literacy performance. The aim is to provide early childhood educators and policymakers a single up-to-date resource they can use to make curricular choices for use in state-supported prekindergarten programs and other agencies, increase knowledge regarding evidence-based practices for kindergarten readiness, and inform professional development efforts. Box 1 defines the outcomes explored in this review, box 2 defines other key terms, and box 3 summarizes data sources, the sample, and the methods used to describe study findings.

Research questions

The review addresses one primary research question and several related subquestions to better understand the nature of the interventions studied:

- What rigorous evidence exists that early literacy curricula, lesson packages, instructional practices, and technology programs effectively improve students' language, phonological awareness, print knowledge, decoding, early writing, or general literacy performance?
  - Which instructional domains are taught in the studied interventions?
  - How do implementation characteristics (intervention type, intervention duration, implementer type, group size, and the presence of professional development with or without ongoing support) vary among the studied interventions?
  - How effective are the studied interventions in promoting performance on taught and untaught outcome domains?
  - How do effects differ between researcher-developed outcome measures that assess skills similar to those taught and standardized outcome measures that assess broader skills?
  - Which instructional features effectively promote early literacy performance in each outcome domain?

Box 1. Language and literacy domains of interest

The review explored six language and literacy domains: language, phonological awareness, print knowledge, decoding, early writing, and general literacy. The term *instructional domain* is used when discussing the domain that was taught, and the term *outcome domain* is used when discussing the effects of an intervention.

- **Language.** The ability to comprehend or use spoken language, which can include vocabulary, listening comprehension, syntax, or narrative understanding and production.
- **Phonological awareness.** The awareness of the sound units of spoken language, such as phonemes, onset-rimes, syllables, or words. Phonological awareness tasks include producing rhyming words or words that share common sound units; segmenting larger units into smaller ones (for example, words into phonemes or words into syllables); and identifying, deleting, and blending the separate sounds of a word.
- **Print knowledge.** The knowledge of the names and sounds of the letters of the alphabet and the knowledge of concepts about print.
- **Decoding.** The ability to translate a word from print to speech, usually by understanding sound-symbol correspondences; also, the act of deciphering a new word by sounding it out.
- **Early writing.** The knowledge of letter or name writing, spelling, and conveying meaning through writing.
- **General literacy.** Outcome measure that combines two or more outcome domains (that is, language, phonological awareness, print knowledge, decoding, and early writing), provides a summary score across domains, or assesses kindergarten readiness.

Box 2. Key terms

Implementation characteristics. The implementation characteristics of interest included intervention type, intervention duration, implementer type, group size, and the presence of professional development with or without ongoing support. Implementation characteristics were coded for each intervention as a whole and therefore do not provide information about implementation for each instructional domain taught unless the intervention taught only one instructional domain.

Instructional features. The core components of an intervention, including its essential practices, its structural elements, and the contexts in which it was implemented and tested. Instructional features were coded by instructional domain to better understand the instructional content of each domain. For example, instructional features coded in the language domain include the occurrence of shared book reading with or without questions (see appendix A for details about the instructional features that were coded for each instructional domain).

Intervention. A curriculum, a lesson package, an instructional practice, or a technology program (see figure A1 in appendix A for details about these classifications):

- A **curriculum** is a set of activities, materials, or guidance for working with children and is the primary instructional tool or is designed as a supplement to the primary instructional tool.
- A **lesson package** has an identified name and includes lesson plans (it can combine two or more named interventions).
- An **instructional practice** is a specific teaching method that guides the instructional interaction with children.
- A **technology program** is a program that uses some form of technology (such as, a computer or audio player) to deliver instruction to students. An intervention was coded as a technology program when the intervention exclusively comprised a single or multiple technology programs.

Outcome measure. Two types of outcome measures are discussed. A researcher-developed outcome measure is an assessment developed by researchers that does not have norm-referenced scores and might not be commercially available. A standardized outcome measure is an assessment that has established administration and scoring procedures that are often documented in a technical manual and is often commercially available. Standardized outcome measures typically include a norm-referenced group to which the study sample is compared. Outcome measures that include a modification to a standardized outcome measure (for example, purposively selecting items) are considered researcher-developed in this report. Researcher-developed outcome measures are often more closely aligned to instruction, whereas standardized outcome measures often assess broader skills (Hill et al.,

2008; Marulis & Neuman, 2010, 2013; National Institute of Child Health and Human Development, 2000). Researcher-developed outcome measures are also often more sensitive to changes in student performance that are directly related to the instruction being studied than standardized outcome measures are (National Institute of Child Health and Human Development, 2000).

Use of What Works Clearinghouse (WWC) evidence standards. The evidence standards used to evaluate the quality of an evaluation study and assess the strength of its claims about whether an intervention caused the estimated impact on student achievement. Studies that are rated as *high-quality experimental studies*—comparable to studies meeting WWC evidence standards without reservations—are considered to provide the strongest empirical evidence on the effectiveness of an intervention on student achievement. Studies that are rated as *high-quality quasi-experimental studies*—comparable to studies meeting WWC evidence standards with reservations—provide a lower degree of empirical evidence on the effectiveness of an intervention on student achievement. Studies rated as high-quality experimental or quasi-experimental studies are collectively referred to as *high-quality impact studies* in this report. Studies that are not rated as *high-quality impact studies*—comparable to studies not meeting WWC evidence standards—are not able to provide causal evidence to support the effectiveness of an intervention on student achievement. Because the online study review guide for entering version 4.0 reviews into the WWC database as required by the Institute of Education Sciences (IES) for its contractors starting in late 2017 was not ready at the time of these reviews, the study team could not use an official WWC protocol for the reviews. Consequently, the reviews, while conducted using the latest WWC standards available at the time, were not entered into the database of official WWC reviews maintained by IES. Therefore, the studies discussed in this report cannot be described as meeting WWC evidence standards with or without reservations or as not meeting WWC evidence standards.

Weighted effect size. The magnitude of the effect of a characteristic or feature that is shared among a subset of the 132 interventions evaluated by studies that met the evidence standards (see below) discussed in this report. Effect sizes are weighted so that effects based on smaller sample sizes do not contribute as much to the calculated weighted effect size as effects based on larger sample sizes (see appendix A for the formulas used to estimate the weighted effect sizes). Interventions evaluated by studies using a single-case design were excluded from these calculations because effect size estimates were not calculated for these studies.

Note: Additional key terms are defined in box B1 in appendix B.

