What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Early Childhood Education September 2009[1]

Let's Begin with the Letter People®

Program Description[2]

Let's Begin with the Letter People® is an early education curriculum that uses 26 thematic units to develop children's language and early literacy skills. A major focus is phonological awareness, including rhyming, word play, alliteration, and segmentation. Children are encouraged to learn as individuals, in small groups, and in a whole-class environment. Teacher resource books and a set of classroom books and other program materials are available as a program kit.

Research[3]

Two studies of *Let's Begin with the Letter People*® meet What Works Clearinghouse (WWC) evidence standards and no studies meet WWC evidence standards with reservations. The two studies include children in 49 classrooms in 25 preschools in Houston, Texas, and southeastern New York State.[4]

Based on these two studies, the WWC considers the extent of evidence for *Let's Begin with the Letter People*® to be medium to large for oral language and print knowledge and small for phonological processing and math. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Let's Begin with the Letter People*® in the early reading and writing or cognition domains.

Effectiveness

Let's Begin with the Letter People® was found to have no discernible effects on oral language, print knowledge, phonological processing, or math.

	Oral language	*Print knowledge*	*Phonological processing*	*Early reading and writing*	*Cognition*	*Math*
Rating of effectiveness	No discernible effects	No discernible effects	No discernible effects	na	na	No discernible effects
Improvement index[5]	Average: +1 percentile point Range: –1 to +3 percentile points	Average: +6 percentile points Range: +1 to +12 percentile points	Average: –5 percentile points	na	na	Average: +3 percentile points Range: –4 to +8 percentile points

na = not applicable

1. This report has been updated to include reviews of two studies that were released since 2007, a review of one study that was released in 2002 but was not reviewed for the previous report, and a rereview of two studies that were included in the previous report. Of the five studies, two meet evidence standards and three were within the scope of the protocol but did not meet evidence standards. A complete list and disposition of all studies reviewed are provided in the references. The findings described in the previous *Let's Begin with the Letter People*® intervention report were based in part on a study by Assel et al. (2007). A rereview of that study for the present report revealed that the subcluster attrition rate of children exceeded standards, as specified in the Early Childhood Education protocol. Hence, results from the Assel et al. (2007) study were not considered when preparing the present intervention report.
2. The descriptive information for this program was obtained from publicly available sources: the program's website (www.abramslearningtrends.com/lets_begin_with_letter_people.aspx, downloaded July 2009) and the research literature (Assel et al., 2007; Fischel et al., 2007). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
3. The studies in this report were reviewed using WWC Evidence Standards, Version 1.0 (see the WWC Standards).
4. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
5. These numbers show the average and range of student-level improvement indices for all findings across the studies.

Absence of conflict of interest

The PCER Consortium (2008) study summarized in this intervention report had numerous contributors, including staff of Mathematica Policy Research, Inc. (MPR). Because the principal investigator for the WWC Early Childhood Education review is also an MPR staff member, the study was rated by Chesapeake Research Associates, who also prepared the intervention report. The report was then reviewed by the principal investigator, a WWC Quality Assurance reviewer, and an external peer reviewer.

Additional program information

Developer and contact

Abrams Learning Trends is the developer and distributor for *Let's Begin with the Letter People*®. Address: P.O. Box 10025, Waterbury, CT 06725. Email: customerservice@abramslearningtrends.com. Web: www.abramslearningtrends.com. Telephone: (800) 227-9120.

Scope of use

According to the developer, approximately 750,000 children have used the full program since its initial publication in 1999. Several million more children have used parts of the curriculum to supplement other preschool curricula. Information is not available on the number of centers and classrooms using this program or on the demographics of children using this program.

Teaching

Let's Begin with the Letter People® includes 26 units arranged around five thematically organized Teacher Resource books: All About Me; Animals, Animals, Animals; Everyone Has Needs; Getting Along with Others; and Nature All Around Us. Each of the Teacher Resource books offers varied teaching strategies and suggested activities. The units in each book have a Classroom Floor Plan Model, which includes suggestions for Interest Centers (individual and small-group time) and Meeting Circle (whole-class time), providing teachers with a number of choices for teaching knowledge and skills in language and literacy, as well as in science, math, art, music, social development, and motor skills. Through the Interest Centers, children are able to explore, investigate, construct, and apply knowledge. Skills are integrated in the classroom's daily events and are taught using a number of materials such as Letter People Huggables®, children's literature, Big Books, Little Books and story tapes, songs and rhymes, Just Listen™ computer program, Ready to Read PREdecodable books, Me Bag™ (for sharing special items), Letter People Stickables™, Puppet Patterns, and Family Activity Pages. Teachers introduce concepts during Meeting Circle time that are then explored in the Interest Centers and other group activities. For instance, the Letter People Huggables® (for example, Mr. N) are used to introduce letters, sounds, stories, colors, shapes, and characteristics. Blueprint for Learning, the program guide for *Let's Begin with the Letter People*®, provides an overview of the program and its components and includes information teachers can use for setting up their classrooms, as well as various instructional strategies. Teachers are trained during professional development activities and with other resources such as the Teacher Resource books.

Cost

Let's Begin with the Letter People® products can be purchased separately or in various combinations. The introductory set is available for $1,930 and includes the Teacher Resource File ($695), Letter People Huggables® ($585), Meeting and Greeting Cards ($110), Let's Sing with the Letter People CD ($30), Big and Little Books ($338), a read-along CD ($30), and a set of Letter People Virtual Books ($199). Packages that include additional components at extra cost are also available. Information about the cost of professional development is not available. Additional pricing information for separate products from the *Let's Begin with the Letter People*® is available on the website: www.abramslearningtrends.com/lets_begin_with_letter_people.aspx.

Research

Five studies reviewed by the WWC investigated the effects of *Let's Begin with the Letter People*®. Two studies (Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller, 2007; PCER Consortium, 2008) are randomized controlled trials that meet WWC evidence standards. The remaining three studies do not meet WWC evidence standards.

