What Works Clearinghouse

Literacy Express

Program description[1]

Literacy Express is a comprehensive preschool curriculum designed for three- to five-year-old children. It is structured around units on oral language, emergent literacy, basic math, science, general knowledge, and socio-emotional development. It can be used in half- or full-day programs with typically developing children and children with special needs. It provides professional development opportunities for staff, teaching materials, suggested activities, and recommendations for room arrangement, daily schedules, and classroom management.

Research

Two studies of *Literacy Express* met the What Works Clearinghouse (WWC) evidence standards.[2] The studies included more than 900 three- to five-year-old children attending preschools in Florida and California.[3] The WWC considers the extent of evidence for *Literacy Express* to be moderate to large for oral language, print knowledge, and phonological processing, and small for cognition and math. No studies that met the WWC evidence standards with or without reservations addressed early reading/writing.

Effectiveness

Literacy Express was found to have positive effects on print knowledge and phonological processing, potentially positive effects on oral language and math, and no discernible effects on cognition.

	Oral language	*Print knowledge*	*Phonological processing*	*Early reading/ writing*	*Cognition*	*Math*
Rating of effectiveness	Potentially positive effects	Positive effects	Positive effects	na	No discernible effects	Potentially positive effects
Improvement index[4]	Average: +14 percentile points Range: +12 to +18 percentile points	Average: +16 percentile points Range: +13 to +30 percentile points	Average: +17 percentile points Range: +6 to +29 percentile points	na	Average: +1 percentile point Range: -5 to +5 percentile points	Average: +18 percentile points Range: +14 to +23 percentile points

na = not applicable

1. The descriptive information for this program was obtained from publicly available sources: the research literature (Lonigan, 2006; Lonigan, Farver, Clancy-Menchetti, & Phillips, 2005) and from the developer. 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.
2. To be eligible for the WWC's review, the Early Childhood Education (ECE) intervention had to be implemented in English in center-based settings with children aged three to five or in preschool.
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available. *Literacy Express* is being studied under the Preschool Curriculum Evaluation Research (PCER) Grants administered through the U.S. Department of Education's Institute of Education Sciences. The final PCER reports were not released in time to be reviewed for this report.
4. These numbers show the average and range of student-level improvement indices for all findings across the studies.

Absence of conflict of interest

Literacy Express was developed in part by Dr. Lonigan, one of the two Principal Investigators for the WWC ECE review, and he has received income from sales of this curriculum. Dr. Lonigan was the primary author on both studies reviewed for this WWC intervention report. He also developed the P-CTOPPP, one of the outcome measures used in this report. Dr. Lonigan was not involved in the coding, reconciliation, or discussion of the included studies. Additionally, he was not involved in writing or reviewing the WWC intervention report. Dr. Kisker, the second Principal Investigator for the review, led all study review and report writing activities for *Literacy Express.*

Additional program information[1]

Developer and contact

Developed by Christopher J. Lonigan, Ph.D., Jeanine Clancy-Menchetti, Ph.D., Beth M. Phillips, Ph.D., and colleagues, *Literacy Express* is currently available in limited commercial distribution through *Literacy Express*. Email: LiteracyExpress@hotmail.com.

Scope of use

Literacy Express has been implemented by preschool and child-care programs in Florida, California, Texas, Massachusetts, and New Mexico. These programs have included typically developing children, children with special needs, and English language learners. However, information is not available on the number or demographics of children or centers using this program.

Teaching

Literacy Express can be implemented in various early childhood settings using daily individual, small-group, and large-group activities and a balance of teacher-initiated and child-initiated activities. The complete curriculum package includes a teacher's manual, 10 thematic unit guides, a unit guide for augmentative or summer activities ("Off to Kindergarten"), and key curriculum materials such as shape materials, more than 80 thematically linked picture books and alphabet books, and phonological awareness activity picture cards, letters, and numbers.

Literacy Express is structured around a number of three- to five-week thematic units that can be integrated into classroom activities. The units are sequenced in order of increasing complexity, each building on the previous one. In each unit, teachers use three specific and brief small-group activities daily (dialogic reading, phonological awareness activities, and print knowledge activities) with groups of three to five children. Teachers also select from a number of suggested large-group activities for each unit that involve active child participation to use skills learned in the small-group activities in new contexts. The curriculum includes both teacher- and child-directed math and science activities that follow a developmental scope and sequence. The curriculum also provides suggested activities—art, cooking, science and discovery, and gross and fine motor activities—for many independent play centers. Teachers participate in a two-day professional development workshop at the beginning of the school year and in either two additional half-day workshops or one additional full-day workshop.

Cost

The current cost for a complete *Literacy Express* classroom package is $2,300. Professional development fees vary by the size of group and number of trainers.