Box 3. Data sources, sample, and methods

Data sources. This review included data obtained from published (but not necessarily peer-reviewed) studies, identified as a result of a comprehensive literature search of experimental research evaluating the effectiveness of interventions (curricula, lesson packages, instructional practices, and technology programs) intended to improve language and literacy development. To be eligible, each study had to have:

- Been published between January 1, 1997, and December 31, 2017.
- Included a randomized controlled design, a quasi-experimental design, or a single-case design.
- Been conducted in a school or childcare center in the United States or a similar country (that is, in which English is the predominant language).
- Evaluated an intervention intended for children ages 36–71 months who were not yet in kindergarten, were primarily native English speakers, and were not eligible for special education services under Parts B and C of the Individuals with Disabilities Education Act (except for children with a language or speech impairment).
- Included an intervention delivered by a researcher or practitioner (for example, school- or center-based personnel, a speech-language pathologist, or a paraprofessional).
- Included an intervention in which 75 percent or more of instruction is delivered in English.
- Included a reliable and valid outcome in the language, phonological awareness, print knowledge, decoding, early writing, or general literacy domain that was not considered overaligned with the intervention.[1]

Sample. More than 74,000 unique studies were initially identified by the search procedures (see figure A2 in appendix A). Of these, 357 met eligibility criteria and were evaluated against version 4.0 of the What Works Clearinghouse (WWC) standards by WWC-certified reviewers (see appendix F for a list of these studies).[2] The study team determined that 109 of the 357 eligible

studies were high-quality impact studies, representing 132 interventions (see appendixes C–E for intervention descriptions). The studies included in the current report have limited overlap with the studies included in either National Early Literacy Panel (2008; 5 percent overlap) or evaluated by the WWC (17 percent overlap). This report includes only outcome measures and contrasts between intervention and comparison groups that met eligibility and evidence standards. For example, it does not include outcome measures that the study team considered overaligned. Furthermore, outcomes that did not meet the standards were excluded from the calculation of the average effect size for each domain and from determinations of intervention effectiveness.

Methodology. The study team analyzed the 132 interventions evaluated by 109 studies that it determined were high-quality impact studies in three ways:

1. *Evaluating statistical significance in each intervention for researcher-developed and standardized outcome measures by outcome domain.* The study team determined statistical significance based on the presence of at least one statistically significant effect size in each outcome domain and type of outcome measure (that is, researcher-developed and standardized) after adjusting for multiple comparisons using the Benjamini-Hochberg correction.[3] In addition, the team calculated an average effect size for each outcome domain and type of outcome measure in each intervention. Categories for outcome domains in interventions included effective, inconclusive, or not effective and were based on the statistical significance of the effects.[4]

2. *Calculating weighted effect sizes across interventions that share specific components to gauge their collective effect on each outcome domain.* Effect sizes are weighted so that effects based on smaller sample sizes do not contribute as much to the weighted effect size as effects based on larger sample sizes do (see appendix A for the formulas used to estimate the weighted effect sizes). In single-case design studies the effect sizes and statistical significance are not estimated; visual analysis is used instead to evaluate the effectiveness of an intervention. Single-case design studies are therefore not included in weighted effect size estimates.

3. *Examining the descriptions of the interventions.* The study team coded intervention descriptions according to a common set of codes identifying the instructional domain or domains, the specific materials and instructional features, the implementation characteristics of the individuals delivering the intervention, the setting, and the duration and intensity of the intervention. Members of the study team were trained to reliably code implementation characteristics and instructional features using a codebook (see appendix A).[5] The purpose of this coding was to look for patterns of characteristics or features that were associated with increased language and literacy performance.

Details pertaining to the literature search, eligibility criteria, screening, and review processes are in appendix A.

Notes

1. An outcome measure is considered overaligned if it contains content or materials provided to subjects in one group but not the other or others. Overaligned outcome measures might provide the intervention group with an unfair advantage over the comparison group, such that the effect size would not reflect a true representation of the intervention's effect. For example, an outcome measure comprising vocabulary words that only the children in the intervention group were exposed to would be considered overaligned with the intervention group. Therefore, only studies that include outcome measures not considered overaligned can meet the evidence standards.

2. Although this report relied heavily on version 4.0 of the WWC procedures and standards (What Works Clearinghouse, 2017a, 2017b) and reviews were conducted by WWC certified reviewers, this report is not a WWC product.

3. Significance was not determined using the p-value associated with the calculated average effect size among all outcomes in an outcome domain and type of outcome measure because this approach was considered too conservative and would diminish the effectiveness of interventions demonstrating statistically significant effects on individual outcomes. For example, if one of two researcher-developed outcome measures for print knowledge was statistically significant after a Benjamini-Hochberg correction, the print knowledge domain was identified as demonstrating positive effects even if the p-value of the calculated average effect size did not reach significance.

4. An effective outcome is one that shows a statistically significant positive effect on at least one outcome in an outcome domain and type of outcome measure after a Benjamini-Hochberg correction. An inconclusive outcome is one that does not demonstrate any statistically significant individual effects in an outcome domain after a Benjamini-Hochberg correction. A not effective outcome is one that shows a statistically significant negative effect on at least one outcome in an outcome domain and type of outcome measure after a Benjamini-Hochberg correction.

5. Importantly, no minimum amount of instruction in a particular instructional domain was required to apply the code. Therefore, interventions coded as including multiple instructional domains likely spent differing amounts of time on each coded instructional domain.

Findings

This section presents the findings across the 132 interventions evaluated by studies that the study team determined were high-quality impact studies. Appendix B provides technical results, and appendixes C–E provide detailed descriptions of each of the 132 studied interventions.

Rigorous evidence exists on effective early literacy interventions

Of the 132 interventions evaluated by high-quality impact studies, 38 demonstrated effectiveness in at least one language or literacy outcome domain relative to the comparison group. This means that the intervention group significantly outperformed the comparison group. Table C1 in appendix C provides a list of the interventions (denoted by a filled circle) by outcome domain that effectively improved performance.

All other studied interventions except one (93 of 132) demonstrated inconclusive effects. Inconclusive effects mean that language and literacy performance was considered statistically comparable between the intervention and comparison groups on all outcome domains assessed. Studied interventions demonstrating inconclusive effects need further investigation to better understand their effectiveness. In some cases studies of interventions demonstrating inconclusive effects might have demonstrated effectiveness if the study evaluating those interventions had included a larger sample size (Seftor, 2016).

Most interventions taught multiple instructional domains, and nearly all the rest taught the language domain exclusively

Of the 132 interventions evaluated by high-quality impact studies, 77 taught two or more instructional domains (see table B3 in appendix B). Of the 55 interventions that taught a single instructional domain, 50 taught language. Language was by far the most frequently taught instructional domain (113 of 132), followed by print knowledge (69) and phonological awareness (61). Few interventions taught early writing (30) or decoding (11). Among the interventions that taught two or more instructional domains, phonological awareness and print knowledge instruction co-occurred most frequently (52).