Meets evidence standards

One study reviewed by the WWC (PCER Consortium, 2008) assessed the effectiveness of *Let's Begin with the Letter People*® as part of the Preschool Curriculum Evaluation Research (PCER) effort.[6] The PCER Consortium (2008) used a randomized controlled trial design in which 19 preschool programs in Houston, Texas, were randomly assigned to implement *Let's Begin with the Letter People*®, to implement Doors to Discovery™, or to a control group. For the *Let's Begin with the Letter People*® versus control study, data were collected on 184 children (95 *Let's Begin with the Letter People*® and 89 control). Fifty-five percent of the children were male; 43% were Hispanic, 30% were Caucasian, and 13% were African-American; and 12% were reported to have a disability. Pretest data were collected in the fall and posttest data were collected in the spring of the preschool year. Follow-up kindergarten data were collected in the spring of the following year. The study investigated effects on oral language, print knowledge, phonological processing, and math. The control condition consisted of teacher-developed, nonspecific curricula with a focus on basic school readiness.

A second study (Fischel et al., 2007) examined the effectiveness of *Let's Begin with the Letter People*® and the Waterford Early Reading™ Level One curricula using a randomized controlled trial design in 27 full-day Head Start classrooms in six Head Start centers in southeastern New York State. In each of the two intervention conditions, one of the experimental curricula was used in conjunction with the High/Scope® program that had been used by the Head Start programs for more than 10 years. Control classrooms used only the High/Scope® program. Classrooms were randomly assigned in each of three years, and data were collected in each year. In the first year, three classrooms were assigned to each group. In the second year, intervention group classrooms continued in the same group, the control group classrooms were randomly assigned to an intervention group, and eight new classrooms were randomly assigned to the three groups. In the third year, intervention group classrooms again continued in the same group, control group classrooms were randomly assigned to an intervention group, and three new classrooms formed the control group. The original number of children in these classrooms over the three years was 507. The mean age was 4.3 years at pretest. Participating children were 42% African-American, 41% Hispanic, 8% multiracial, 7% Caucasian, and 2% other race/ethnicity. Approximately 14% of the children were identified as Spanish-language dominant at the outset of the study. Pretest data were collected in the fall and posttest data were collected in the spring of the preschool year. The study investigated effects on oral language and print knowledge. The WWC includes the data from children participating in classrooms that were newly-randomized to study groups in each year, because the classrooms that continued in the same study group in later years might have been chosen by some parents because of their curriculum. The WWC thus includes data for 132 children in eight *Let's Begin with the Letter People*® classrooms and 149 children in 11 control classrooms over the three-year period.

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across

6. The Preschool Curriculum Evaluation Research Consortium (2008) evaluated a total of 14 preschool curricula, including *Let's Begin with the Letter People*®, in comparison to the respective control conditions.

the studies that meet WWC evidence standards with or without reservations.[7]

The WWC considers the extent of evidence for *Let's Begin with the Letter People*® to be medium to large for oral language and print knowledge, and small for phonological processing and math. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Let's Begin with the Letter People*® in the early reading and writing or cognition domains.

Effectiveness

Findings

The WWC review of interventions for Early Childhood Education addresses child outcomes in six domains: oral language, print knowledge, phonological processing, early reading and writing, cognition, and math. The studies included in this report cover four domains: oral language, print knowledge, phonological processing, and math. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Let's Begin with the Letter People*® on children.[8]

Oral language. The PCER Consortium (2008) analyzed the effectiveness of *Let's Begin with the Letter People*® on oral language using the Peabody Picture Vocabulary Test–III (PPVT-III) and the Test of Language Development–Primary: III (TOLD-P:3). The authors report, and the WWC confirms, that differences between *Let's Begin with the Letter People*® and the control group are not statistically significant or substantively important (that is, an effect size of at least 0.25) on any of these measures. Fischel et al. (2007) analyzed the effectiveness of *Let's Begin with the Letter People*® on oral language using two measures: the PPVT-III and Comprehension. The authors report, and the WWC confirms, that differences between the *Let's Begin with the Letter People*® and the control groups on these measures are not statistically significant or large enough to be substantively important. According to WWC criteria, these two studies show no discernible effects on oral language.

Print knowledge. The PCER Consortium (2008) analyzed the effectiveness of *Let's Begin with the Letter People*® on the Test of Early Reading Ability (TERA-3), the Woodcock-Johnson–III (WJ-III) Letter-Word Identification subtest, and the WJ-III Spelling subtest. The authors report, and the WWC confirms, that differences between the *Let's Begin with the Letter People*® and control groups are not statistically significant or large enough to be substantively important on any of these measures. Fischel et al. (2007) analyzed the effectiveness of *Let's Begin with the Letter People*® on print knowledge using six measures: the Woodcock-Johnson–Revised (WJ-R) Letter-Word Identification subtest, the WJ-R Dictation subtest, the Get Ready to Read! screening test, Letters Known, Book Knowledge, and Print Conventions. The study reports significant differences favoring *Let's Begin with the Letter People*® on two measures: the WJ-R Dictation subtest and the Get Ready to Read! screening test. The WWC was unable to confirm statistically significant findings for any outcomes in this domain. Furthermore, the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria. According to WWC criteria, these two studies show no discernible effects on print knowledge.

7. 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 the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Let's Begin with the Letter People* is in Appendix A6.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. No correction for clustering was needed for the study by the PCER Consortium (2008) because its analysis corrected for clustering by using HLM, but a correction for multiple comparisons was needed, so the significance levels in this report may differ from those reported in the original study. In the case of Fischel et al. (2007), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Effectiveness *(continued)*

Phonological processing. The PCER Consortium (2008) analyzed the effectiveness of *Let's Begin with the Letter People*® on phonological processing using the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest. The authors report, and the WWC confirms, that differences between the *Let's Begin with the Letter People*® and control groups are not statistically significant or substantively important on any of these measures. According to WWC criteria, this study shows no discernible effects on phonological processing.