Research

Two studies reviewed by the WWC investigated the effects of *Literacy Express* in center-based settings. Both studies (Lonigan, 2006; Lonigan, Farver, Clancy-Menchetti, & Phillips, 2005) were randomized controlled trials that met WWC evidence standards.

Lonigan (2006) included 17 public preschool centers in northern Florida. Almost two-thirds of the children were African-American. The study author compared oral language, print knowledge, phonological processing, and math outcomes using data from year two of the study for children participating in a *Literacy Express* intervention group, a *DLM Express plus Open Court Pre-K* intervention group, or a business-as-usual comparison group.[5]

Lonigan et al. (2005) included 48 preschools from Tallahassee, Florida, and Los Angeles, California. A majority of the preschools were Head Start centers and more than three-quarters of the children were African-American or Hispanic. The study authors compared oral language, print knowledge, phonological processing, and cognition outcomes for children who participated in a *Literacy Express* with professional development via workshops group (the "workshop group") or a *Literacy Express* with professional development via workshops plus mentoring group (the "mentoring group"), or a business-as-usual comparison group.[6]

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate to large (see the What Works Clearinghouse Extent of Evidence Categorization Scheme). The extent of evidence takes into account the number of studies and the total sample size across the studies that met WWC evidence standards with or without reservations.[7]

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

5. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Literacy Express* intervention group with the business-as-usual comparison group. The WWC does not include the *DLM Express plus Open Court Pre-K* versus business-as-usual comparison in a separate WWC intervention report because the effects of *DLM Express* and *Open Court Pre-K* on children's outcomes cannot be disentangled. The WWC does not include the head-to-head comparison of *Literacy Express* and *DLM Express plus Open Court Pre-K*, but interested readers can examine that comparison using the data provided in the original article.
6. The study authors combined the two intervention groups, and the WWC used the combined data to determine the rating of effectiveness for this WWC intervention report. The study authors' analyses for the comparisons between the individual intervention groups and the business-as-usual comparison group are included in Appendices A4.1–A4.4 and the comparison between the mentoring and workshop groups is included in the report and in Appendices A5.1–A5.4.
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.

Effectiveness

Findings

The WWC review of interventions for early childhood education addresses children's outcomes in six domains: oral language, print knowledge, phonological processing, early reading/writing, cognition, and math. Lonigan (2006) and Lonigan et al. (2005) addressed outcomes in the oral language, print knowledge, phonological processing, cognition, and math outcome domains. The findings below present the authors' and the WWC-calculated estimates of the size and statistical significance of the effects of *Literacy Express* on children's performance.[8]

Oral language. Lonigan (2006) analyzed group differences between the *Literacy Express* and business-as-usual comparison groups for two measures of oral language—the Peabody Picture Vocabulary Test-III (PPVT-III) and the Preschool Comprehensive Test of Phonological and Print Processes (P-CTOPPP) Definitional Vocabulary subtest. The differences between the groups were not statistically significant for either outcome; however, the average effect size across measures was large enough to be considered substantively important according to WWC criteria (that is, at least 0.25).

Lonigan et al. (2005) analyzed group differences between the combined *Literacy Express* group and the business-as-usual comparison group for one measure of oral language—the Preschool Language Scales-IV (PLS-IV) Expressive Communication subscale. The difference between groups was statistically significant and favored children in the *Literacy Express* group.

Print knowledge. Lonigan (2006) found a statistically significant difference favoring children in the *Literacy Express* group on three of the five outcome measures assessed in this domain—the Test of Early Reading Ability-3 (TERA-3) Alphabet and Meaning subtests and the Woodcock-Johnson III (W-J III) spelling subtest. These effects were confirmed to be statistically significant by the WWC. The author found no statistically significant differences between the intervention and business-as-usual comparison groups on the other two measures (TERA-3 Print Conventions subtest and P-CTOPPP Print Knowledge subtest). The average effect size across measures was large enough to be considered substantively important according to WWC criteria (that is, at least 0.25).

Lonigan et al. (2005) found a statistically significant difference favoring children in the combined *Literacy Express* group on the one outcome measure assessed in this domain (P-CTOPPP Print Knowledge subtest).

Phonological processing. Lonigan (2006) analyzed group differences between the *Literacy Express* and business-as-usual comparison groups for two measures of phonological processing (P-CTOPPP Blending subtest and P-CTOPPP Elision subtest) and found statistically significant differences favoring the intervention group for both measures. The statistical significance of these effects was confirmed by the WWC.

