What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Early Childhood Education Revised April 30, 2007

Sound Foundations

Program description

Sound Foundations, a literacy curriculum designed to teach phonological awareness to preliterate children, focuses exclusively on phoneme identity (that is, different words can start and end with the same sound). It works from the principle that phonemic awareness is necessary but not sufficient to reading, which depends on the alphabetic principle (that is, the association of sounds with letters and using those sounds to form words). The curriculum is self-contained and can be used by teachers, parents, or teaching assistants.

Research

One study of *Sound Foundations* met the What Works Clearinghouse (WWC) evidence standards.[1] This study included 26 preschool children and examined intervention effects on children's phonological processing and early reading/writing. This report focuses on immediate posttest findings to determine the effectiveness of the intervention; however, follow-up findings provided by the study authors are included in the technical appendices.[2]

The WWC considers the extent of evidence for *Sound Foundations* to be small for phonological processing and early reading/writing. No studies that met WWC evidence standards with or without reservations addressed oral language, print knowledge, cognition, or math.

Effectiveness

Sound Foundations was found to have potentially positive effects on phonological processing and early reading/writing.

	Oral language	*Print knowledge*	*Phonological processing*	*Early reading/ writing*	*Cognition*	*Math*
Rating of effectiveness	na	na	Potentially positive effects	Potentially positive effects	na	na
Improvement Index[3]	na	na	na	na	na	na

na = not applicable

1. To be eligible for the WWC's review, the Early Childhood Education (ECE) interventions had to be implemented in English in center-based settings with children ages 3 to 5 or in preschool. One additional study is not included in the overall effectiveness rating because the intervention included a combination of *Sound Foundations* and *Dialogic Reading,* which does not allow the effects of *Sound Foundations* alone to be determined. See the section titled "Findings for *Sound Foundations plus Dialogic Reading*" and Appendices A4.1–A4.4 for findings from this and a related document.
2. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available. There are three follow-up manuscripts to Byrne and Fielding-Barnsley (1991). Because these report on the same sample, the four manuscripts are treated as one study.
3. Student-level improvement indices could not be computed for the outcome domains studied, phonological processing and early reading/writing.

Absence of conflict of interest

The WWC ECE topic team works with two Principal Investigators (PIs): Dr. Ellen Eliason Kisker and Dr. Christopher Lonigan. The studies on *Sound Foundations* reviewed by the ECE team included two studies on which Dr. Grover Whitehurst, director of the Institute for Education Sciences, was either the primary or a secondary author. Dr. Whitehurst's financial interests are not affected by the success or failure of *Sound Foundations,* and he does not receive any royalties or other monetary return from the use of *Sound Foundations.* In all instances where Dr. Whitehurst was a study author, he was not involved in the decision to include the study in the review, and he was not involved in the coding, reconciliation, or discussion of the included study. Drs. Kisker and Lonigan led all review activities related to those studies. The decision to review *Sound Foundations* was made by the PIs in collaboration with the rest of the ECE team following prioritization of interventions based on the results from the literature review. This report on *Sound Foundations* was reviewed by a group of independent reviewers, including members of the WWC Technical Review Team and external peer reviewers.

Additional program information

Developer and contact

Developed by Brian Byrne and Ruth Fielding-Barnsley. Correspondence concerning *Sound Foundations* should be addressed to Brian Byrne, Department of Psychology, University of New England, Armidale, New South Wales 2351, Australia. E-mail: bbyrne@une.edu.au.[4]

Scope of use

Sound Foundations was developed in the late 1980s. The first published study appeared in 1991 (Byrne & Fielding-Barnsley, 1991). Information is not available on the number or demographics of children or centers using this intervention.

Teaching

In center-based settings, teachers can use *Sound Foundations* with individual children or in small groups.

Sound Foundations emphasizes nine phonemes: seven consonants (/s/, /ʃ/ (as in ship), /l/, /m/, /p/, /t/, /g/) and two vowels (/ae/ and /e/). The first sound of each of the seven consonants is represented on one poster by a series of pictures that start with the same sound (e.g., sea, sailor, and sand) and, on another poster, the last sound of each of the seven consonants is represented by a series of pictures that end with the same sound (e.g., bus, octopus, and hippopotamus). The two vowels are also represented on a poster by pictures that focus exclusively on beginning sounds. Worksheets containing outlines of objects and characters representing the nine key phonemes, as well as the other letters of the alphabet, are also provided. Additional worksheets are available to focus on the end sounds of the nine key phonemes. *Sound Foundations* also employs two card games—dominoes and "Snap"—to emphasize four sounds (/s/, /p/, /t/, and /l/). The domino cards depict two objects on either end of the card, while "Snap" uses cards with one item depicted on each card (Byrne & Fielding-Barnsley, 1991).

Cost

Published *Sound Foundations* procedures are freely available to the public. Information is not available about the costs of teacher training and implementation of *Sound Foundations.*

Research

Four studies reviewed by the WWC investigated the effects of *Sound Foundations* in center-based settings. One study (Byrne & Fielding-Barnsley, 1991) was a randomized controlled trial that met WWC evidence standards. One additional study met the WWC evidence standards (Whitehurst, Epstein, Angell, Payne, Crone, & Fischel, 1994[5]) and is included in this report; however, Whitehurst et al. (1994) examined a combination of *Sound Foundations* and *Dialogic Reading,* which does not allow the effects

4. For more information on *Sound Foundations,* refer to Byrne, B., and Fielding-Barnsley, R. (1991). *Sound foundations: An introduction to prereading skills.* Sydney, Australia: Leyden Educational Publishers.