Math. The PCER Consortium (2008) analyzed the effectiveness of *Let's Begin with the Letter People*® on math using the WJ-III Applied Problems subtest, the Child Math Assessment–Abbreviated, and Shape Composition task. The authors report, and the WWC confirms, that differences between the *Let's Begin with the Letter People*® and control groups are not statistically significant or large enough to be substantively important on any of these measures. According to WWC criteria, this study shows no discernible effects on math.

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 (as calculated by the WWC), the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Procedures and Standards Handbook, Appendix E).

The WWC found *Let's Begin with the Letter People*® to have no discernible effects on oral language, print knowledge, phonological processing, and math

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 WWC Procedures and Standards Handbook, Appendix F). 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 for the intervention report.

Based on two studies, the average improvement index for *Let's Begin with the Letter People*® on four measures of oral language is +1 percentile point with a range of –1 to +3 percentile points across findings. The average improvement index for nine measures of print knowledge is +6 percentile points across the two studies, with a range of +1 to +12 percentile points across findings. Based on one study, the average improvement index for *Let's Begin with the Letter People*® on one measure of phonological processing is –5 percentile points, and the average improvement index on three measures of math is +3 percentile points, with a range of –4 to +8 percentile points.

Summary

The WWC reviewed five studies of *Let's Begin with the Letter People*®. Two of these studies meet WWC evidence standards. Three studies do not meet either WWC evidence standards or eligibility screens. Based on the two studies, the WWC found no discernible effects of *Let's Begin with the Letter People*® on oral language, print knowledge, phonological processing, and math. The conclusions presented in this report may change as new research emerges.

References

Meets WWC evidence standards

Fischel, J. E., Bracken, S. S., Fuchs-Eisenberg, A., Spira, E. G., Katz, S., & Shaller, G. (2007). Evaluation of curricular approaches to enhance preschool early literacy skills. *Journal of Literacy Research*, *39*(4), 471–501.

Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). Doors to Discovery and *Let's Begin with the Letter People.* In *Effects of preschool curriculum programs on school readiness* (pp. 85–98). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Studies that fall outside the Early Childhood Education protocol or do not meet WWC evidence standards

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2007). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading & Writing, 20*(5; 5), 463–494. The study does not meet WWC evidence standards because the overall attrition rate exceeds WWC standards for this area.

Hayes, K., Maddahian, E., & Fernandez, A. (2002). *An evaluation of pre-K reading programs* (Planning, Assessment, and Research Division Publication No. 137). Los Angeles: Los Angeles United School District. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Judkins, D., St. Pierre, R., Gutmann, B., Goodson, B., von Glatz, A., Hamilton, J., et al. (2008). *A study of classroom literacy interventions and outcomes in Even Start* (NCEE 2008-4028). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Appendix

Appendix A1.1 Study characteristics: Preschool Curriculum Evaluation Research Consortium, 2008 (randomized controlled trial)

Characteristic	Description
Study citation	Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). Doors to Discovery and *Let's Begin with the Letter People*. In *Effects of preschool curriculum programs on school readiness* (pp. 85–98). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
Participants	The study, conducted during the 2003–04 and 2004–05 school years, included two intervention groups (*Let's Begin with the Letter People*® and Doors to Discovery™) and a control group. Nineteen full-day Head Start and public prekindergarten preschools were recruited for the study. From these 19 preschools, 95 teachers/classrooms were recruited, of which 76 were included in random assignment. The researchers randomly assigned the 19 preschools to three treatment conditions (*Let's Begin with the Letter People*®, Doors to Discovery™, and control) with all classrooms within a preschool being assigned to the same treatment condition. The resulting sample of teachers/classrooms included 24 *Let's Begin with the Letter People*® classrooms, 25 Doors to Discovery™ classrooms, and 27 control classrooms. Forty-five of the 76 classrooms were then randomly selected to participate in the PCER study. One of the 45 classrooms dropped out, leaving 15 *Let's Begin with the Letter People*®, 14 Doors to Discovery™, and 15 control classrooms. Seven children (whose parents had provided consent to participate in the study) were randomly selected from each classroom for a total of 308 children. The parental consent rate was 65% for the treatment group (combined *Let's Begin with the Letter People*® and Doors to Discovery™) and 55% for the control group. The total number of participating children in the study at baseline was 297 (100 *Let's Begin with the Letter People*®, 101 Doors to Discovery™, and 96 control). At baseline, children in the study averaged 4.6 years of age; 55% were male; and 43% were Hispanic, 30% were white, and 13% were African-American. The analysis sample for the *Let's Begin with the Letter People*® study included 184 children (95 *Let's Begin with the Letter People*® and 89 control). Depending on the outcome, child-level attrition ranged from 6% to 7%.
Setting	The *Let's Begin with the Letter People*® study was conducted with children in 19 preschools in Houston, Texas. The sample included 30 Head Start and public prekindergarten (Title I and non-Title I) classrooms (15 *Let's Begin with the Letter People*® and 15 control).
Intervention	*Let's Begin with the Letter People*® is a comprehensive prekindergarten curriculum that is organized thematically. Literacy learning is integrated across topic areas, including science, health and safety, art, math, spatial concepts, and music, as well as development of large and small motor skills. The curriculum focuses on literacy and language skills, including oral language, phonological and phonemic awareness, and letter knowledge. The curriculum lessons address the development of letter knowledge in various contexts (for example, circle time, small group, large group) and activities (for example, center activities, story times). Classroom practices include teacher-directed activities, application of skills, and independent practice with activities that are tied to the curriculum. The *Let's Begin with the Letter People*® classroom includes interest centers (for example, Paint Corner, Blocks, Drama Center, Mathematics). Curriculum materials include Letter People Huggables®. Each Letter Person represents a letter of the alphabet and has distinguishing characteristics that are readily associated with the sound represented by the letter. Each classroom's fidelity to the curriculum was rated on a four-point scale ranging from "not at all" (0) to "high" (3). The average score for *Let's Begin with the Letter People*® classrooms was 1.86 on this measure.
Comparison	Business-as-usual using teacher-developed, nonspecific curricula. Control teachers' classrooms were rated with the same fidelity measure used in the *Let's Begin with the Letter People*® classrooms, which ranged from 0 to 3. The average score for the control classrooms using this measure was 1.0.