Lonigan et al. (2005) analyzed group differences between the combined *Literacy Express* group and business-as-usual comparison groups for the same two measures of phonological processing and found statistically significant differences favoring the intervention group for the Elision subtest. The WWC confirmed the statistical significance of this effect. The authors found no statistically significant difference between groups on the Blending subtest; however, the average effect size across these two measures was large enough to be considered substantively important according to WWC criteria (that is, at least 0.25).

Cognition. Lonigan et al. (2005) found no statistically significant differences between the combined *Literacy Express* group and the business-as-usual comparison group for any of the three measures assessed in this domain (P-CTOPPP Non-Word Repetition subtest; P-CTOPPP Word Span subtest; P-CTOPPP Rapid Object Naming subtest). Additionally, the average effect size across

8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of *Literacy Express,* a correction for multiple comparisons was needed.

Effectiveness *(continued)*

measures was not large enough to be considered substantively important according to WWC criteria (that is, at least 0.25).

Math. Lonigan (2006) found a statistically significant difference favoring children in the *Literacy Express* group on one of the two outcome measures assessed in this domain—Children's Math Assessment (CMA), Abbreviated Version—and this effect was confirmed to be statistically significant by the WWC. The author found no statistically significant difference between the intervention and business-as-usual comparison groups on the W-J III Applied Problems subtest. However, the average effect size across these two measures was large enough to be considered substantively important according to WWC criteria (that is, at least 0.25).

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,[8] the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Intervention Rating Scheme).

The WWC found *Literacy Express* to have positive effects on print knowledge and phonological processing, potentially positive effects on oral language and math, and no discernible effects on cognition

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see Technical Details of WWC-Conducted Computations). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is based entirely on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analyses. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.

The average improvement index for oral language is +14 percentile points across the two studies, with a range of +12 to +18 percentile points across findings. The average improvement index for print knowledge is +16 percentile points across the two studies, with a range of +13 to +30 percentile points across findings. The average improvement index for phonological processing is +17 percentile points across the two studies, with a range of +6 to +29 percentile points across findings. The average improvement index for cognition is +1 percentile point for the one study, with a range of –5 to +5 percentile points across findings.

The average improvement index for math is +18 percentile points for the one study, with a range of +14 to +23 percentile points across findings.

Findings for comparisons between *Literacy Express* workshop and *Literacy Express* mentoring

The comparison of *Literacy Express* workshop and *Literacy Express* mentoring was included in the Lonigan et al. (2005) study, but does not contribute to the overall rating of effectiveness because it involves two different implementations of *Literacy Express,* which does not allow the effects of *Literacy Express* to be determined. However, the WWC believes that the findings from this comparison provide useful information to practitioners who may be interested in comparing the effects of different implementations of the same program. The WWC reports the findings from this comparison here and in Appendices A5.1-A5.4. Lonigan et al. (2005) analyzed group differences between the mentoring and workshop groups for the same outcome measures listed above in the domains of oral language, print knowledge, phonological processing, and cognition. A statistically significant effect in favor of the mentoring group was found in the print knowledge domain, but outcomes in the other domains did not differ significantly between the workshop and mentoring groups.

The WWC found *Literacy Express* to have positive effects on print knowledge and phonological processing, potentially positive effects on oral language and math, and no discernible effects on cognition *(continued)*

Summary

The WWC reviewed two studies on *Literacy Express* and both studies met WWC evidence standards. Based on these studies, the WWC found positive effects for print knowledge and phonological processing, potentially positive effects for oral language and math, and no discernible effects for cognition.

Additional findings that were not considered for the rating of effectiveness indicate that implementing *Literacy Express* using standard workshops plus mentoring may be beneficial for certain print knowledge outcomes when compared with using *Literacy Express* with only standard workshops. The evidence presented in this report may change as new research emerges.

References

Met WWC evidence standards

Lonigan, C. J. (2006, July). *Impact of preschool literacy curricula: Results of a randomized evaluation in a public prekindergarten program.* Paper presented at the 13th annual meeting of the Society for the Scientific Study of Reading, Vancouver, British Columbia, Canada.

Additional source:

Lonigan, C. J. (2005, December). *Impact of preschool literacy curricula: Results of a randomized evaluation in a public prekindergarten program.* Paper presented at the annual meeting of the National Association for the Education of Young Children, Washington, DC.

Lonigan, C. J., Farver, J. M., Clancy-Menchetti, J., & Phillips, B. M. (2005, June). *Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models.* Paper presented at the 12th annual meeting of the Society for the Scientific Study of Reading, Toronto, Ontario, Canada.

Additional source:

Lonigan, C. J., Farver, J. M., Clancy-Menchetti, J., & Phillips, B. M. (2005, April). *Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models.* Paper presented at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.