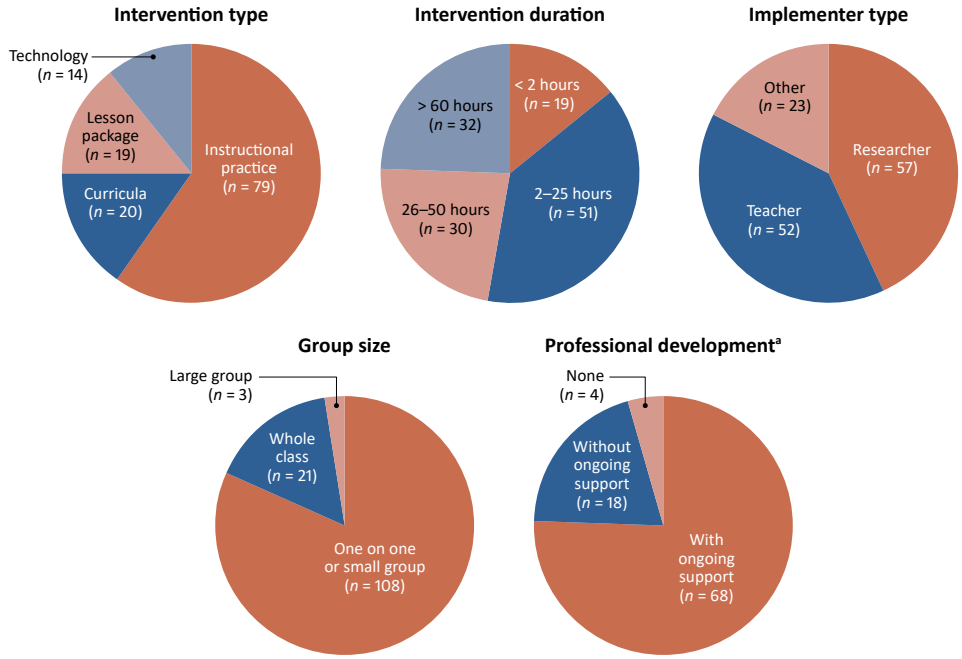
Implementation characteristics varied among the 132 interventions

The implementation characteristics varied across the 132 interventions evaluated by high-quality impact studies (figure 1). The implementation characteristics of interest included:

- Intervention type (curriculum, lesson package, instructional practice, or technology program).
- Intervention duration (less than 2 hours, 2–25 hours, 26–50 hours, or more than 60 hours).
- Implementer type (teacher, researcher, or other).
- Group size (one on one or small group, large group only, or whole class only).
- Professional development (with or without ongoing support or no professional development).

The number of each intervention type varied, with 79 of 132 classified as instructional practices. Total instructional time for the interventions studied ranged from 15 minutes to 400 hours, with the largest number of interventions lasting 2–25 hours. Researchers (57) and teachers (52) implemented a similar number of interventions, and the rest were implemented by other personnel (for example, a speech-language pathologist or a paraprofessional). Of the 132 interventions, 108 included at least some one-on-one or small-group instruction, and the rest were conducted exclusively in either whole-class (21) or large-group (3) configurations. Of the 90 interventions that reported information about professional development, 67 provided ongoing support.

Figure 1. Implementation characteristics varied among the 132 interventions evaluated by high-quality impact studies



a. Includes only the 90 interventions that provided sufficient information about professional development.

Source: Authors' compilation.

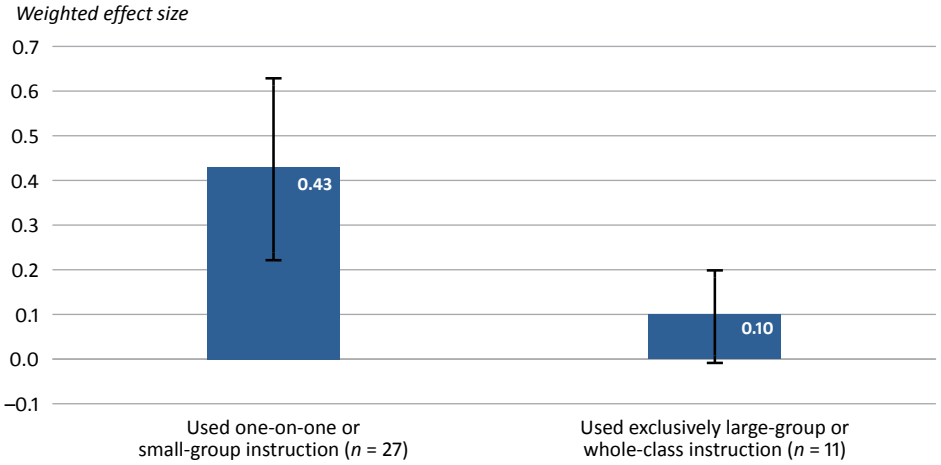
The study team could not code implementation characteristics by instructional domain because most intervention descriptions lacked sufficient detail about each instructional domain taught. Intervention descriptions often included information about teaching more than one instructional domain and more than one group size but did not specify which group size corresponded to which instructional domain. Similarly, intervention duration was often reported for the intervention as a whole and not for each instructional domain taught. As such, implementation characteristics were coded for each intervention as a whole and therefore do not provide information about implementation for each instructional domain taught unless the intervention taught only one instructional domain. As a result, implementation characteristics could be further explored for only the 38 interventions that taught language only, evaluated effectiveness on language outcomes, and included sufficient information to derive an effect size estimate (see figure B1 in appendix B).

Among interventions that taught language exclusively, some implementation characteristics were associated with significantly larger improvements in language performance. Among interventions that taught language exclusively, interventions classified as instructional practices produced a significantly larger weighted effect size on language performance than interventions classified as curricula (0.43 versus 0.02; see table B6 in appendix B). An effect size of 0.43 is equivalent to a 17 percentile point increase in language performance for an average student in the intervention group relative to an average student in the comparison group.

The weighted effect size on language performance was significantly larger when language instruction included some one-on-one or small-group implementation than when it used large-group or whole-class configurations exclusively (0.43 versus 0.10; figure 2). Furthermore, interventions that used large-group or whole-class configurations exclusively did not improve language performance relative to the comparison group.

These findings highlight the benefit of language-focused instructional practices and the importance of one-on-one and small-group language instruction for preschool students.

Figure 2. Among interventions that taught the language domain exclusively, those that included one-on-one or small-group configurations led to significantly better language performance than interventions that used large-group or whole-class configurations exclusively



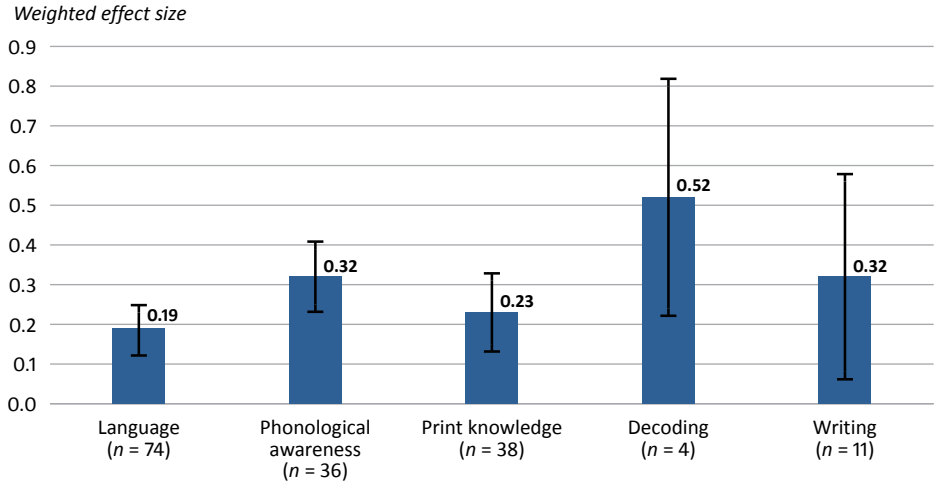
Note: The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. If there is no overlap between the 95 percent confidence intervals, the difference between the weighted effect sizes is considered statistically significant.