Research *(continued)*

of *Sound Foundations* alone to be determined. Therefore, this study is discussed separately, and the findings are not included in the intervention ratings. The remaining two studies did not meet WWC evidence screens.

Byrne and Fielding-Barnsley (1991) included 126 four- to five-year-old children from four preschools in Australia. Byrne and Fielding-Barnsley compared *Sound Foundations* to a comparison group trained on the identification of semantic (i.e., word meaning) categories.[6] This report focuses on the comparison of phonological processing outcomes and early reading/writing outcomes between the *Sound Foundations* group and the comparison group.[7]

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate to large (see the What Works Clearinghouse Extent of Evidence Categorization Scheme). The extent of evidence takes into account the number of studies and the total sample size across the studies that met WWC evidence standards with or without reservations.[8]

The WWC considers the extent of evidence for *Sound Foundations* to be small for phonological processing and early reading/writing. No studies that met WWC evidence standards with or without reservations addressed oral language, print knowledge, cognition, or math.

Effectiveness

Findings

The WWC review of interventions for early childhood education addresses children's outcomes in six domains: oral language, print knowledge, phonological processing, early reading/writing, cognition, and math.[9]

Phonological processing. Byrne and Fielding-Barnsley (1991) analyzed findings for four measures in this outcome domain,[10] all of which favored the *Sound Foundations* group and were statistically significant as calculated by the WWC. In this study, the effect was statistically significant and positive, according to WWC criteria.

Early reading/writing. Byrne and Fielding-Barnsley (1991) analyzed findings for one measure in this outcome domain.[10] The finding favored the *Sound Foundations* group and was statistically significant. The statistical significance of this effect was confirmed by the WWC. In this study, the effect was statistically significant and positive, according to WWC criteria.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings,[9] the size of the difference between participants in the intervention condition and the comparison condition, and the consistency in findings across studies (see the WWC Intervention Rating Scheme).

5. Zevenbergen, Whitehurst, and Zevenbergen (2003) report additional results from the sample first reported in Whitehurst et al. (1994), so the WWC reviewed the two manuscripts as a single study.
6. Byrne and Fielding-Barnsley (1991) implemented a modified version of the *Sound Foundations* curriculum, introducing fewer phonemes to children than are specified in the full curriculum (i.e., six phonemes instead of nine).
7. The letter knowledge outcome was not included in this review because it was used to test the prediction that both phoneme identity and letter knowledge are necessary conditions for acquisition of the alphabetic principle. It was not used to test the effects of the intervention.
8. The Extent of Evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of the *Sound Foundations* report, corrections for clustering and multiple comparisons were needed.
10. The authors also reported results from the one-, two-, three-, and six-year follow-up tests. Because the primary focus of this review is on the immediate posttest results, the follow-up results are not discussed here but are included in Appendices A5.1–A5.5.

The WWC found *Sound Foundations* to have potentially positive effects for phonological processing and early reading/writing

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see Technical Details of WWC-Conducted Computations). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results. A student-level average improvement index and range of improvement indices could not be computed.

Findings for *Sound Foundations plus Dialogic Reading*

The study described below does not contribute to the overall rating of effectiveness because the intervention included a combination of *Sound Foundations* and *Dialogic Reading,* which does not allow the effects of *Sound Foundations* alone to be determined. However, the WWC believes that the findings from this combined intervention may provide useful information to practitioners who are making a determination about the merits of combining *Sound Foundations* with a specific interactive shared book reading practice *(Dialogic Reading)*. The WWC reports the individual study findings here and in Appendices A4.1–A4.4.

Whitehurst et al. (1994) included 167 at-risk low-income four-year-old children from four Head Start centers in Suffolk County, New York. The study compared oral language, print knowledge, phonological processing, and early reading/writing outcomes for children participating in an adapted *Sound Foundations* curriculum combined with *Dialogic Reading* to outcomes for children in a no-treatment comparison group participating in their regular Head Start services.[11]

Oral language. Whitehurst et al. (1994) found no statistically significant difference between the intervention group and the comparison group on oral language as measured by the Language factor.[12] Zevenbergen, Whitehurst, and Zevenbergen (2003), a second report on the same study, reported findings on four additional oral language measures from the same study, none of which were statistically significant as calculated by the WWC. The average effect across the five measures was neither statistically significant nor large enough to be considered substantively important, according to WWC criteria. The average improvement index for oral language is +6 percentile points, with a range of –12 to +19 percentile points across findings.

Print knowledge. Whitehurst et al. (1994) reported a statistically significant difference favoring the intervention group on the Print Concepts factor.[12] The statistical significance of this effect was confirmed by the WWC. The improvement index for print knowledge is +24 percentile points for the one print knowledge outcome in this study.

Phonological processing. Whitehurst et al. (1994) reported neither statistically significant nor substantively important effects on the Linguistic Awareness factor.[12] The improvement index for phonological processing is +1 percentile point for the one phonological processing outcome in this study.

Early reading/writing. Whitehurst et al. (1994) reported a statistically significant difference favoring the intervention group on the Writing factor.[12] The statistical significance of this effect was confirmed by the WWC. The improvement index for early reading/writing is +20 percentile points for the one early reading/writing outcome in this study.

11. Whitehurst et al. (1994) implemented a modified version of the *Sound Foundations* curriculum. Changes included substituting the card games with other games and extension activities and omitting the audio tapes. This study is also included in the WWC *Dialogic Reading* intervention report.
12. To reduce data the study authors conducted a principal components analysis on the 21 measures. The WWC presents results only for the four factor scores (Language factor, Print Concepts factor, Linguistic Awareness factor, and Writing factor) because effect sizes could not be computed for the individual measures.