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

Appendix

Appendix A1.1 Study characteristics: Lonigan, 2006 (randomized controlled trial)

Characteristic	Description
Study citation	Lonigan, C. J. (2006, July). *Impact of preschool literacy curricula: Results of a randomized evaluation in a public prekindergarten program.* Paper presented at the 13th annual meeting of the Society for the Scientific Study of Reading, Vancouver, British Columbia, Canada. *Additional source:* Lonigan, C. J. (2005, December). *Impact of preschool literacy curricula: Results of a randomized evaluation in a public prekindergarten program.* Paper presented at the annual meeting of the National Association for the Education of Young Children, Washington, DC.
Participants	Fifteen of 17 preschools were rank ordered according to state letter grade and teacher experience and placed in groups of three that were then randomly assigned within triad to two intervention groups (*Literacy Express* or *DLM Express plus Open Court Pre-K*) or to a business-as-usual comparison group.[1] The two other preschools were randomly assigned to condition without going through this process due to factors outside of the researcher's control. This procedure resulted in five preschools (13 classrooms) in the *Literacy Express* group and six preschools (nine classrooms) in the business-as-usual comparison group.[2] In the intervention preschools with at least two classrooms, classrooms were randomly assigned to receive standard professional development or enhanced professional development; however, the study author found no effects for these conditions and did not include them in further analyses. Across groups, the study began with 406 preschool children with ages ranging from 33 to 64 months (mean age = 52.06 months). At posttest, 350 children remained in the sample (119 in the *Literacy Express* group and 91 in the business-as-usual comparison group) and ranged in age from 33 to 64 months (mean age = 52.05 months). The final sample included 60.5% African-American, 30.5% Caucasian children, and 9% children of other races and ethnicities. Forty-one percent of the children were female[3] and all children scored in the low-average range on a pretest of receptive vocabulary skills.
Setting	The study took place in 17 public preschool centers (35 classrooms) from two local school districts in northern Florida.
Intervention	Classrooms in the *Literacy Express* intervention group used the *Literacy Express* curriculum, which was taught in 11 three- to five-week thematic units using individual and whole-group activities, as well as small-group activies such as dialogic reading, phonological awareness activities, and print knowledge activities.[4]
Comparison	The business-as-usual comparison group classrooms used *High/Scope,* the standard curriculum used in their regular preschool centers. *High/Scope* prescribes a daily routine of outdoor play, planning, work, cleanup, recall, large-group time, and small-group time.
Primary outcomes and measurement	The primary outcome domains assessed were children's oral language, print knowledge, phonological processing, and mathematical knowledge. Oral language was assessed with two standardized measures: the Peabody Picture Vocabulary Test-III (PPVT-III) and the Preschool Comprehensive Test of Phonological and Print Processes (P-CTOPPP) Definitional Vocabulary subtest. Print knowledge was assessed with subtests from three standardized measures: the P-CTOPPP Print Knowledge subtest; the Test of Early Reading Ability-3 (TERA-3) Alphabet, Meaning, and Print Conventions subtests; and the Woodcock-Johnson III (W-J III) Spelling subtest. Phonological processing was assessed with the Blending and Elision subtests from the P-CTOPPP. Mathematical knowledge was assessed with two standardized measures: the W-J III Applied Problems subtest and the Children's Math Assessment, Abbreviated Version (see Appendices A2.1–A2.5 for more detailed descriptions of outcome measures).[5]

(continued)

Appendix A1.1 Study characteristics: Lonigan, 2006 (randomized controlled trial) *(continued)*

Characteristic	Description
Teacher training	The research staff provided all materials and training for the *Literacy Express* group. Two weeks prior to the beginning of preschool classes, staff in the *Literacy Express* group participated in separate two-day professional development workshops. Throughout the school year, staff in this group also attended separate two-hour professional development meetings every other month to troubleshoot and discuss curriculum implementation. At mid-year, a one-day professional development session was conducted with each group to provide training on the curriculum activities that would be used in the second part of the school year. All classrooms in the "enhanced" professional development condition were visited by the project's teacher mentor every other week for the school year.[6]