Source: Authors' analysis of primary data collected for the review; see appendix E.

Early literacy interventions improved language and literacy performance in taught domains

Results from the 112 interventions evaluated by high-quality impact studies and included at least one effect size estimate indicate that for the five domains explored in this review, instruction in a domain often improved outcomes in that domain (figure 3; see also figure B2 in appendix B). For the language, phonological awareness, and decoding domains, interventions that taught the domain exclusively or that taught the domain in combination with other

Figure 3. Interventions that taught language, phonological awareness, print knowledge, decoding, or early writing were likely to improve performance in the taught domain



Note: The weighted effect size represents the overall performance (averaging researcher-developed and standardized outcome measures) in each domain. The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. Includes only interventions evaluated in high-quality impact studies and included at least one effect size estimate (n = 112). Interventions could teach more than one domain.

Source: Authors' analysis of primary data collected for the review; see appendix E.

domains improved performance in that domain. For the print knowledge and early writing domains, interventions that taught either domain exclusively did not improve performance in the taught domain. In addition, interventions that taught a given domain did not consistently improve performance in untaught domains (see figure B3).

Language

This section first discusses the instructional features of the 86 interventions that taught language and were evaluated in studies that included language outcomes. It then examines the effect of teaching language on language outcomes and explores the effects of instructional features in a subset of 74 interventions evaluated in high-quality impact studies and included at least one effect size estimate (see figure B4 in appendix B).

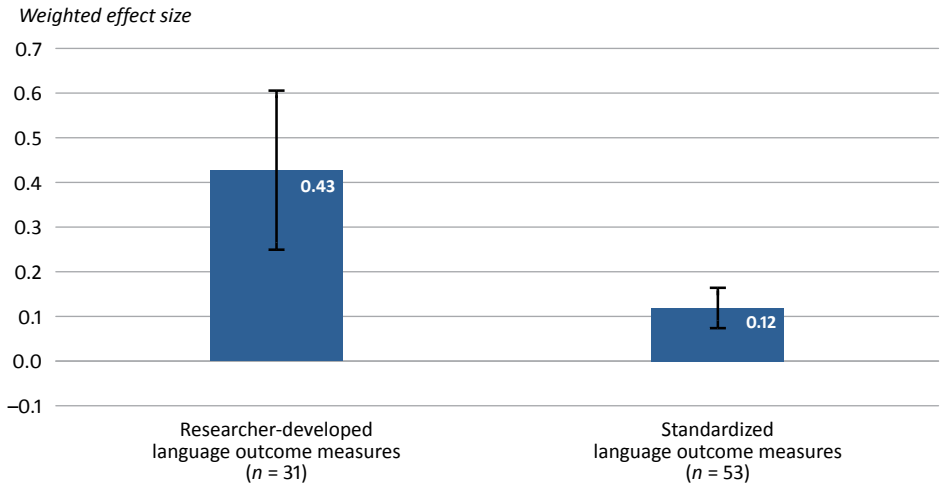
Of the 86 interventions that were evaluated in studies that taught language and included language outcomes, 67 included shared book reading (see figure B5 in appendix B). Implementers used a variety of books, including narrative, expository, and wordless picture books. Interventions that included shared book reading were more likely to have the implementer ask students questions (49 of 67) than to have just passive listening (18 of 67). Implementers who asked questions did so before, during, or after reading and incorporated a variety of question types, including asking students to connect story events to their own experiences, to recall information or story events, and to make inferences beyond the book.

Other components of language instruction that were coded were comprehension, vocabulary, extending language, morphology, speech production, and pragmatics (see appendix A for descriptions; see also figure B6 in appendix B). Of the 86 language interventions, 78 included instruction focused on vocabulary, and 71 included instruction on comprehension (which always co-occurred with vocabulary instruction). Less than half (32) focused on extending language, which almost always (29) co-occurred with vocabulary instruction. Very few interventions were devoted to speech production (4), pragmatics (3), or morphology (2).

Interventions that taught language improved language performance, especially when skills similar to those taught were assessed. Teaching language resulted in a statistically significant weighted effect size of 0.19 on language outcomes, among the 74 interventions evaluated in studies that included at least one effect size estimate (see figure 3 and table B7 in appendix B). That effect size is equivalent to an 8 percentile point increase in language performance for an average student in the intervention group relative to an average student in the comparison group. The weighted effect size was significantly larger for researcher-developed outcome measures than for standardized ones (0.43 versus 0.12; figure 4; see also table B7 in appendix B). This is likely because researcher-developed outcome measures often represent skills that are more similar to those being taught, whereas standardized outcome measures often represent broader skills (Hill et al., 2008; Marulis & Neuman, 2010, 2013; National Institute of Child Health and Human Development, 2000). These findings suggest that when practitioners are seeking interventions to use or purchase, they should keep in mind that some interventions are unlikely to yield sizable effects for standardized outcome measures.

All language instructional features resulted in comparable language performance. Several pairs of language instructional features were compared to identify which features were more effective at improving language performance. For example, language performance was compared between interventions that included shared book reading in which the implementer asked questions and interventions that included shared book reading without questions. Both types of shared book reading yielded significant and statistically comparable improvements in language performance (0.26 versus 0.09; see table B8 in appendix B). In addition, language interventions that included both comprehension and vocabulary were compared with interventions that did not include both features. Both intervention types yielded statistically comparable effect sizes (0.20 versus 0.14). However, interventions that included both features significantly improved language performance, whereas interventions that did not include both features did not significantly improve language performance. Although most of the language

Figure 4. Interventions that taught language improved performance on researcher-developed language outcomes, which often represent skills similar to those taught, more than they improved performance on standardized language outcomes, which often represent broader language skills



Note: The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. If there is no overlap between the 95 percent confidence intervals, the difference between the weighted effect sizes is considered statistically significant. Includes only interventions evaluated in high-quality impact studies and included at least one effect size estimate (n = 74). Ten studied interventions included both standardized and researcher-developed outcome measures.

Source: Authors' analysis of primary data collected for the review; see appendix E.

instructional features explored led to improvements in language performance, it remains unclear which features are most effective at improving performance.