The WWC found *Sound Foundations* to have potentially positive effects for phonological processing and early reading/writing *(continued)*

Summary

The WWC reviewed four studies on *Sound Foundations.* One of the studies met WWC evidence standards. One additional study that met WWC evidence standards is described in this report but is not included in the overall rating of effectiveness. The remaining two studies did not meet WWC evidence screens. Based on the one study included in the overall rating of effectiveness, the WWC found potentially positive effects for both phonological processing and early reading/writing. Based on the study that included a *Sound Foundations plus Dialogic Reading* intervention, the WWC found no discernible effects on oral language, potentially positive effects on print knowledge, no discernible effects on phonological processing, and potentially positive effects on early reading/writing.

Although this report focuses on immediate posttest findings to determine the effectiveness of the intervention, longer term follow-up findings of the Byrne and Fielding-Barnsley (1991) study (i.e., Byrne & Fielding-Barnsley, 1993, 1995; Byrne, Fielding-Barnsley, & Ashley, 2000) are reported in the technical appendices.

The evidence presented in this report may change as new research emerges.

References

Met WWC evidence standards

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83*(4), 451–455.

Additional sources:

Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*(1), 104–111.

Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*(3), 488–503.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92*(4), 659–667.

Whitehurst, G. J., Epstein, J. N., Angell, A. L., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology, 86*(4), 542–555.

Additional sources:

Epstein, J. N. (1994). Accelerating the literacy development of disadvantaged preschool children: An experimental evaluation of a Head Start emergent literacy curriculum. *Dissertation Abstracts International, 55*(11), 5065B. (UMI No. 9510085)

Zevenbergen, A. A., Whitehurst, G. J., & Zevenbergen, J. A. (2003). Effects of a shared-reading intervention on the inclusion of evaluative devices in narratives of children from low-income families. *Journal of Applied Developmental Psychology, 24,* 1–15.

Did not meet WWC evidence screens

Elliott, J., Prior, M., Merrigan, C., & Ballinger, K. (2002). Evaluation of a community intervention programme for preschool behaviour problems. *Journal of Pediatric Child Health, 38,* 41–50.[13]

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology, 91*(2), 267–272.[14]

For more information about specific studies and WWC calculations, please see the WWC *Sound Foundations* Technical Appendices.

13. The outcome measures are not relevant to this review.
14. Complete data were not reported: the WWC could not compute effect sizes.

Appendix

Appendix A1.1 Study characteristics: Byrne & Fielding-Barnsley, 1991 (randomized controlled trial)

Characteristic	Description
Study citation	Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83*(4), 451–455. *Additional sources:* Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*(1), 104–111 (randomized controlled trial with attrition problems). Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*(3), 488–503 (randomized controlled trial with attrition problems). Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92*(4), 659–667 (randomized controlled trial with attrition problems).
Participants	The study began with 128 children; two comparison group children left the study, leaving a sample of 126 children. The mean age of the children in the intervention condition was 55.4 months and the mean age of the children in the comparison condition was 55.0 months. Forty-five percent of the sample was female. The authors reported that the children were randomly assigned to the intervention and comparison conditions with the provision that the number of children from each preschool was equally distributed across groups.
Setting	The study took place in four preschools in Australia.
Intervention	Children in the intervention condition were trained in groups[1] of four to six for a 12-week period. The weekly training sessions were 25–30 minutes long. In the first 11 weeks children were taught five consonants (/s/, /m/, /t/, /l/, and /p/ in initial and final positions) and one vowel (/ae/ in initial position). Individual phonemes were taught in two consecutive weeks. The first week focused on the phoneme in initial positions and the second week focused on phonemes in final positions. In each session, worksheets with outline drawings, where children identified and colored the critical items, were introduced following the teaching of any particular phoneme. In the 12th week of the intervention, the researchers introduced card games, dominoes and "Snap," which focused on four phonemes (/s/, /t/, /l/, and /p/) in initial and final positions.
Comparison	Children in the comparison condition were trained in groups of four to six for a 12-week period. The weekly training sessions were 25–30 minutes long. This training focused on teaching children to find semantic categories in worksheets and posters after hearing a story. Children in this condition did not receive phoneme training.

1. Although students were randomly assigned to intervention and comparison conditions, the authors conducted the intervention and comparison activities with small groups of children (teaching groups) and used the teaching group as the unit of analysis. Student-level data were not available to the WWC. The authors' analysis provides a conservative test of the intervention's effectiveness. There were 12 teaching groups in the intervention condition and 12 teaching groups in the comparison condition.

(continued)

Appendix A1.1 Study characteristics: Byrne & Fielding-Barnsley, 1991 (randomized controlled trial) *(continued)*