1. The data included in the study are from the second year of the Preschool Curriculum Evaluation Research project. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Literacy Express* group with the business-as-usual comparison group. The WWC does not include the *DLM Express plus Open Court Pre-K* versus business-as-usual comparison in a separate WWC intervention report because the effects of *DLM Express* and *Open Court Pre-K* on children's outcomes cannot be disentangled. The WWC does not include the head-to-head comparison of *Literacy Express* and *DLM Express plus Open Court Pre-K,* but interested readers can examine that comparison using the data provided in the original article.
2. This same process resulted in six preschools (13 classrooms; 140 children) in the *DLM Express plus Open Court Pre-K* group. The data about the number of preschools for each group were provided by the study author upon WWC request.
3. In the article, the study author reported that there were 203 males and 141 females (N = 344). The study author provided corrected sample sizes by gender (207 males and 143 females) upon WWC request.
4. Classrooms in the other intervention group used a combination of *DLM Express* (a comprehensive child-centered program that focuses on the development of the whole child through carefully sequenced lessons focused on language acquisition and other early reading skills, math, general knowledge, and socio-emotional development) and *Open Court Pre-K* (a supplemental curriculum focusing on the development of phonological awareness, phonics, early decoding, and comprehension skills primarily through whole-group instruction and activities).
5. The study author administered the Grammatical Understanding subtest of the Test of Oral Language Development (TOLD-GU), but this measure was used as a covariate, not a posttest outcome measure. The study author also trained research staff to observe the classrooms about three times a year to determine implementation fidelity and administer two general measures of classroom language and literacy, but these measures are not discussed further in this WWC intervention report. For further details about the outcomes included in the ECE review please see the Early Childhood Education Protocol.
6. The school districts provided all materials and training for the *High/Scope* comparison group. At the beginning of the preschool year, classroom staff were visited by *High/Scope* personnel and participated in a week-long "*High/Scope* Institute." Supplemental training was provided throughout the year by *High/Scope* trainers and school district staff. Periodic classroom visits were made by *High/Scope* trainers for additional training and assistance.

(continued)

Appendix A1.2 Study characteristics: Lonigan, Farver, Clancy-Menchetti, and Phillips, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Lonigan, C. J., Farver, J. M., Clancy-Menchetti, J., & Phillips, B. M. (2005, June). *Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models.* Paper presented at the 12th annual meeting of the Society for the Scientific Study of Reading, Toronto, Ontario, Canada. *Additional source:* Lonigan, C. J., Farver, J. M., Clancy-Menchetti, J., & Phillips, B. M. (2005, April). *Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models.* Paper presented at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.
Participants	Forty-eight preschools were randomly assigned within state to a *Literacy Express* workshop group, a *Literacy Express* workshop plus mentoring group, or the business-as-usual comparison condition, which resulted in 15 preschools in each of the two *Literacy Express* groups and 18 preschools in the business-as-usual comparison group. The study began with 808 preschool children ranging in age from 36 to 69 months (mean age = 50.63 months). At posttest 722 children remained in the sample and ranged in age from 36 to 69 months (mean age = 50.71 months). The final sample included 55.7% African-American children, 35.2% Latino/Hispanic children, 7.9% Caucasian children, and 1.1% children of other races and ethnicities. Forty-nine percent of the children were female; 52% of the children in the California sites and 1% of the children in the Florida sites were Spanish-speaking English language learners. All children were considered at-risk for academic difficulties as determined by pretest scores on a measure of cognitive performance.
Setting	The study took place in 18 preschools in Tallahassee, Florida, and 30 preschools in Los Angeles, California. The majority of the preschools (77%) were Head Start programs.
Intervention	Preschools in the intervention group participated in a *Literacy Express* plus professional development via workshops group ("workshop group") or a *Literacy Express* plus professional development via workshops and mentoring group ("mentoring group"). *Literacy Express* was taught in 11 three- to five-week thematic units using individual and whole-group activities, as well as small-group activities such as dialogic reading, phonological awareness activities, and print knowledge activities. The workshop group participated in two-day workshops at the beginning of the school year and three half-day workshops during the school year. The mentoring group participated in the same workshops and received regular classroom visits by a trained project mentor.[1]
Comparison	Preschools in the business-as-usual comparison group participated in the preschool's standard curriculum, which in most cases was *High/Scope* or *Creative Curriculum*.
Primary outcomes and measurement[2]	The primary outcome domains assessed were children's oral language, print knowledge, phonological processing, and cognition, all of which were assessed with standardized measures. Oral language was assessed with the Expressive Communication subscale from the Preschool Language Scales-IV (PLS-IV). Print knowledge was assessed with the Print Knowledge subtest from the Preschool Comprehensive Test of Phonological and Print Processes (P-CTOPPP). Phonological processing was assessed with the Blending and Elision subtests from the P-CTOPPP. Cognition was assessed with three subtests from the P-CTOPPP: Non-Word Repetition, Word Span, and Rapid Object Naming (see Appendices A2.1–A2.5 for more detailed descriptions of outcome measures).
Teacher training	The research staff provided all materials and training for the *Literacy Express* intervention groups. Classrooms teachers and aides attended a two-day curriculum-specific professional development workshop at the start of the school year as well as three half-day curriculum-specific professional development workshops throughout the school year. In all workshops, staff participated in both teacher-directed and hands-on components. Classroom staff in the mentoring group received regular classroom visits throughout the school year from trained project teacher-mentors in addition to the professional development activities.