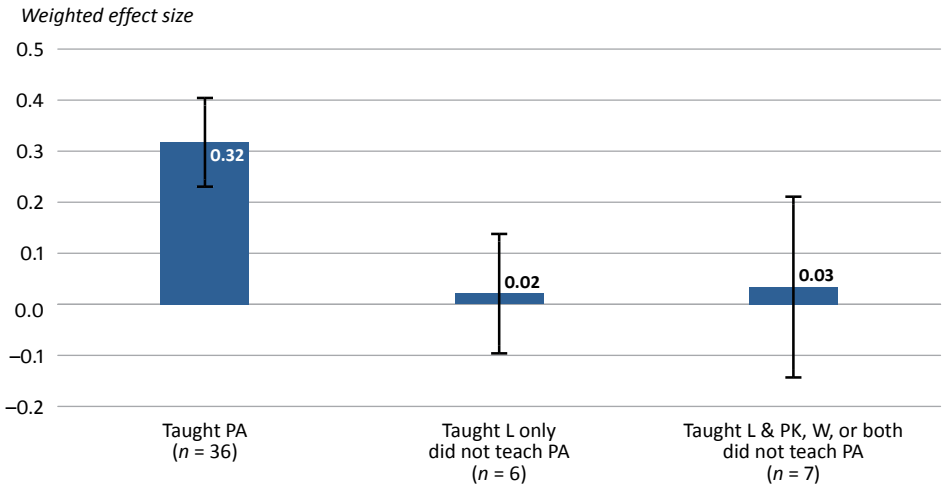
Phonological awareness

This section first discusses the instructional features of the 43 interventions that taught phonological awareness and were evaluated in studies that included phonological awareness outcomes. It then examines the effect of teaching phonological awareness on phonological awareness outcomes and explores the effects of instructional features in a subset of 36 interventions evaluated in high-quality impact studies and included at least one effect size estimate (see figure B7 in appendix B). Finally, it explores the effect on phonological awareness performance for 13 studied interventions that did not teach phonological awareness.

Of the 43 studied interventions that were evaluated in studies that taught phonological awareness and included phonological awareness outcomes, 29 included at least two of the following tasks: identification, matching, blending, counting, segmenting, or production (see figure B7 in appendix B). Interventions that included identification tasks asked students to identify the initial phoneme or rime unit in an orally presented word or in a word depicted in a picture. Interventions that included matching tasks asked students to match or sort words that shared a common rime unit or the initial phoneme. Blending tasks included practice with orally presented words, syllables, rime units, and phonemes. Segmenting tasks included practice breaking orally presented words into smaller words, syllables, and phonemes. All the interventions that included production tasks asked students to produce words that rhymed, and most also asked students to produce words that shared the first phoneme.

Interventions that taught phonological awareness improved phonological awareness performance. Teaching phonological awareness resulted in a significant weighted effect size of 0.32 on phonological awareness outcomes, among the 36 interventions evaluated in studies that included at least one effect size estimate (see figure 3). That effect size is equivalent to a 13 percentile point increase in performance for an average student in the intervention group relative to an average student in the comparison group. Additionally, teaching phonological awareness

Figure 5. Interventions that taught phonological awareness improved phonological awareness performance more than interventions that did not teach phonological awareness



L is language. PA is phonological awareness. PK is print knowledge. W is early writing.

Note: The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. If there is no overlap between the 95 percent confidence intervals, the difference between the weighted effect sizes is considered statistically significant. Includes only interventions evaluated in high-quality impact studies and included at least one effect size estimate.

Source: Authors' analysis of primary data collected for the review; see appendix E.

resulted in a significantly larger weighted effect size than not teaching phonological awareness (figure 5; see also table B9 in appendix B).

Phonological awareness performance was comparable for researcher-developed outcome measures and standardized outcome measures. Among the 36 studied interventions that taught and evaluated effects on phonological awareness and included an effect size estimate, the weighted effect size was 0.43 for researcher-developed outcome measures and 0.28 for standardized outcome measures; the difference was not statistically significant (see table B10 in appendix B). This suggests that interventions that teach phonological awareness are likely to improve performance regardless of the type of outcome measure assessed.

Phonological awareness instruction that included identification, matching, blending, counting, segmenting, production, or a combination of these tasks significantly improved phonological awareness performance, as evidenced by a weighted effect size of 0.38 (see table B11 in appendix B). This suggests that including multiple types of phonological awareness tasks in an early literacy program is likely to improve phonological awareness performance.

Print knowledge

This section first discusses the instructional features of the 42 interventions that taught print knowledge and were evaluated in studies that included print knowledge outcomes. It then examines the effect of teaching print knowledge on print knowledge outcomes and explores the effects of instructional features in a subset of 38 interventions evaluated in high-quality impact studies and included at least one effect size estimate (see figure B9 in appendix B).

Of the 42 interventions that were evaluated in studies that taught print knowledge and included print knowledge outcomes, 26 included instruction in both letter names and letter sounds, and 10 taught either letter names or letter sounds. Of the 26 interventions that taught both letter names and sounds, 19 taught them simultaneously (see figure B10 in appendix B). Instruction in concepts about print was included in 16 of the 42 studied

interventions and co-occurred with letter name instruction or letter name and sound instruction in 11 of them. Implementers used a variety of materials when teaching print knowledge, including books (18), letter cards (with or without pictures; 12), picture cards (11), and letter-shaped manipulatives (10; see figure B11 in appendix B).

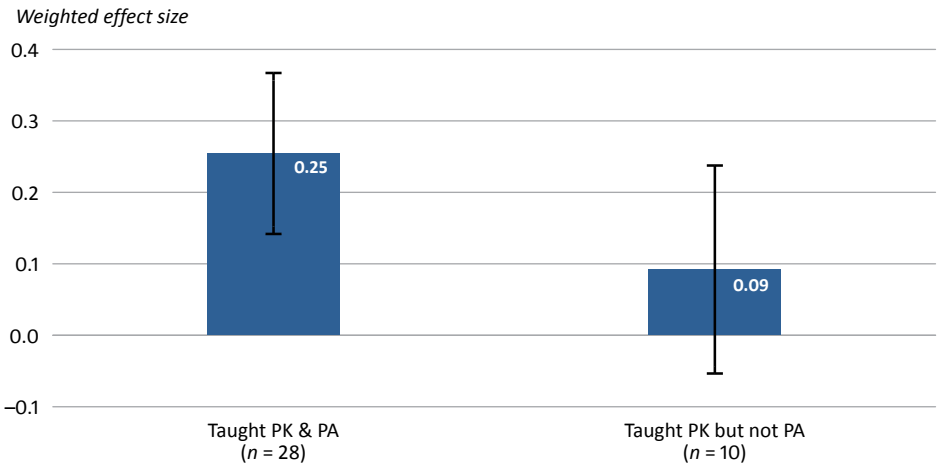
Interventions that taught both print knowledge and phonological awareness improved print knowledge performance, but interventions that taught print knowledge without teaching phonological awareness did not. Interventions that taught print knowledge resulted in a significant weighted effect size of 0.23 on print knowledge outcomes, among the 38 interventions evaluated in studies that included at least one effect size estimate (see figure 3). That effect size is equivalent to a 9 percentile point increase in performance for an average student in the intervention group relative to an average student in the comparison group. In contrast, the weighted effect size for interventions that did not teach print knowledge was 0.10 (see table B12 in appendix B).

Interventions that taught both print knowledge and phonological awareness improved print knowledge performance (weighted effect size of 0.25). On average, interventions that taught both domains significantly improved print knowledge performance; however, interventions that taught print knowledge but not phonological awareness did not (figure 6; see also table B12 in appendix B). This suggests that teaching both print knowledge and phonological awareness can benefit print knowledge performance.