(continued)

Appendix A1.1 Study characteristics: Preschool Curriculum Evaluation Research Consortium, 2008 (randomized controlled trial) *(continued)*

Characteristic	Description
Primary outcomes and measurement	The outcome domains assessed were children's oral language, print knowledge, phonological processing, and math. Oral language was assessed with the Peabody Picture Vocabulary Test–III (PPVT-III) and the Test of Language Development–Primary: III (TOLD-P:3) Grammatic Understanding subtest. Print knowledge was assessed with the Test of Early Reading Ability–III (TERA-3), the Woodcock-Johnson–III (WJ-III) Letter-Word Identification subtest, and the WJ-III Spelling subtest. Phonological processing was assessed with the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest. Math was assessed with the WJ-III Applied Problems subtest, the Child Math Assessment–Abbreviated (CMA-A), and the Shape Composition task. For a more detailed description of these outcome measures, see Appendices A2.1–A2.4.
Staff/teacher training	Teachers received curriculum training prior to the start of the 2003–04 school year. This was the second year of implementation of the treatment, and most of the teachers had been trained prior to the start of the 2002–03 school year. New teachers each received 12 hours of training, and returning teachers each received 6 hours of training.

Appendix A1.2 Study characteristics: Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller, 2007 (randomized controlled trial)

Characteristic	Description
Study citation	Fischel, J. E., Bracken, S. S., Fuchs-Eisenberg, A., Spira, E. G., Katz, S., & Shaller, G. (2007). Evaluation of curricular approaches to enhance preschool early literacy skills. *Journal of Literacy Research, 39*(4), 471–501.
Participants	Twenty-seven unique Head Start preschool classrooms in six centers were randomly assigned to one of two curricular approaches overlaid onto their standard curriculum (High/Scope® Educational Approach) and a business-as-usual control group that used only the High/Scope® Educational Approach over three years of the study.[1] The two curricula were *Let's Begin with the Letter People®* and the Waterford Early Reading™ Level One. In year one of the study, three classrooms were assigned to *Let's Begin with the Letter People®* and three to the control group. In year two, the three control group classrooms from year one were randomly assigned to one of the intervention groups, and an additional eight new classrooms joined the study and were randomly assigned to groups; thus, three new classrooms were assigned to *Let's Begin with the Letter People®* and five new classrooms participated as the control group. In addition, two randomly selected *Let's Begin with the Letter People®* classrooms from year one participated again in year two. In year three, the five control classrooms from the previous year were randomly assigned to one of the intervention groups and three new classrooms participated as the control group. Thus, two new classrooms were assigned to *Let's Begin with the Letter People®* and three to the control group, and two randomly selected *Let's Begin with the Letter People®* classrooms from previous years also participated in the study.[2] Classrooms were divided as follows: *Let's Begin with the Letter People®* (12), Waterford Early Reading™ Level One (12), and comparison (11). A total of 507 children attending full-day classes for five days a week participated in the study. Children in the study sample had a mean age of 4 years, 4 months at the time of pretest. The sample of children included African-American (42%), Hispanic (41%), multiracial (8%), Caucasian (7%), and other race/ethnicity (2%). Approximately 14% of the total sample were Spanish-language dominant at Head Start entry.
Setting	The study was conducted in 27 unique Head Start classrooms in six centers in southeastern New York State (four centers in year one; one additional center in year two; and one additional center in year three). All centers were part of the same Head Start grantee.
Intervention	The intervention group classrooms used the *Let's Begin with the Letter People®* curriculum overlaid on the existing High/Scope® curriculum, which all programs had used for at least 10 years before the study. *Let's Begin with the Letter People®* is a comprehensive prekindergarten curriculum that is organized thematically. Literacy learning is integrated across topic areas, including science, health and safety, art, math, spatial concepts, and music, as well as development of large and small motor skills. The curriculum focuses on literacy and language skills, including oral language, phonological and phonemic awareness, and letter knowledge. The curriculum lessons address the development of letter knowledge in various contexts (for example, circle time, small group, large group) and activities (for example, center activities, story times). Classroom practices include teacher-directed activities, application of skills, and independent practice with activities that are tied to the curriculum. *Let's Begin with the Letter People®* classroom is based on centers (for example, Paint Corner, Block, Drama Center, Mathematics). Curriculum materials include Letter People Huggables®. Each Letter Person represents a letter of the alphabet and has distinguishing characteristics that are readily associated with the sound represented by the letter.
Comparison	The business-as-usual comparison group classrooms used the standard classroom curriculum (High/Scope®), which prescribes a daily routine (planning time, work time, cleanup time, time for recall, large-group time, small-group time, and outdoor play) and aligns well with Head Start's performance standards, focusing on language, literacy, and other school readiness skills, such as numeracy, reasoning, problem solving, and decisionmaking.

1. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Let's Begin with the Letter People®* group to the business-as-usual comparison group; however, results for the comparison between the curricula are included in Appendices A4.5 and A4.6. The WWC includes the Waterford Early Reading™ Level One versus business-as-usual comparison in a separate Waterford Early Reading™ Level One intervention report. Both intervention groups used the studied intervention in conjunction with the High/Scope® curriculum, which was the standard curriculum used by the classrooms prior to the study.
2. The same process yielded three Waterford Early Reading™ Level One classrooms in year one, five Waterford Early Reading™ Level One classrooms (three new classrooms and two repeat classrooms) in year two, and four Waterford Early Reading™ Level One classrooms (two new classrooms and two repeat classrooms) in year three. The WWC includes the data only from children in classrooms new to their study group in each year because families of children attending in the second year of the intervention may have selected the classroom because of its curriculum.