1. The study authors combined the two intervention groups to examine the effectiveness of *Literacy Express,* and the WWC uses the combined data to determine the rating of effectiveness for this WWC intervention report. The study authors also provided the data for the comparisons between the individual intervention groups and the business-as-usual comparison group and the WWC includes the results from these analyses in Appendices A4.1–A4.4. The WWC also includes the comparison between the mentoring and workshop groups in Appendices A5.1–A5.4.
2. The study authors also trained research staff to observe the classrooms for three hours twice a year to determine implementation fidelity and administer two general measures of classroom language and literacy, but these measures are not discussed further in this WWC intervention report. For further details about the outcomes included in the ECE review, please see the Early Childhood Education Protocol.

Appendix A2.1 Outcome measures in the oral language domain

Outcome measure	Description
Preschool Comprehensive Test of Phonological and Print Processes (P-CTOPPP) Definitional Vocabulary subtest	A subtest from a standardized measure that assesses children's expressive vocabulary by requiring children to name the items shown in pictures and provide simple definitions for the expressive vocabulary items (as cited in Lonigan, 2006).
Peabody Picture Vocabulary Test-III (PPVT-III)	A standardized measure of children's single-word receptive vocabulary that requires children to identify pictures that correspond to spoken words (as cited in Lonigan, 2006).
Preschool Language Scales-IV (PLS-IV) Expressive Communication subscale	A subtest from a standardized measure that assesses children's expressive communication skills in multiple areas of language development (vocal development, social communication, semantics, structure, and integrative thinking) (as cited in Lonigan et al., 2005).

Appendix A2.2 Outcome measures in the print knowledge domain

Outcome measure	Description
Test of Early Reading Ability-3 (TERA-3) Alphabet subtest	A subtest from a standardized measure that assesses children's knowledge of the alphabet and its uses (as cited in Lonigan, 2006).
TERA-3 Meaning subtest	A subtest from a standardized measure that assesses children's construction of meaning from print (as cited in Lonigan, 2006).
TERA-3 Print Conventions subtest	A subtest from a standardized measure that assesses children's knowledge of the conventions of print (as cited in Lonigan, 2006).
Woodcock-Johnson III (W-J III) Spelling subtest	A subtest from a standardized measure that assesses children's ability to write letter forms, write letters, and complete simple spelling tasks (as cited in Lonigan, 2006).
P-CTOPPP Print Knowledge subtest	A subtest from a standardized measure that assesses children's early print concepts, alphabet recognition, letter-name knowledge, and letter-sound knowledge (as cited in Lonigan et al., 2005).

Appendix A2.3 Outcome measures in the phonological processing domain

Outcome measure	Description
P-CTOPPP Blending subtest	A subtest from a standardized measure that requires children to combine word sounds to make a new word and uses both recognition and expressive formats (as cited in Lonigan, 2006; Lonigan et al., 2005).
P-CTOPPP Elision subtest	A subtest from a standardized measure that requires children to take away a sound from a word to create a new word and uses both recognition and expressive formats (as cited in Lonigan, 2006; Lonigan et al., 2005).

Appendix A2.4 Outcome measures in the cognition domain

Outcome measure	Description
P-CTOPPP Non-Word Repetition subtest	A subtest from a standardized measure that assesses children's auditory short-term memory by having children repeat non-words built from English phonology that grow increasingly longer throughout the assessment (as cited in Lonigan et al., 2005).
P-CTOPPP Word Span subtest	A subtest from a standardized measure that assesses children's auditory short-term memory by having children repeat one- to seven-word lists of common words (as cited in Lonigan et al., 2005).
P-CTOPPP Rapid Object Naming subtest	A subtest from a standardized measure that assesses children's lexical access by measuring the speed with which children can name pictures of five common objects that are arranged randomly within rows (as cited in Lonigan et al., 2005). To make effect size estimates consistent across measures, the WWC reversed the direction of the effect so that a higher score reflected a better outcome.

Appendix A2.5 Outcome measures in the math domain

Outcome measure	Description
W-J III Applied Problems subtest	A subtest from a standardized measure that assesses children's math skills by asking children to count small sets and to solve simple addition and subtraction questions using pictures (as cited in Lonigan, 2006).
Children's Math Assessment, Abbreviated Version	A standardized measure that assesses children's informal math knowledge in numerous skill areas, including knowledge of number, arithmetic, space and geometry, patterns, and nonstandard measurement (as cited in Lonigan, 2006).