Print knowledge performance was comparable for researcher-developed and standardized outcome measures. The weighted effect sizes for researcher-developed and standardized outcome measures were statistically comparable (0.30 versus 0.24; see table B13 in appendix B). This suggests that interventions that teach print knowledge can improve print knowledge performance regardless of the type of outcome measure assessed.

All print knowledge instructional features resulted in comparable print knowledge performance. Print knowledge performance was statistically comparable across mutually exclusive groupings of print knowledge instructional features. For example, instruction that included both letter names and sounds resulted in statistically equivalent print knowledge effects compared with instruction that included either letter names or letter sounds but not

Figure 6. Interventions that taught print knowledge and phonological awareness significantly improved performance in print knowledge



PA is phonological awareness. PK is print knowledge.

Note: The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. If the confidence intervals overlap, the difference between the weighted effect sizes is not statistically significant. Includes only interventions evaluated in high-quality impact studies and included at least one effect size estimate.

Source: Authors' analysis of primary data collected for the review; see appendix E.

both (0.20 versus 0.16; see table B14 in appendix B). In addition, instruction that focused on concepts about print resulted in a weighted effect size of 0.32, which did not differ significantly from the weighted effect size of interventions that did not teach concepts about print (0.18). This suggests that more research is needed to identify the instructional features that are most likely to improve print knowledge performance.

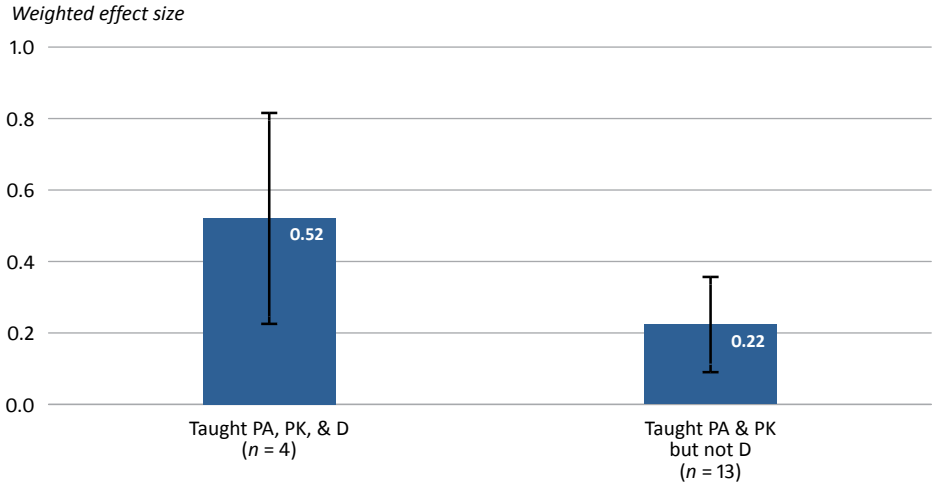
Decoding

This section first discusses the instructional features of the five interventions that taught decoding and were evaluated in studies that included decoding outcomes. It then examines the effect of teaching decoding on decoding outcomes in a subset of four high-quality impact studies and included at least one effect size estimate (see figure B12 in appendix B). Finally, it explores the effect on decoding performance for 13 studied interventions that did not teach decoding but taught phonological awareness and print knowledge, which provide the foundation for students to learn to decode (Foorman et al., 2016).

Of the five interventions that were evaluated in studies that included decoding outcomes, four included blending phonemes in printed words and used narrative books during decoding instruction (see figure B13 in appendix B). Two of the five interventions included segmenting printed words into phonemes. The study team was unable to further explore these instructional features because so few interventions taught decoding.

Interventions that taught phonological awareness and print knowledge improved decoding performance, even when decoding was not taught. Although teaching both phonological awareness and print knowledge without teaching decoding improved decoding performance, the weighted effect size on decoding was more than twice the size when all three domains were taught (figure 7). All five of the studied decoding interventions also taught phonological awareness and print knowledge. Thirteen other interventions taught phonological awareness and print knowledge but not decoding and were evaluated in studies that included decoding outcomes; these interventions significantly improved decoding performance (weighted effect size 0.22; see table B15 in appendix B).

Figure 7. Interventions that taught phonological awareness and print knowledge, with or without decoding, significantly improved decoding performance



D is decoding. PA is phonological awareness. PK is print knowledge.

Note: The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. If the confidence intervals overlap, the difference between the weighted effect sizes is not statistically significant. Includes only interventions evaluated in high-quality impact studies and included at least one effect size estimate.

Source: Authors' analysis of primary data collected for the review; see appendix E.

This suggests that improved decoding performance can occur even without decoding instruction. However, when phonological awareness, print knowledge, and decoding instruction co-occurred, the weighted effect size for decoding increased to 0.52. Although these effect sizes (0.52 and 0.22) were not significantly different, they suggest that teaching decoding along with phonological awareness and print knowledge might be more effective. Given the small number of interventions that taught and evaluated effectiveness on decoding, this finding should be interpreted with caution.

Decoding performance was comparable for researcher-developed and standardized outcome measures. The weighted effect sizes for researcher-developed and standardized outcome measures were statistically comparable (0.56 versus 0.45; see table B16 in appendix B). This suggests that interventions that teach decoding can improve decoding performance regardless of the type of outcome measure assessed.

Early writing

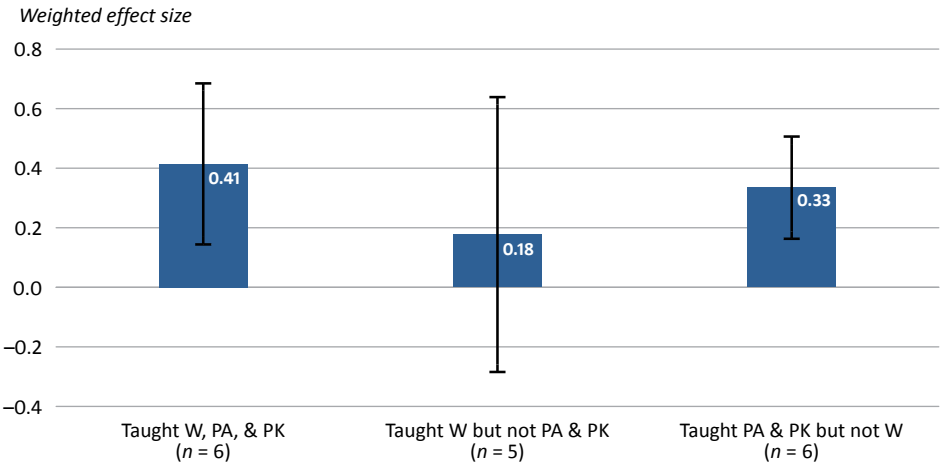
This section first discusses the instructional features of the 12 interventions that taught early writing and were evaluated in studies that included early writing outcomes. It then examines the effect of teaching early writing on early writing outcomes and explores the effects of instructional features in a subset of 11 interventions evaluated in high-quality impact studies and included at least one effect size estimate (see figure B14 in appendix B). Finally, it explores the effect on early writing performance for six studied interventions that did not teach early writing but taught phonological awareness and print knowledge, which provide the foundation for students' early writing development (Foorman et al., 2016).