Characteristic	Description
Primary outcomes and measurement	The primary outcome domains assessed were children's phonological processing and early reading/writing. Phonological processing was assessed with four nonstandardized measures: phoneme initial trained, phoneme initial untrained, phoneme final trained, and phoneme final untrained. Early reading/writing was assessed with two nonstandardized measures: word choice and letter knowledge. The letter knowledge measure was not considered in this review because it was used to test the prediction that both phoneme identity and letter knowledge are necessary conditions for acquisition of the alphabetic principle. It was not used to test the effects of the intervention. (See Appendices A2.3–A2.4 for more detailed descriptions of outcome measures.) Below are the details of the measures used in the follow-up studies of this intervention. Although the results of the follow-up studies are not part of the WWC effectiveness ratings, they are reported in Appendices A5.1–6. Byrne and Fielding-Barnsley (1993) assessed the outcome domains of children's print knowledge, phonological processing, and early reading/writing. Print knowledge was assessed with a nonstandardized measure of alphabet knowledge. Phonological processing was assessed with four nonstandardized measures: phoneme identity initial, phoneme identity final, phoneme elision initial, and phoneme elision final. Early reading/writing was assessed with two nonstandardized measures (pseudoword identification and spelling) and one standardized measure (the Woodcock Reading Mastery Test-Revised, Form G–word identification). Byrne and Fielding-Barnsley (1995) assessed the outcome domains of children's oral language (grade 1), print knowledge (grade 1), phonological processing (grade 1), early reading/writing (grades 1 and 2), and math (grade 2). Oral language was assessed with a nonstandardized measure of listening comprehension. Print knowledge was assessed with a nonstandardized measure of alphabet knowledge, but it is not included in this report because there is not sufficient information to compute an effect size. Phonological processing was assessed with a nonstandardized test of phoneme identity, but it is not included in this report because there is not sufficient information to compute an effect size. Early reading/writing was assessed in grade 1 with three nonstandardized tests of word identification and reading (reading regular words, reading irregular words, and reading pseudowords) and three nonstandardized tests of spelling (spelling regular words, spelling irregular words, and spelling pseudowords). Early reading/writing was assessed in grade 2 with a series of nonstandardized tests assessing number names, pseudowords, regular words, irregular words, and reading comprehension. Math was measured with a nonstandardized test of number identification to determine children's ability to recognize nonalphabetic symbols. The researchers also utilized tests of rapid naming and title recognition. The rapid naming test is not included in this report because it does not test the effects of the intervention, and the title recognition test is not included because it is not relevant to the WWC review. Byrne et al. (2000) assessed the outcome domain of children's early reading/writing. This domain was assessed with five standardized measures (word attack and word identification subtests of the Woodcock Reading Mastery Tests-Revised, Castles' list nonwords, Castles' list regular words, and Castles' list irregular words) and one nonstandardized measure (South Australian Test of Written Spelling). The researchers also used a test of title recognition, but it is not included in this report because it is not relevant to the WWC review.
Teacher training	Implementation of both the intervention and comparison conditions was conducted by the second author. The WWC found no reasons to believe that the person implementing the intervention and comparison condition was not equally trained and motivated to implement each condition.

Appendix A1.2 Study characteristics: Whitehurst, Epstein, Angell, Payne, Crone, & Fischel, 1994 and Zevenbergen, Whitehurst, & Zevenbergen, 2003 (randomized controlled trial)

Characteristic	Description
Study citation	Whitehurst, G. J., Epstein, J. N., Angell, A. L., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology, 86*(4), 542–555. *Additional sources:* Epstein, J. N. (1994). Accelerating the literacy development of disadvantaged preschool children: An experimental evaluation of a Head Start emergent literacy curriculum. *Dissertation Abstracts International, 55*(11), 5065B. (UMI No. 9510085) Zevenbergen, A. A., Whitehurst, G. J., & Zevenbergen, J. A. (2003). Effects of a shared-reading intervention on the inclusion of evaluative devices in narratives of children from low-income families. *Journal of Applied Developmental Psychology, 24,* 1–15. This study is not included in the overall effectiveness rating because the intervention included a combination of *Sound Foundations* and *Dialogic Reading,* which does not allow the effects of *Sound Foundations* alone to be determined. The study is also included in the WWC *Dialogic Reading* intervention report.
Participants	The study began with 207 four-year-old at-risk low socioeconomic status children. Forty children did not complete the study, leaving 167 children in the final sample. The final sample of children was 46% Caucasian, 45% African-American, 8% Latin American, and 1% Asian. Forty-four percent of the sample was female. Fifteen classrooms[1] were randomly assigned to the intervention and comparison conditions.
Setting	The study took place in 15 classrooms from four Head Start centers in Suffolk County, New York.
Intervention	Children in the intervention condition participated in an emergent literacy program at school (*Dialogic Reading* plus an adaptation of *Sound Foundations*) and one-on-one *Dialogic Reading* at home. *Dialogic Reading* occurred over a 30-week period and consisted of reading to children in small groups three to five times a week in the classroom and one-on-one reading at home with the same book. A different book was used each week and the researchers added hints to each book (e.g., wh- and recall prompts). *Sound Foundations* occurred in the classroom at least two times a week for no more than 45 minutes a week over a 16-week period. Children were introduced to seven consonant sounds at the beginning and ending of words, to two vowel sounds at the beginning of words, and to manuscript letters that correspond to curriculum sounds.
Comparison	Children in the no-treatment comparison group participated in their regular "business as usual" Head Start program.
Primary outcomes and measurement	Whitehurst et al. (1994) examined outcomes in the oral language, phonological processing, print knowledge, and early reading/writing domains. Children's oral language was measured by three standardized measures: Peabody Picture Vocabulary Test-Revised (PPVT-R), Expressive One Word Picture Vocabulary Test-Revised (EOWPVT-R), and the Illinois Test of Psycholinguistic Abilities-Verbal Expression Subscale (ITPA-VE). Children's literacy was measured by 18 subscales from the Developing Skills Checklist. Because of the large number of outcome measures (21), the study authors conducted a principal components analysis to reduce the data to four factors for the analyses: language (oral language domain), print concepts (print knowledge domain), linguistic awareness (phonological processing domain), and writing (early reading/writing domain). (See Appendices A2.1–A2.4 for more detailed descriptions of outcome measures.) Zevenbergen et al. (2003) tested additional oral languages outcomes from the same study. They assessed children's narrative understanding by asking children to retell a story about a bus immediately after hearing the story. The narrative was transcribed and coded for general content (information) and children's use of narrative devices (references to character states, dialogue, and causal states—all in the oral language domain). (See Appendix A2.1 for more detailed descriptions of outcome measures.)
Teacher training	Parents and teachers were trained by the authors on *Dialogic Reading* using a 20-minute video, which was combined with role-playing and discussion after viewing the video. Training occurred once at the beginning of the school year. Teachers and aides in the intervention classrooms were asked to keep a daily log of the reading activities. To observe compliance and provide feedback, each classroom was visited at least once every two weeks by one of the study authors. Specific training for *Sound Foundations* is not reported.