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

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
P-CTOPPP Definitional Vocabulary subtest	3–4 year olds	11[9]/210	58.47 (15.79)	51.25 (15.86)	7.22	0.45	ns	+18
PPVT-III	3–4 year olds	11[9]/210	90.38 (15.45)	83.67 (17.91)	6.71	0.40	ns	+16
Average[10] for oral language (Lonigan, 2006)						0.43	ns	+17
Lonigan et al., 2005 (randomized controlled trial)[11]								
PLS-IV Expressive Communication subscale	3–5 year olds	48/722	53.35 (8.78)	50.66 (9.71)	2.69	0.30	Statistically significant	+12
Average[10] for oral language (Lonigan et al., 2005)						0.30	Statistically significant	+12
Domain average[10] for oral language across all studies						0.36	na	+14

ns = not statistically significant
na = not applicable
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes
PPVT-III = Peabody Picture Vocabulary Test-III
PLS-IV = Preschool Language Scales-IV

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.1 and A5.1.
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. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. 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 size.
11. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed.

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

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
P-CTOPPP Print Knowledge subtest	3–4 year olds	11[9]/210	23.54 (10.42)	19.37 (9.60)	4.17	0.41	ns	+16
TERA-3 Alphabet subtest	3–4 year olds	11[9]/210	9.29 (3.03)	7.63 (2.71)	1.66	0.57	Statistically significant	+22
TERA-3 Meaning subtest	3–4 year olds	11[9]/210	9.93 (2.18)	8.07 (2.33)	1.86	0.83	Statistically significant	+30
TERA-3 Print Conventions subtest	3–4 year olds	11[9]/210	8.22 (2.54)	7.31 (2.80)	0.91	0.34	ns	+13
W-J III Spelling subtest	3–4 year olds	11[9]/210	57.53 (11.74)	51.83 (10.95)	5.70	0.50	Statistically significant	+19
Average[10] for print knowledge (Lonigan, 2006)						0.53	Statistically significant	+20
Lonigan et al., 2005 (randomized controlled trial)[11]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	48/722	17.55 (9.30)	14.70 (7.83)	2.85	0.32	Statistically significant	+13
Average[10] for print knowledge (Lonigan et al., 2005)						0.32	Statistically significant	+13
Domain average[10] for print knowledge across all studies						0.43	na	+16

ns = not statistically significant
na = not applicable
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes
TERA-3 = Test of Early Reading Ability-3
W-J III = Woodcock-Johnson III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.2 and A5.2.
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. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.

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

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 the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), a correction for multiple comparisons was needed, so the significance levels may differ from those found by the study author.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. 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 size.
11. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed.

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

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
P-CTOPPP Blending subtest	3–4 year olds	11[9]/210	16.82 (4.31)	13.29 (4.58)	3.53	0.79	Statistically significant	+29
P-CTOPPP Elision subtest	3–4 year olds	11[9]/210	10.58 (5.78)	8.15 (4.32)	2.43	0.47	Statistically significant	+18
Average[10] for phonological processing (Lonigan, 2006)						0.63	Statistically significant	+24
Lonigan et al., 2005 (randomized controlled trial)[11]								
P-CTOPPP Blending subtest	3–5 year olds	48/722	14.15 (4.47)	13.47 (4.47)	0.68	0.15	ns	+6
P-CTOPPP Elision subtest	3–5 year olds	48/722	8.86 (3.93)	7.43 (3.54)	1.43	0.38	Statistically significant	+15
Average[10] for phonological processing (Lonigan et al., 2005)						0.26	ns	+10
Domain average[10] for phonological processing across all studies						0.45	na	+17

ns = not statistically significant
na = not applicable
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.3 and A5.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), a correction for multiple comparisons was needed, so the significance levels may differ from those found by the study author.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. 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 size.
11. In the case of Lonigan et al. (2005), a correction for multiple comparisons was needed, so the significance levels may differ from those found by the study authors.

Appendix A3.4 Summary of study findings included in the rating for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	48/722	9.07 (4.16)	9.55 (4.36)	–0.48	–0.11	ns	–5
P-CTOPPP Word Span subtest	3–5 year olds	48/722	8.86 (2.56)	8.54 (2.54)	0.32	0.13	ns	+5
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	48/722	48.21 (17.12)	49.57 (16.93)	1.36[8]	0.08	ns	+3
Domain average[9] for cognition						0.03	ns	+1