Of the 12 interventions that were evaluated in studies that included early writing outcomes, 8 included instruction focused on individual letter formation, 7 included instruction focused on whole words in isolation, and 3 included instruction focused on whole words in connected text (see figure B15 in appendix B). During instruction focused on letter formation, the implementer typically modeled how to write the letter or the students were asked to trace or copy the letter. During instruction focused on whole words in isolation, the implementer asked students to copy or trace the words provided or to spell the words orally presented. Finally, during instruction focused on whole words in connected text, the implementer transcribed students' verbally expressed thoughts.

Interventions that taught phonological awareness and print knowledge, with or without early writing, improved early writing performance. On average, interventions that taught early writing, phonological awareness, and print knowledge significantly improved early writing performance, among the 11 interventions evaluated in studies that included at least one effect size estimate (weighted effect size of 0.41; figure 8; see also table B17 in appendix B). However, interventions that taught early writing but not phonological awareness and print knowledge did not improve early writing performance. Furthermore, six interventions that taught phonological awareness and print knowledge but not early writing demonstrated statistically comparable effects on early writing performance (weighted effect size of 0.33) relative to studied interventions that taught all three domains. This suggests that phonological awareness and print knowledge instruction can benefit early writing performance even without early writing instruction. As with decoding, these findings are based on a small number of interventions and should be interpreted with caution.

Interventions that taught early writing improved early writing performance when skills similar to those taught were assessed. On average, interventions that taught early writing significantly improved early writing performance on researcher-developed outcome measures, among the 11 interventions evaluated in high-quality impact studies and included at least one effect size estimate (weighted effect size of 0.36; figure 9; see also table B18 in appendix B). That effect size is equivalent to a 14 percentile point increase in performance for an average student in the intervention group compared with an average student in the comparison group. The weighted effect size for standardized outcome measures was 0.23, which was not statistically significant. This suggests that interventions that teach early writing improve performance on researcher-developed outcome measures but not

Figure 8. Interventions that taught phonological awareness and print knowledge, with or without early writing, improved early writing performance

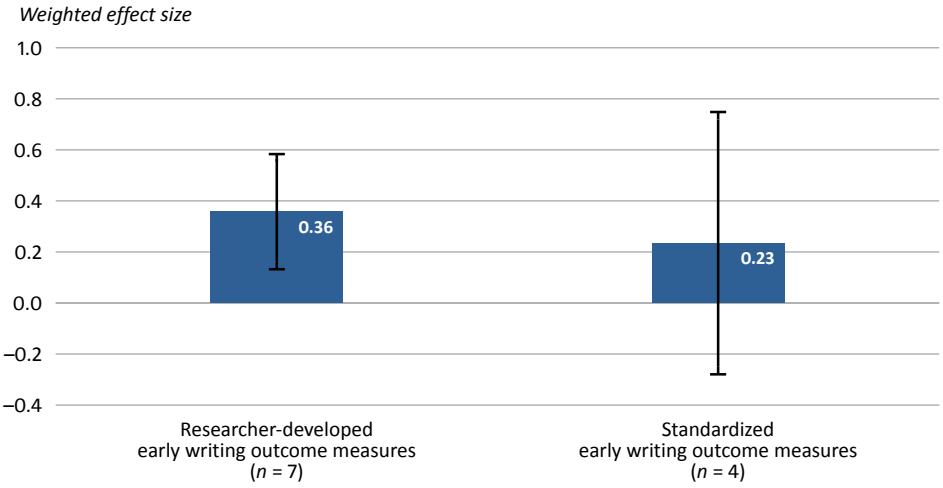


PA is phonological awareness. PK is print knowledge. W is early writing.

Note: The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. If the confidence intervals overlap, the difference between the weighted effect sizes is not statistically significant. Includes only interventions evaluated in high-quality impact studies and included at least one effect size estimate.

Source: Authors' analysis of primary data collected for the review; see appendix E.

Figure 9. Interventions that taught early writing improved performance on research-developed language outcome measures, which often represent skills similar to those taught



Note: The error bar represents the 95 percent confidence interval, meaning that there is a 95 percent probability that the "true" effect size lies between the lower and upper limits. If the interval includes 0, the weighted mean effect size is not statistically significant. If the confidence intervals overlap, the difference between the weighted effect sizes is not statistically significant. Includes only interventions evaluated in high-quality impact studies and included at least one effect size estimate.

Source: Authors' analysis of primary data collected for the review; see appendix E.

standardized outcome measures. Again, these findings are based on a small number of interventions and should thus be interpreted with caution.

All early writing instructional features resulted in comparable early writing performance. Early writing performance was statistically comparable across mutually exclusive groupings of early writing instructional features. For example, instruction that included individual letter formation or whole words in isolation resulted in average weighted effect sizes that reached practical importance (0.35 and 0.43) but did not exceed the effect size of interventions that did not include either (0.62 and 0.34; see table B19 in appendix B). This suggests that more research is needed to identify the instructional features that are most likely to improve early writing performance.

Implications

Early childhood educators can use these findings to compare their current teaching practices and materials to the evidence-based interventions described in this report that likely support the development of early literacy skills. The results from this systematic review suggest that intentional instruction in the important early literacy domains of language, phonological awareness, print knowledge, decoding, and early writing can meaningfully benefit students.

The findings can help early childhood educators improve the alignment between the focus of their instruction and the skill domains they seek to support among students. Evidence identified in this review suggests that instruction in language, phonological awareness, and decoding increases the likelihood of positively impacting performance in the domain taught. Instruction in print knowledge and early writing domains was effective when provided in combination with instruction in other domains (see figures 6 and 8). The review also provides some evidence that instruction in one domain might not necessarily impact other domains. For example, teaching only language did not positively impact skills in any other outcome domains (see figure B3 in appendix B). Similarly, instruction in domains other than language did not improve language outcomes. In contrast, there is some evidence that teaching the related skills of phonological awareness and print knowledge might also lead to improvement in decoding and early writing.

Early childhood educators can also use the findings to identify effective, evidence-based early literacy curricula, lesson packages, instructional practices, and technology programs that can be used in state-supported prekindergarten programs and in other settings. Table C1 in appendix C lists effective interventions by outcome domain. That information can be used with the information in table D1 in appendix D on implementation characteristics and information in tables E1 and E2 in appendix E on effect sizes to identify effective early literacy interventions. When reviewing these tables, it is important to keep in mind that studied interventions demonstrating inconclusive effects need further investigation to better understand their effectiveness. In some cases, studies of interventions demonstrating inconclusive effects (that is, effects that are not statistically significant) might have demonstrated effectiveness if the study evaluating those interventions had included a larger sample size (Seftor, 2016).

When selecting an intervention, educators should carefully consider the implementation characteristics used in the studied intervention. This is important because the evidence for intervention effectiveness described in this report is based on the intervention as implemented in the study being evaluated, including the specific instructional domains, participant sample, setting in which the intervention was implemented, and level of implementation and support provided. Deviations in any of these components could result in intervention impacts that differ from those described in this report.