Appendix A2.1 Outcome measures in the oral language domain

Outcome measure	Description
Language factor	A factor derived from a number of outcome measures (subscales from the Developing Skills Checklist; Illinois Test of Psycholinguistic Abilities-Verbal Expression Subscale, ITPA-VE; Peabody Picture Vocabulary Test-Revised, PPVT-R; and Expressive One Word Picture Vocabulary Test-Revised, EOWPVT-R) using a principal components analysis. Factor loadings for the language factor were high for EOWPVT-R, PPVT-R, ITPA-VE, Tell a Story in Sequence, and Identify Function of Words-Numbers (as cited in Whitehurst et al., 1994).
Reference to character states	Children heard an adapted version of the Bus Story (Renfrew, 1969 as cited in Zevenbergen et al., 2003) and then retold the story. Researchers coded transcripts of the children's narrative to determine the number of times children referred to the internal states of the characters in the story.
Dialogue	Children heard an adapted version of the Bus Story (Renfrew, 1969 as cited in Zevenbergen et al., 2003) and then retold the story. Researchers coded transcripts of the children's narrative to determine their usage of dialogue.
Causal statements	Children heard an adapted version of the Bus Story (Renfrew, 1969 as cited in Zevenbergen et al., 2003) and then retold the story. Researchers coded transcripts of the children's narrative to determine their usage of causal statements.
Information/general content score	Children heard an adapted version of the Bus Story (Renfrew, 1969 as cited in Zevenbergen et al., 2003) and then retold the story. Researchers coded transcripts of the children's narrative to rate the general content of the story.

Appendix A2.2 Outcome measure in the print knowledge domain

Outcome measure	Description
Print concepts factor	A factor derived from a number of outcome measures (subscales from the Developing Skills Checklist, ITPA-VE, PPVT-R, and EOWPVT-R) using a principal components analysis. Factor loadings for the print concepts factor were high for Name Letters, Blend CVC Words, Rhyming, Identify People Reading, Distinguish Words-Pictures-Numbers, Identify Functions of Words-Numbers, and Identify Components of Writing (as cited in Whitehurst et al., 1994).

Appendix A2.3 Outcome measures in the phonological processing domain

Outcome measure	Description
Phoneme initial trained	A researcher-developed 12-item test in which children were asked to identify which word had the same initial phoneme as the target (for example, lamp: shoe, lock, heart). Four of the phonemes selected were part of the identity-training procedure (/s/, /m/, /t/, /l/) (as cited in Byrne & Fielding-Barnsley, 1991).
Phoneme initial untrained	A researcher-developed 12-item test in which children were asked to identify which word had the same initial phoneme as the target (for example, lamp: shoe, lock, heart). Four of the phonemes selected were not part of the identity-training procedure (/f/, /n/, /b/, /k/) (as cited in Byrne & Fielding-Barnsley, 1991).
Phoneme final trained	A researcher-developed 12-item test in which children were asked to identify which word had the same final phoneme as the target (for example, drum: horse, swim, kite). Four of the phonemes selected were part of the identity-training procedure (/s/, /m/, /t/, /l/) (as cited in Byrne & Fielding-Barnsley, 1991).
Phoneme final untrained	A researcher-developed 12-item test in which children were asked to identify which word had the same final phoneme as the target (for example, drum: horse, swim, kite). Four of the phonemes selected were not part of the identity-training procedure (/f/, /n/, /b/, /k/) (as cited in Byrne & Fielding-Barnsley, 1991).

(continued)

Appendix A2.3 Outcome measures in the phonological processing domain *(continued)*

Outcome measure	Description
Linguistic awareness factor	A factor extracted from a number of outcome measures (subscales from the Developing Skills Checklist, ITPA-VE, PPVT-R, and EOWPVT-R) using a principal components analysis. Factor loadings for the linguistic awareness factor were high on Identify Sounds and Letters, Identify Same-Different Sounds, Segment Sentences, and Segment Words (as cited in Whitehurst et al., 1994).

Appendix A2.4 Outcome measures in the early reading/writing domain

Outcome measure	Description
Word choice	A researcher-developed measure in which children were shown 10 words that either began or ended with a letter that was taught during the intervention (i.e., sat, mat, pam, lam, tap, sap, map, pat, lap, and pal) and asked to say what each word said (as cited in Byrne & Fielding-Barnsley, 1991).
Writing factor	A factor derived from a number of outcome measures (subscales from the Developing Skills Checklist, ITPA-VE, PPVT-R, and EOWPVT-R) using a principal components analysis. Factor loadings for the writing factor were high for Print in Left-Right Progression, Print First Name, and Write Message Mechanics (as cited in Whitehurst et al., 1994).