(continued)

Appendix A1.2 Study characteristics: Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller, 2007 (randomized controlled trial) *(continued)*

Characteristic	Description
Primary outcomes and measurement	The outcome domains assessed were children's oral language and print knowledge. Oral language was assessed with the Peabody Picture Vocabulary Test–III (PPVT-III; a standardized measure) and the Story Comprehension subtest (a nonstandardized measure). Print knowledge was assessed using six measures: the Woodcock-Johnson–Revised (WJ-R) Letter-Word Identification subtest, the WJ-R Dictation subtest (both standardized measures), the Get Ready to Read! (GRTR) screening instrument (a nonstandardized measure), and the Letter Knowledge, Book Knowledge, and Print Conventions subtests of the storybook assessment developed for the Head Start FACES study (nonstandardized measures). For a more detailed description of these outcome measures, see Appendices A2.1–A2.4.
Staff/teacher training	Teachers and teacher assistants in the *Let's Begin with the Letter People®* group participated in a three-day curriculum training each August conducted by a professional trainer from Abrams and Company (the developer and distributor of this curriculum). The trainer visited each classroom in the *Let's Begin with the Letter People®* condition in the fall and spring of each intervention year and provided individual feedback to teachers. Fidelity was measured during these visits using a checklist to assess the degree of implementation in two domains: Classroom Organization and Teacher Behavior. Implementation by all teachers in each year of the study was determined to be accurate and appropriate. Fischel et al. (2007) reported that additional training was offered by the trainer; however, details of the frequency, content, or degree of participation in these trainings were not provided. Teachers and assistants in the *Let's Begin with the Letter People®* group and the business-as-usual comparison group participated in a week-long in-service High/Scope® curriculum training at the beginning of the school year. Support was provided in the classroom by educational and child development specialists throughout the school year.

Appendix A2.1 Outcome measures for the oral language domain

Outcome measure	Description
Peabody Picture Vocabulary Test–III (PPVT-III)	A standardized measure of children's receptive vocabulary in which children show understanding of a spoken word by pointing to a picture that best represents the meaning (as cited in Fischel et al., 2007, and PCER Consortium, 2008).
Test of Language Development–Primary: III (TOLD-P:3) Grammatic Understanding subtest	A standardized measure of children's ability to comprehend the meaning of sentences by selecting pictures that most accurately represent the sentence (as cited in PCER Consortium, 2008).
Comprehension	An adaptation of prereading assessments developed by Clay (1979), Teale (1988, 1990), and Mason and Stewart (1989) for use in the Family and Child Experiences Study (FACES). The child is handed a storybook and asked a series of questions about the characters and plot (as cited in Fischel et al., 2007).

Appendix A2.2 Outcome measures for the print knowledge domain

Outcome measure	Description
Test of Early Reading Ability–III (TERA-3)	A standardized measure of children's developing reading skills with three subtests: Alphabet, Conventions, and Meaning (as cited in PCER Consortium, 2008).[1]
Woodcock Johnson–III (WJ-III) Letter-Word Identification subtest	A standardized measure of identification of letters and reading of words (as cited in PCER Consortium, 2008).
Woodcock-Johnson–Revised (WJ-R) Letter-Word Identification subtest	A standardized measure of identification of letters and reading of words (as cited in Fischel et al., 2007).
Woodcock-Johnson–III (WJ-III) Spelling subtest	A standardized measure that assesses children's prewriting skills, such as drawing lines, tracing, and writing letters (as cited in PCER Consortium, 2008).
Woodcock-Johnson–Revised (WJ-R) Dictation subtest	A standardized measure that assesses children's prewriting skills, such as drawing lines, tracing, and writing letters (as cited in Fischel et al., 2007).
Get Ready to Read! (GRTR) Screen	A 20-question, nonstandardized screening test designed to measure emergent writing skills (identifying best picture exemplars), linguistic awareness (rhyming, segmenting words, and deletion of sounds), and print knowledge (differentiating print from pictures, letter naming, and identifying letter sounds) (as cited in Fischel et al., 2007).
Letters Known	A test developed for FACES requiring the child to identify as many letters as possible from uppercase arrays of letters (as cited in Fischel et al., 2007).
Book Knowledge	A test developed for FACES in which a child is handed a book that is inverted and backwards and asked a series of questions about book knowledge (for example, where is the front of the book, demonstrate how to open a book, and locate the title and the author's name) (as cited in Fischel et al., 2007).
Print Conventions	A test developed for FACES using a storybook, in which a child is asked to identify print conventions (for example, reading from left to right and top to bottom, location of page start, page turning) (as cited in Fischel et al., 2007).

1. By name, this measure sounds as if it should be captured under the Early Reading and Writing domain; however, the description of the measure identifies constructs that are pertinent to Print Knowledge, such as knowing the alphabet, understanding print conventions, and environmental print.

Appendix A2.3 Outcome measures for the phonological processing domain

Outcome measure	Description
Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest	A measure of children's ability to identify and manipulate sounds in spoken words, using word prompts and picture plates for the first nine items and word prompts only for later items (as cited in PCER Consortium, 2008).

Appendix A2.4 Outcome measures for the math domain

Outcome measure	Description
Woodcock-Johnson–III (WJ-III) Applied Problems subtest	A standardized measure of children's ability to solve numerical and spatial problems, presented verbally with accompanying pictures of objects (as cited in PCER Consortium, 2008).
Child Math Assessment–Abbreviated (CMA-A) Composite Score	The average of four subscales: (1) solving addition and subtraction problems using visible objects, (2) constructing a set of objects equal in number to a given set, (3) recognizing shapes, and (4) copying a pattern using objects that vary in color and identity from the model pattern (as cited in PCER Consortium, 2008).
Building Blocks, Shape Composition task	Modified for PCER from the Building Blocks assessment tools. Children use blocks to fill in a puzzle and are assessed on whether they fill the puzzle without gaps or hangovers (as cited in PCER Consortium, 2008).