Policymakers can use the findings when updating their state's early learning standards and practice standards for early childhood educators. In many states standards and practice documents include examples of instructional activities aligned to specific standards. These examples can be reviewed and revised as needed to ensure consistency with evidence-based instruction for each outcome domain as represented in the findings of this report.

Professional development designed for early childhood educators can draw on the findings to support educators' understanding of evidence-supported ways to improve children's performance in those domains. For example, educators designing professional development could draw on the specific information provided on the efficacy of each study (see appendix C) and the key features of the instruction in those studies (see appendix D) to identify examples of evidence-based instruction for each outcome domain.

Limitations

The findings presented are based on a large body of research; however, the search excluded interventions that primarily or exclusively targeted a specific student population (such as English learner students or students with disabilities) and policies (such as evaluations of the Head Start policy). Those interventions might be relevant to the concerns of some districts or states. Moreover, this review did not explore potential differences in effects for different populations of students, nor did it consider the cost associated with implementing each intervention.

The coding of implementation characteristics and instructional features was based solely on the description provided in the study evaluating each intervention and might not reflect all activities or instructional domains included in each intervention. Additionally, there was no minimum instruction on a particular instructional domain required for the code to be applied. For interventions that taught multiple domains, the coding of a domain does not indicate the relative emphasis that the intervention as a whole placed on the domain.

It is difficult to infer effects for a single domain from interventions that incorporate multiple domains. Almost 60 percent of the interventions in this review included instruction in more than one domain. As a result, it is often unclear whether or how instruction in one domain impacted student's performance in a different domain. Instruction that taught a specific domain might have led to a different impact on performance if it had been delivered outside the context of the other instruction being provided at the same time within the multifaceted intervention.

This review highlights several areas that could be explored in future research. The findings suggest that more research is needed to identify the instructional features that are most likely to improve print knowledge performance. Additionally, the small number of studies exploring decoding and early writing instruction that met the evidence standards makes it difficult to draw informative conclusions about instructional features that relate to effectiveness in these areas. Given the substantial variability in instructional design among the interventions deemed effective, systematic combinations of implementation characteristics and instructional features in each domain could be investigated to determine which are necessary or sufficient to generate improved performance in language, phonological awareness, print knowledge, decoding, and early writing.

This review describes only studies of interventions and outcomes that the study team deemed high quality, that is, likely to meet version 4.0 of the WWC evidence standards with or without reservations. Studies receiving this designation provide the highest quality evidence for determining the effectiveness of an intervention on student achievement. This does not mean that an intervention is ineffective if it is evaluated in a study rated as not meeting the evidence standards. It simply means that the evaluation was not implemented in a way that rigorously tested the intervention's effectiveness. Therefore, it is possible that some of the studies in this report that are not rated as high-quality impact studies include effective interventions. It is also possible that there are other effective interventions being used in prekindergarten settings that have not been empirically evaluated.

Approximately 60 studies did not meet the evidence standards because the study authors were unable to provide additional information needed to complete a full review, primarily because the original data were no longer accessible to them. In some cases completing a full review required authors to provide information beyond what was reported in the original studies. When authors were unable to do so, the study team rated studies using only

the limited information available. It is unclear whether the ratings for these studies would have changed if the authors had provided the requested information.

Lastly, although the current review included both published peer-reviewed research and other available research, some publication bias could still be present. Publication bias occurs when unfavorable results in the form of non-significant findings from experimental research influence the decision to disseminate or share findings (Cooper et al., 2009). Excluding these nonsignificant findings could result in overestimating the weighted effect size.

References

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166–176. https://eric.ed.gov/?id=EJ874842.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.

Dale, P. S., McMillan, A. J., Hayiou-Thomas, M. E., & Plomin, R. (2014). Illusory recovery: Are recovered children with early language delay at continuing elevated risk? *American Journal of Speech-Language Pathology*, *23*(3), 437–447.

Denton, C., Flanagan, K., & McPhee, C. (2009). *The children born in 2001 at kindergarten entry: First findings from the kindergarten data collections of the Early Childhood Longitudinal Study—Birth Cohort (ECLS-B)* (NCES No. 2010–005). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

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 No. 2016–4008). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved June 4, 2019, from http://whatworks.ed.gov.

Furnes, B., & Samuelsson, S. (2009). Preschool cognitive and language skills predicting kindergarten and grade 1 reading and spelling: A cross-linguistic comparison. *Journal of Research in Reading*, *32*(3), 275–292. https://eric.ed.gov/?id=EJ846628.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.

Kim, Y. S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition—a cross-classified multilevel model approach. *Journal of Educational Psychology*, *102*(2), 313–326. https://eric.ed.gov/?id=EJ884856.

Lerner, M. D., & Lonigan, C. J. (2016). Bidirectional relations between phonological awareness and letter knowledge in preschool revisited: A growth curve analysis of the relation between two code-related skills. *Journal of Experimental Child Psychology*, *144*(1), 166–183.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335. https://eric.ed.gov/?&id=EJ906930.

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6*(3), 223–262.

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352.

Moiduddin, E., Aikens, N., Tarullo, L., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009* (OPRE Report No. 2012–37a). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. National Institute for Literacy. https://eric.ed.gov/?id=ED504224.

National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 004754). U.S. Government Printing Office. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf.

Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-K. *The Reading Teacher, 62*(5), 384–392.

Phillips, B., Zhao, Y., & Weekley, M. K. (2018). Teacher language in the preschool classroom: Initial validation of a classroom environment observation tool. *Early Education and Development*, *29*(3), 379–397.

Piasta, S. B., Justice, L. M., O'Connell, A. A., Mauck, S. A., Weber-Mayere, M., Schachter, R. E., Farley, K. S. & Spear, C. F. (2017). Effectiveness of large-scale, state-sponsored language and literacy professional development on early childhood educator outcomes. *Journal of Research on Education Effectiveness*, *10*(2), 354–378.

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open, 1*(4), 1–31.

Reynolds, M. E., & Fish, M. (2010). Language skills in low-SES rural Appalachian children: Kindergarten to middle childhood. *Journal of Applied Developmental Psychology*, *31*(3), 238–248.

Seftor, N. (2016). What does it mean when a study finds no effects? (REL 2017–265). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/REL_2017265/pdf/REL_2017265.pdf.

What Works Clearinghouse. (2017a). *What Works Clearinghouse procedures handbook version 4.0.* U.S. Department of Education, Institute of Education Sciences.

What Works Clearinghouse. (2017b). *What Works Clearinghouse standards handbook version 4.0.* U.S. Department of Education, Institute of Education Sciences.

REL 2021–084

July 2021

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0011 by the Regional Educational Laboratory Southeast administered by the Florida Center for Reading Research, Florida State University. 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.

Herrera, S., Phillips, B. M., Newton, Y., Dombek, J. L., & Hernandez, J. A. (2021). *Effectiveness of early literacy instruction: Summary of 20 years of research* (REL 2021–084). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.