Appendix A3.1 Summary of study findings included in the rating for the phonological processing domain[1]

Outcome measure	Study sample	Sample size[3] (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group[4]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[4]	WWC calculations: Mean difference[5] (*Sound Foundations* – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Byrne & Fielding-Barnsley, 1991 (randomized controlled trial)[9]								
Phoneme initial trained	4 year olds	24	11.07 (nr)	7.87 (nr)	3.20	na[11]	Statistically significant	na[11]
Phoneme initial untrained	4 year olds	24	10.12 (nr)	7.49 (nr)	2.63	na[11]	Statistically significant	na[11]
Phoneme final trained	4 year olds	24	10.40 (nr)	6.29 (nr)	4.11	na[11]	Statistically significant	na[11]
Phoneme final untrained	4 year olds	24	9.83 (nr)	6.34 (nr)	3.49	na[11]	Statistically significant	na[11]
Domain average[10] for phonological processing						na[11]	Statistically significant	na[11]

na = not applicable
nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the improvement indices. Follow-up findings from the same study are not included in these ratings, but are reported in Appendix A5.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the study authors.
3. The sample size reported is the teaching group.
4. The means were computed using the teaching group as the unit of analysis; this information was provided by the study authors.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne and Fielding-Barnsley (1991), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.
11. Student-level standard deviations were not available for this study. Teaching group standard deviations for the intervention group were 0.60 for phoneme initial trained, 0.89 for phoneme initial untrained, 0.87 for phoneme final trained, and 0.73 for phoneme final untrained. Teaching group standard deviations for the comparison group were 1.78 for phoneme initial trained, 1.58 for phoneme initial untrained, 1.70 for phoneme final trained, and 1.25 for phoneme final untrained. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A3.2 Summary of study findings included in the rating for the early reading/writing domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group[4]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[4]	WWC calculations: Mean difference[5] (*Sound Foundations* – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Byrne & Fielding-Barnsley, 1991 (randomized controlled trial)[9]								
Word Choice	4 year olds	24	8.14 (nr)	6.25 (nr)	1.89	na[11]	Statistically Significant	na[11]
Domain average[10] for early reading/writing						na[11]	Statistically significant	na[11]

na = not applicable
nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the improvement indices. Follow-up and subscale findings from the same study are not included in these ratings, but are reported in Appendix A5.4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the study authors.
3. The sample size reported is the teaching group.
4. The means were computed using the teaching group as the unit of analysis; this information was provided by the study authors.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne and Fielding-Barnsley (1991), no corrections for clustering or multiple comparisons were needed.
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.
11. Student-level standard deviations were not available for this study. The teaching group standard deviation was 1.12 for the intervention group and 1.38 for the comparison group. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A4.1 Summary of findings for *Sound Foundations plus Dialogic Reading* for the oral language domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations + Dialogic Reading* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Sound Foundations + Dialogic Reading* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Whitehurst et al., 1994 and Zevenbergen et al., 2003 (randomized controlled trials)[8]								
Language factor	4 year olds	15/167	–0.02 (1.00)	–0.10 (1.00)	0.08	0.08	ns	+3
Character states	4 year olds	16/123	1.42 (1.82)	0.67 (0.86)	0.75	0.50	ns	+19
Dialogue	4 year olds	16/123	1.56 (1.44)	0.96 (0.92)	0.60	0.48	ns	+18
Causal state	4 year olds	16/123	0.18 (0.41)	0.33 (0.58)	–0.15	–0.30	ns	–12
Information/general content	4 year olds	16/123	87.54 (14.32)	87.40 (11.50)	0.14	0.01	ns	0
Domain average[9] for oral language						0.15	ns	+6

ns = not statistically significant

1. This appendix presents a summary of study findings for measures that fall in the oral language domain for a study that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. For Whitehurst et al. (1994) the standard deviations are not reported by the study author but are reported as 1.00 by the WWC because standardized factor scores have a mean of zero and a standard deviation of one.
3. For Whitehurst et al. (1994), the intervention and comparison group means were estimated by the WWC from the y-axis of figure 2 in the Whitehurst et al. (1994) article. For Zevenbergen et al. (2003), the posttest means are covariate-adjusted means provided by the study authors.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Whitehurst et al. (1994), a correction for clustering was needed, so the significance levels may differ from those reported in the original studies.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.2 Summary of findings for *Sound Foundations plus Dialogic Reading* for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations + Dialogic Reading* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Sound Foundations + Dialogic Reading* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Whitehurst et al., 1994 (randomized controlled trial)[8]								
Print concepts factor	4 year olds	15/167	0.26 (1.00)	–0.38 (1.00)	0.64	0.64	Statistically significant	+24
Domain average[9] for print knowledge						0.64	Statistically significant	+24

1. This appendix presents a summary of study findings for measures that fall in the print knowledge domain for a study that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were not reported by the study author but are reported as 1.00 by the WWC because standardized factor scores have a mean of zero and a standard deviation of one.
3. The intervention and comparison group means were estimated by the WWC from the y-axis of figure 2 in the Whitehurst et al. (1994) article.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Whitehurst et al. (1994), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.3 Summary of findings for *Sound Foundations plus Dialogic Reading* for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations + Dialogic Reading* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Sound Foundations + Dialogic Reading* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Whitehurst et al., 1994 (randomized controlled trial)[8]								
Linguistic awareness factor	4 year olds	15/167	0.08 (1.00)	0.06 (1.00)	0.02	0.02	ns	+1
Domain average[9] for phonological processing						0.02	ns	+1

ns = not statistically significant

1. This appendix presents a summary of study findings for measures that fall in the phonological processing domain for a study that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were not reported by the study author but are reported as 1.00 by the WWC because standardized factor scores have a mean of zero and a standard deviation of one.
3. The intervention and comparison group means were estimated by the WWC from the y-axis of figure 2 in the Whitehurst et al. (1994) article.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Whitehurst et al. (1994), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.4 Summary of findings for *Sound Foundations plus Dialogic Reading* for the early reading/writing domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations + Dialogic Reading* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Sound Foundations + Dialogic Reading* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Whitehurst et al., 1994 (randomized controlled trial)[8]								
Writing factor	4 year olds	15/167	0.20 (1.00)	–0.34 (1.00)	0.54	0.54	Statistically significant	+20
Domain average[9] for early reading/writing						0.54	Statistically significant	+20