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®* - comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
PPVT-III	Preschoolers	30/184	90.72 (19.18)	91.33 (18.12)	–0.61	–0.03	ns	–1
TOLD-P:3 Grammatic Understanding	Preschoolers	30/184	9.74 (2.73)	9.33 (2.71)	0.41	0.08	ns	+3
Average for oral language (PCER Consortium, 2008)[9]						**0.03**	**ns**	**+1**
Fischel et al., 2007 (randomized controlled trial)[8, 10]								
PPVT-III	Preschoolers	19/272	86.59 (13.80)	85.72 (13.68)	0.87	0.06	ns	+3
Comprehension	Preschoolers	19/277	0.89 (0.77)	0.90 (0.74)	–0.01	–0.01	ns	–1
Average for oral language (Fischel et al., 2007)[9]						**0.02**	**ns**	**+1**
Domain average for oral language across all studies[9]						**0.03**	**na**	**+1**

ns = not statistically significant
na = not applicable
PPVT-III = Peabody Picture Vocabulary Test–III
TOLD-P:3 = Test of Language Development–Primary-III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language domain. Kindergarten follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.1.
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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.

(continued)

Appendix A3.1 Summary of study findings included in the rating for the oral language domain[1] *(continued)*

5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the authors (Cohen's *d* based on a repeated measures analysis).
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant. In the case of Fischel et al. (2007), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.
10. The WWC analysis of Fischel et al. (2007) focused on teachers new to their study group, whereas the original study reported findings based on analysis of new and experienced teachers; this also may cause the significance levels reported to differ from those reported in the original study. The child-level posttest sample sizes, and posttest means and standard deviations, were provided by the study authors upon WWC request.

Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
TERA-3	Preschoolers	30/183	92.94 (16.06)	92.76 (17.86)	0.18	0.02	ns	+1
WJ-III Letter-Word	Preschoolers	30/184	108.72 (12.54)	106.04 (13.82)	2.68	0.10	ns	+4
WJ-III Spelling	Preschoolers	30/184	101.34 (13.01)	97.37 (12.63)	3.97	0.17	ns	+7
Average for print knowledge (PCER Consortium, 2008)[9]						**0.10**	**ns**	**+4**
Fischel et al., 2007 (randomized controlled trial)[8, 10]								
WJ-R Letter-Word Identification	Preschoolers	19/235	98.08 (12.06)	96.69 (11.90)	1.39	0.12	ns	+5
WJ-R Dictation	Preschoolers	19/194	93.48 (15.48)	88.93 (15.03)	4.55	0.30	ns	+12
Get Ready to Read! Screen	Preschoolers	19/281	12.62 (3.70)	11.59 (3.83)	1.03	0.27	ns	+11
Letters Known	Preschoolers	19/277	17.80 (9.01)	15.86 (9.68)	1.94	0.21	ns	+8
Book Knowledge	Preschoolers	19/277	2.85 (1.37)	2.53 (1.27)	0.32	0.24	ns	+10

(continued)

Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain[1] *(continued)*

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Print Conventions	Preschoolers	19/277	0.43 (0.74)	0.27 (0.60)	0.16	0.24	ns	+9
Average for print knowledge (Fischel et al., 2007)[9]						**0.23**	**ns**	**+9**
Domain average for print knowledge across all studies[9]						**0.16**	**na**	**+6**

ns = not statistically significant
na = not applicable
TERA-3 = Test of Early Reading Ability–III
WJ-III = Woodcock-Johnson–III
WJ-R = Woodcock-Johnson–Revised

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge domain. Kindergarten follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.2.
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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the authors (Cohen's *d* based on a repeated measures analysis).
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant. In the case of Fischel et al. (2007), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.
10. The WWC analysis of Fischel et al. (2007) focused on teachers new to their study group, whereas the original study reported findings based on analysis of new and experienced teachers; this also may cause the significance levels reported to differ from those reported in the original study. The child-level posttest sample sizes, and posttest means and standard deviations, were provided by the study authors upon WWC request.

Appendix A3.3 Summary of study findings included in the rating for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
Pre-CTOPPP Elision subtest	Preschoolers	30/184	9.35 (5.07)	10.11 (4.64)	–0.76	–0.13	ns	–5
Domain average for phonological processing[9]						**–0.13**	**ns**	**–5**

ns = not statistically significant

Pre-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processing

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the phonological processing domain. Kindergarten follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.3.
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.
3. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.
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 A3.4 Summary of study findings included in the rating for the math domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
WJ-III Applied Problems subtest	Preschoolers	30/184	96.75 (13.25)	99.28 (16.60)	–2.53	–0.10	ns	–4
CMA-A Composite	Preschoolers	30/184	0.69 (0.22)	0.65 (0.24)	0.04	0.15	ns	+6
Shape Composition	Preschoolers	30/184	1.92 (0.95)	1.72 (0.69)	0.20	0.21	ns	+8
Domain average for math[9]						**0.09**	**ns**	**+3**

ns = not statistically significant
WJ-III = Woodcock-Johnson–III
CMA-A = Child Math Assessment–Abbreviated

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math domain. Kindergarten follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.4.
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.
3. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.
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 A4.1 Summary of follow-up findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
PPVT-III	Kindergarten	nr/150	93.95 (18.03)	94.00 (16.01)	–0.05	0.00	ns	0
TOLD-P:3 Grammatic Understanding subtest	Kindergarten	nr/151	9.47 (3.12)	10.08 (2.80)	–0.61	–0.12	ns	–5

ns = not statistically significant
nr = not reported
PPVT-III = Peabody Picture Vocabulary Test–III
TOLD-P:3 = Test of Language Development Primary–III

1. This appendix presents follow-up findings considered for measures that fall in the oral language domain. End-of-preschool scores 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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A4.2 Summary of follow-up findings for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
TERA-3	Kindergarten	nr/151	92.65 (18.50)	93.96 (16.47)	–1.31	–0.13	ns	–5
WJ-III Letter-Word Identification	Kindergarten	nr/151	104.75 (13.44)	109.53 (13.57)	–4.78	–0.18	ns	–7
WJ-III Spelling	Kindergarten	nr/151	101.91 (15.68)	103.46 (13.14)	–1.55	–0.06	ns	–2

ns = not statistically significant
nr = not reported
TERA-3 = Test of Early Reading Ability–III
WJ-III = Woodcock-Johnson–III

1. This appendix presents follow-up findings for measures that fall in the print knowledge domain. End-of-preschool scores 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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A4.3 Summary of follow-up findings for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
CTOPP Elision subtest	Kindergarten	nr/151	4.52 (3.66)	5.04 (4.24)	–0.52	–0.13	ns	–5

ns = not statistically significant
nr = not reported
CTOPP = Comprehensive Test of Phonological Processing

1. This appendix presents follow-up findings for measures that fall in the phonological processing domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.3.
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.
3. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A4.4 Summary of follow-up findings for the math domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Let's Begin with the Letter People®*-comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
WJ-III Applied Problems	Kindergarten	nr/151	99.18 (13.42)	102.40 (11.38)	–3.22	–0.13	ns	–5
CMA-A Composite	Kindergarten	nr/151	0.70 (0.18)	0.72 (0.14)	–0.02	–0.07	ns	–3
Shape Composition	Kindergarten	nr/151	2.45 (0.77)	2.51 (0.69)	–0.06	–0.06	ns	–2

ns = not statistically significant
nr = not reported
WJ-III = Woodcock-Johnson–III
CMA-A = Child Math Assessment–Abbreviated

1. This appendix presents follow-up findings for measures that fall in the math domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.4.
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.
3. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A4.5 Summary of findings for comparisons between *Let's Begin with the Letter People®* and Waterford Early Reading™ Level One for the oral language domain[1]

Outcome measure	Study sample	Sample size[3] (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[4]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Waterford Early Reading™ Level One group[4]	WWC calculations: Mean difference[5] (*Let's Begin with the Letter People®* – Waterford Early Reading™ Level One)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Fischel et al., 2007 (randomized controlled trial)[9]								
PPVT-III	Preschoolers	16/241	86.59 (13.80)	86.92 (14.39)	–0.33	–0.02	ns	–1
Comprehension	Preschoolers	16/247	0.89 (0.77)	0.85 (0.76)	0.04	0.05	ns	+2

ns = not statistically significant

PPVT-III = Peabody Picture Vocabulary Test–III

1. This appendix reports findings for the head-to-head comparison of *Let's Begin with the Letter People®* and Waterford Early Reading™ Level One. Comparisons of *Let's Begin with the Letter People®* and the business-as-usual comparison group were used for rating purposes and are presented in Appendix A3.1. The WWC includes data from children participating in classrooms that had not participated in previous waves (that is, children from unique classrooms).
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. The child-level posttest sample sizes were provided by the study authors upon WWC request.
4. The posttest means are covariate-adjusted means provided by the study authors upon WWC request.
5. Positive differences and effect sizes favor the *Let's Begin with the Letter People®* group; negative differences and effect sizes favor the Waterford Early Reading™ Level One group.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
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 *Let's Begin with the Letter People®* condition and that of the average student in the Waterford Early Reading™ Level One condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the *Let's Begin with the Letter People®* group.
9. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Fischel et al. (2007), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Appendix A4.6 Summary of findings for comparisons between *Let's Begin with the Letter People®* and Waterford Early Reading™ Level One for the print knowledge domain[1]

Outcome measure	Study sample	Sample size[3] (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Let's Begin with the Letter People®* group[4]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Waterford Early Reading™ Level One group[4]	WWC calculations: Mean difference[5] (*Let's Begin with the Letter People®* – Waterford Early Reading™ Level One)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Fischel et al., 2007 (randomized controlled trial)[9]								
WJ-R Letter-Word Identification subtest	Preschoolers	16/208	98.08 (12.06)	98.69 (11.41)	–0.61	–0.05	ns	–2
WJ-R Dictation	Preschoolers	16/173	93.48 (15.48)	90.37 (14.28)	3.11	0.21	ns	+8
Get Ready to Read! Screen	Preschoolers	16/251	12.62 (3.70)	12.84 (3.87)	–0.22	–0.06	ns	–2
Letters Known	Preschoolers	16/247	17.80 (9.01)	18.03 (8.81)	–0.23	–0.03	ns	–1
Book Knowledge	Preschoolers	16/247	2.85 (1.37)	2.41 (1.37)	0.44	0.32	ns	+13
Print Conventions	Preschoolers	16/247	0.43 (0.74)	0.44 (0.77)	–0.01	–0.01	ns	–1

ns = not statistically significant
WJ-R = Woodcock-Johnson–Revised

1. This appendix reports findings for the head-to-head comparison of *Let's Begin with the Letter People®* and Waterford Early Reading™ Level One. Comparisons of *Let's Begin with the Letter People®* and the business-as-usual comparison group were used for rating purposes and are presented in Appendix A3.2. The WWC includes data from children participating in classrooms that had not participated in previous waves (that is, children from unique classrooms).
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. The child-level posttest sample sizes were provided by the study authors upon WWC request.
4. The posttest means are covariate-adjusted means provided by the study authors upon WWC request.
5. Positive differences and effect sizes favor the *Let's Begin with the Letter People®* group; negative differences and effect sizes favor the Waterford Early Reading™ Level One group.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
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 *Let's Begin with the Letter People®* condition and that of the average student in the Waterford Early Reading™ Level One condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the *Let's Begin with the Letter People®* group.
9. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. For Fischel et al. (2007), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Appendix A5.1 *Let's Begin with the Letter People*® rating for the oral language domain

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

For the outcome domain of oral knowledge, the WWC rated *Let's Begin with the Letter People*® as having no discernible effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. Neither of the two studies of *Let's Begin with the Letter People*® that measured oral language showed a statistically significant or substantively important effect, either positive or negative.