1. This appendix presents a summary of study findings for measures that fall in the early reading/writing domain for a study that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were not reported by the study author but are reported as 1.00 by the WWC because standardized factor scores have a mean of zero and a standard deviation of one.
3. The intervention and comparison group means were estimated by the WWC from the y-axis of figure 2 in the Whitehurst et al. (1994) article.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Whitehurst et al. (1994), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.1 Summary of follow-up study findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (teaching group)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Sound Foundations* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Byrne & Fielding-Barnsley, 1995 (randomized controlled trial with attrition problems)[7]								
Grade 2 comprehension listening	8 year olds	24	0.16 (nr)	–0.09 (nr)	0.25	na[8]	ns	na[8]

na = not applicable
ns = not statistically significant
nr = not reported

1. This appendix reports three-year follow-up findings (the intervention was implemented when the children were in preschool; the data in this table were collected when children were in grade 2) for measures that fall in the oral language domain. Oral language measures were not assessed in the immediate posttests.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne and Fielding-Barnsley (1995), no correction for clustering was needed.
8. Student-level standard deviations were not available for this study. The teaching group standard deviation was 0.33 for the intervention group and 0.62 for the comparison group. The student-level effect size and improvement index could not be computed. However, the statistical significance for this study is comparable to other studies. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A5.2 Summary of follow-up study findings for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (teaching group)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Sound Foundations* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Byrne & Fielding-Barnsley, 1993 (randomized controlled trial with attrition problems)[7]								
Alphabet knowledge	6 year olds	24	23.90 (nr)	23.90 (nr)	0	na[8]	ns	na[8]

na = not applicable
ns = not statistically significant
nr = not reported

1. This appendix reports one-year follow-up findings (the intervention was implemented when the children were in preschool; the data in this table were collected when children were in kindergarten) for measures that fall in the print knowledge domain. Print knowledge measures were not assessed in the immediate posttests.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne and Fielding-Barnsley (1993), no correction for clustering was needed.
8. Student-level standard deviations were not available for this study. The teaching group standard deviation was 2.10 for the intervention group and 2.60 for the comparison group. The student-level effect size and improvement index could not be computed. However, the statistical significance for this study is comparable to other studies. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A5.3 Summary of follow-up study findings for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (teaching group)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Sound Foundations* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Byrne & Fielding-Barnsley, 1993 (randomized controlled trial with attrition problems)[7]								
Phoneme identity initial	6 year olds	24	11.80 (nr)	11.20 (nr)	0.60	na[8]	ns	na[8]
Phoneme identity final	6 year olds	24	11.00 (nr)	9.70 (nr)	1.30	na[8]	Statistically significant	na[8]
Phoneme elision initial	6 year olds	24	5.10 (nr)	4.40 (nr)	0.70	na[8]	ns	na[8]
Phoneme elision final	6 year olds	24	7.90 (nr)	6.90 (nr)	1.00	na[8]	ns	na[8]

na = not applicable
ns = not statistically significant
nr = not reported

1. This appendix reports one-year follow-up findings (the intervention was implemented when the children were in preschool; the data in this table were collected when children were in kindergarten) for measures that fall in the phonological processing domain. Immediate posttest findings were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne and Fielding-Barnsley (1993), no correction for clustering was needed.
8. Student-level standard deviations were not available for this study. The teaching group standard deviations for the intervention group were 0.30 for phoneme identity initial, 0.60 for phoneme identity final, 1.70 for phoneme elision initial, and 1.10 for phoneme elision final. The teaching group standard deviations for the comparison group were 1.40 for phoneme identity initial, 1.60 for phoneme identity final, 1.80 for phoneme elision initial, and 1.40 for phoneme elision final. The student-level effect size and improvement index could not be computed. However, the statistical significance for this study is comparable to other studies. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A5.4 Summary of follow-up study findings for the early reading/writing domain[1]

Outcome measure	Study sample	Sample size (teaching group or children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Sound Foundations* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Byrne & Fielding-Barnsley, 1993 (randomized controlled trial with attrition problems)[8]								
Word identification	6 year olds	24	110.60 (nr)	108.60 (nr)	2.00	na[11]	ns	na[11]
Pseudoword identification	6 year olds	24	12.40 (nr)	10.60 (nr)	1.80	na[11]	Statistically significant	na[11]
Spelling	6 year olds	24	57.40 (nr)	53.40 (nr)	4.00	na[11]	ns	na[11]
Byrne & Fielding-Barnsley, 1995 (randomized controlled trial with attrition problems)[9]								
Grade 1 reading regular words	7 year olds	24	17.47 (nr)	15.86 (nr)	1.61	na[11]	ns	na[11]
Grade 1 reading irregular words	7 year olds	24	11.34 (nr)	10.80 (nr)	0.54	na[11]	ns	na[11]
Grade 1 reading pseudowords	7 year olds	24	14.47 (nr)	10.25 (nr)	4.22	na[11]	Statistically significant	na[11]
Grade 1 spelling regular words	7 year olds	24	28.99 (nr)	28.89 (nr)	0.10	na[11]	ns	na[11]
Grade 1 spelling irregular words	7 year olds	24	30.20 (nr)	29.88 (nr)	0.32	na[11]	ns	na[11]
Grade 1 spelling pseudowords	7 year olds	24	26.72 (nr)	26.46 (nr)	0.26	na[11]	ns	na[11]
Grade 2 number names	8 year olds	24	7.78 (nr)	7.77 (nr)	0.01	na[11]	ns	na[11]
Grade 2 pseudowords list 1	8 year olds	24	7.76 (nr)	7.21 (nr)	0.55	na[11]	Statistically significant	na[11]
Grade 2 words regular	8 year olds	24	28.81 (nr)	28.15 (nr)	0.66	na[11]	ns	na[11]
Grade 2 words irregular	8 year olds	24	24.64 (nr)	23.73 (nr)	0.91	na[11]	ns	na[11]