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. Neither of the two studies of *Let's Begin with the Letter People*® that measured oral language showed a statistically significant positive effect.

AND

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

  Met. Neither of the two studies of *Let's Begin with the Letter People*® that measured oral language showed a statistically significant or substantively important negative effect.

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.

  Not met. Neither of the two studies of *Let's Begin with the Letter People*® that measured oral language showed a statistically significant or substantively important positive effect.

AND

- 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.

  Not met. Neither of the two studies of *Let's Begin with the Letter People*® that measured oral language showed a statistically significant or substantively important negative effect. Neither study showed a statistically significant or substantively important positive effect on oral language.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

(continued)

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured oral language showed a statistically significant or substantively important positive or negative effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. Neither study of *Let's Begin with the Letter People®* that measured oral language showed a statistically significant or substantively important positive or negative effect.

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

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

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured oral language showed a statistically significant or substantively important negative effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured oral language showed a statistically significant or substantively important positive effect. Neither study showed a statistically significant or substantively important negative effect on oral language.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured oral language showed a statistically significant negative effect.

AND

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

Met. Neither of the two studies of *Let's Begin with the Letter People®* that measured oral language showed a statistically significant or substantively important positive effect.

Appendix A5.2 *Let's Begin with the Letter People*® rating for the print knowledge domain

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

For the outcome domain of print knowledge, the WWC rated *Let's Begin with the Letter People*® as having no discernible effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. Neither of the two studies of *Let's Begin with the Letter People*® that measured print knowledge showed a statistically significant or substantively important effect, either positive or negative.

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. Neither of the two studies of *Let's Begin with the Letter People*® that measured print knowledge showed a statistically significant positive effect.

AND

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

  Met. Neither of the two studies of *Let's Begin with the Letter People*® that measured print knowledge showed a statistically significant or substantively important negative effect.

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.

  Not met. Neither of the two studies of *Let's Begin with the Letter People*® that measured print knowledge showed a statistically significant or substantively important positive effect.

AND

- 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.

  Not met. Neither of the two studies of *Let's Begin with the Letter People*® that measured print knowledge showed a statistically significant or substantively important negative effect. Neither study showed a statistically significant or substantively important positive effect on print knowledge.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

(continued)

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured print knowledge showed a statistically significant or substantively important positive or negative effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured print knowledge showed a statistically significant or substantively important positive or negative effect.

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

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

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured print knowledge showed a statistically significant or substantively important negative effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured print knowledge showed a statistically significant or substantively important positive effect. Neither study showed a statistically significant or substantively important negative effect on print knowledge.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. Neither of the two studies of *Let's Begin with the Letter People®* that measured print knowledge showed a statistically significant negative effect.

AND

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

Met. Neither of the two studies of *Let's Begin with the Letter People®* that measured print knowledge showed a statistically significant or substantively important positive effect.

Appendix A5.3 *Let's Begin with the Letter People®* rating for the phonological processing domain

The WWC rates an intervention's effects 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 *Let's Begin with the Letter People®* as having no discernible effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important effect, either positive or negative.

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. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant positive effect.

AND

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

  Met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important negative effect.

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.

  Not met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important positive effect.

AND

- 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.

  Not met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important negative effect. No other studies measured phonological processing.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

(continued)

Appendix A5.3 *Let's Begin with the Letter People®* rating for the phonological processing domain *(continued)*

Not met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important positive or negative effect. No other studies measured phonological processing.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important positive or negative effect. No other studies measured phonological processing.

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

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

Not met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important negative effect. No other studies measured phonological processing.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important positive effect. No other studies measured phonological processing.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant negative effect. No other studies measured phonological processing.

AND

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

Met. The one study of *Let's Begin with the Letter People®* that measured phonological processing showed no statistically significant or substantively important positive effect. No other studies measured phonological processing.

Appendix A5.4 *Let's Begin with the Letter People®* rating for the math domain

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

For the outcome domain of math, the WWC rated *Let's Begin with the Letter People®* as having no discernible effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important effect, either positive or negative.

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. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant positive effect on math.

AND

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

  Met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important negative effects.

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.

  Not met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important positive effect.

AND

- 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.

  Not met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important negative effect.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

(continued)

Not met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important positive or negative effect. No other studies measured math.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important positive or negative effect. No other studies measured math.

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

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

Not met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important negative effect. No other studies measured math.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important positive effect. No other studies measured math.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant negative effect. No other studies measured math.

AND

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

Met. The one study of *Let's Begin with the Letter People®* that measured math showed no statistically significant or substantively important positive effect. No other studies measured math.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Preschool classrooms	Students[2]	
Oral language	2	49	456	Medium to large
Print knowledge	2	49	377	Medium to large
Phonological processing	1	30	184	Small
Early reading or writing	0	na	na	na
Cognition	0	na	na	na
Math	1	30	184	Small

na = not applicable/not studied

1. A rating of "medium 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."
2. The sample size of students shown in this table is based on the smallest number of children with valid posttest measurements within a domain. Posttest responses for the PCER (2008) study ranged from 183 to 184. Posttest responses for the Fischel et al. (2007) study ranged from 194 to 281, primarily because many Spanish-language-dominant children in the sample could not complete the standardized assessments in English.