(continued)

Appendix A5.4 Summary of follow-up study findings for the early reading/writing domain *(continued)*

Outcome measure	Study sample	Sample size (teaching group or children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Sound Foundations* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Grade 2 pseudowords list 2	8 year olds	24	24.09 (nr)	20.42 (nr)	3.67	na[11]	Statistically significant	na[11]
Grade 2 pseudowords list 3 irregular consistent	8 year olds	24	6.97 (nr)	6.34 (nr)	0.63	na[11]	ns	na[11]
Grade 2 pseudowords list 3 total correct	8 year olds	24	17.03 (nr)	15.57 (nr)	1.46	na[11]	Statistically significant	na[11]
Grade 2 comprehension reading	8 year olds	24	0.22 (nr)	–0.19 (nr)	0.41	na[11]	Statistically significant	na[11]
Byrne, Fielding-Barnsley, & Ashley, 2000 (randomized controlled trial with attrition problems)[10]								
Word attack subtests of the WRMT-R	11 year olds	103	105.20 (14.60)	100.50 (12.00)	4.70	0.35	ns	+14
Word identification subtests of the WRMT-R	11 year olds	103	102.30 (22.10)	97.60 (22.30)	4.70	0.21	ns	+8
Castles' list–total	11 year olds	103	–0.13 (0.97)	–0.44 (0.89)	0.31	0.33	ns	+13
South Australian test of written spelling	11 year olds	103	40.40 (7.84)	38.60 (8.60)	1.80	0.22	ns	+9

na = not applicable
nr = not reported
ns = not statistically significant
WRMT-R = Woodcock Reading Mastery Test-Revised

1. This appendix reports one-, two-, three-, and six-year follow-up findings (the intervention was implemented when the children were in preschool; the data in this table were collected when children were in kindergarten, 1993; grade 1 and grade 2, 1995; and grade 5, 2000) for measures that fall in the early reading/writing domain. Immediate posttest findings were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For Byrne and Fielding-Barnsley (1993, 1995), the sample size reported is the teaching group. For Byrne et al. (2000), the sample size reported is the number of children.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.

(continued)

Appendix A5.4 Summary of follow-up study findings for the early reading/writing domain *(continued)*

8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne and Fielding-Barnsley (1993), no correction for clustering was needed.
9. In the case of Byrne and Fielding-Barnsley (1995), no corrrection for clustering was needed.
10. In the case of Byrne et al. (2000), no correction for clustering was needed.
11. Student-level standard deviations were not available for this study. The teaching group standard deviations ranged from 0.28 to 9.10 for the intervention group and from 0.31 to 13.60 for the comparison group, depending on the grade level and test. The student-level effect size and improvement index could not be computed. However, the statistical significance for this study is comparable to other studies. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A5.5 Summary of follow-up study findings for the math domain[1]

Outcome measure	Study sample	Sample size (teaching group)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Sound Foundations* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Byrne & Fielding-Barnsley, 1995 (randomized controlled trial with attrition problems)[7]								
Grade 2 numerals	8 year olds	24	7.85 (0.18)	7.85 (0.29)	0	na[8]	ns	na[8]

na = not applicable

ns = not statistically significant

1. This appendix reports three-year follow-up findings (the intervention was implemented when the children were in preschool; the data in this table were collected when children were in grade 2) for measures that fall in the math domain. Math measures were not assessed in the immediate posttests.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne and Fielding-Barnsley (1995), no correction for clustering was needed.
8. Student-level standard deviations were not available for the study. Teaching group standard deviations for the intervention and comparison groups are included in A5.5.

Appendix A6 Summary of subscale findings for follow-up study findings for the early reading/writing domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Sound Foundations* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Sound Foundations* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Byrne, Fielding-Barnsley, & Ashley, 2000 (randomized controlled trial with attrition problems)[7]								
Castles' list–nonwords	11 year olds	103	–0.23 (1.19)	–0.47 (1.03)	0.24	0.21	ns	+8
Castles' list–regular words	11 year olds	103	–0.07 (0.99)	–0.32 (1.06)	0.25	0.24	ns	+10
Castle's list–irregular words	11 year olds	103	–0.09 (1.11)	–0.52 (1.08)	0.43	0.39	Statistically significant	+15

ns = not statistically significant

1. This appendix reports subscale findings for six-year follow-up findings (the intervention was implemented when the children were in preschool; the data in this table were collected when children were in grade 5) for measures that fall in the early reading/writing domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Byrne et al. (2000), no correction for clustering was needed.

Appendix A7.1 *Sound Foundations* rating for the phonological processing domain

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of phonological processing, the WWC rated *Sound Foundations* as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Sound Foundations* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. The one study that met WWC evidence standards found a statistically significant and positive effect in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important negative effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined effects on phonological processing.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The study did not show statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A7.2 *Sound Foundations* rating for the early reading/writing domain

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of early reading/writing, the WWC rated *Sound Foundations* as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Sound Foundations* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. The one study that met WWC evidence standards found a statistically significant and positive effect in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important negative effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined effects on early reading/writing.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The study did not show statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A8 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Centers	Children	
Oral language	0	0	0	na
Print knowledge	0	0	0	na
Phonological processing	1	4	126	Small
Early reading/writing	1	4	126	Small
Cognition	0	0	0	na
Math	0	0	0	na

na = not applicable/not studied

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."