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.4 and A5.4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings were not statistically significant.
8. For this outcome, the mean difference was calculated so that a positive effect was found when intervention group children took less time to complete the task than comparison group children (comparison group mean minus the intervention group mean).
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 A3.5 Summary of study findings included in the rating for the math domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
W-J III Applied Problems subtest	3–4 year olds	11[9]/210	53.65 (12.80)	49.43 (10.68)	4.22	0.35	ns	+14
CMA Abbreviated Version	3–4 year olds	11[9]/210	31.81 (8.10)	26.59 (9.34)	5.22	0.60	Statistically significant	+23
Domain average[10] for math						0.48	ns	+18

ns = not statistically significant
W-J III = Woodcock-Johnson III
CMA = Comprehensive Math Assessment

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
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. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), a correction multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of subgroup findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	33/442	52.63 (9.68)	50.66 (9.71)	1.97	0.20	ns	+8
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	33/516	54.04 (7.97)	50.66 (9.71)	3.38	0.38	Statistically significant	+15
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	30/198[8]	55.12 (6.40)	51.02 (8.01)	4.10	0.59	Statistically significant[9]	+22
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	30/231[8]	46.62 (10.11)	44.04 (10.76)	2.58	0.25	ns[9]	+10

ns = not statistically significant
PLS-IV = Preschool Language Scales-IV

1. This appendix presents subgroup findings for measures that fall in the oral language domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
9. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A4.2 Summary of subgroup findings for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	33/442	16.19 (9.21)	14.72 (7.83)	1.47	0.17	ns	+7
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	33/516	18.77 (9.16)	14.72 (7.83)	4.05	0.47	Statistically significant	+18
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	30/198[8]	17.40 (9.51)	15.32 (8.30)	2.08	0.23	ns[9]	+9
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	30/231[8]	15.32 (8.73)	11.29 (5.21)	4.03	0.52	Statistically significant[9]	+20

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents subgroup findings for measures that fall in the print knowledge domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
9. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A4.3 Summary of subgroup findings for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
P-CTOPPP Blending subtest	3–5 year olds	33/442	14.23 (4.34)	13.47 (4.47)	0.76	0.17	ns	+7
P-CTOPPP Elision subtest	3–5 year olds	33/442	8.91 (3.79)	7.43 (3.54)	1.48	0.40	Statistically significant	+16
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
P-CTOPPP Blending subtest	3–5 year olds	33/516	14.14 (4.54)	13.47 (4.47)	0.67	0.15	ns	+6
P-CTOPPP Elision subtest	3–5 year olds	33/516	8.86 (4.03)	7.43 (3.54)	1.43	0.37	Statistically significant	+15
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
P-CTOPPP Blending subtest	3–5 year olds	30/198[8]	14.70 (4.49)	14.10 (4.12)	0.60	0.14	ns[9]	+5
P-CTOPPP Elision subtest	3–5 year olds	30/198[8]	9.07 (3.43)	7.75 (3.32)	1.32	0.39	Statistically significant[9]	+15
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
P-CTOPPP Blending subtest	3–5 year olds	30/231[8]	12.94 (4.83)	12.18 (4.38)	0.76	0.16	ns[9]	+6
P-CTOPPP Elision subtest	3–5 year olds	30/231[8]	6.94 (3.40)	6.11 (2.09)	0.83	0.27	Statistically significant[9]	+11

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents subgroup findings for measures that fall in the phonological processing domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.

(continued)

Appendix A4.3 Summary of subgroup findings for the phonological processing domain *(continued)*

7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
9. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A4.4 Summary of subgroup findings for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	33/442	9.29 (3.96)	9.55 (4.36)	–0.26	–0.06	ns	–2
P-CTOPPP Word Span subtest	3–5 year olds	33/442	8.84 (2.42)	8.53 (2.54)	0.31	0.12	ns	+5
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	33/442	48.51 (20.36)	49.57 (16.93)	1.06[8]	0.06	ns	+2
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	33/516	8.89 (4.29)	9.55 (4.36)	–0.66	–0.15	ns	–6
P-CTOPPP Word Span subtest	3–5 year olds	33/516	8.88 (2.66)	8.53 (2.54)	0.35	0.13	ns	+5
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	33/516	47.87 (14.34)	49.57 (16.93)	1.70[8]	0.11	ns	+4
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	30/198[9]	10.62 (4.16)	9.55 (3.98)	1.07	0.26	ns[10]	+10
P-CTOPPP Word Span subtest	3–5 year olds	30/198[9]	9.30 (2.57)	8.57 (2.41)	0.73	0.29	ns[10]	+11
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	30/198[9]	44.89 (14.90)	46.24 (15.11)	1.35[8]	0.09	ns[10]	+4
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	30/231[9]	8.52 (3.89)	8.38 (3.59)	0.14	0.04	ns[10]	+1
P-CTOPPP Word Span subtest	3–5 year olds	30/231[9]	7.47 (2.29)	7.37 (2.32)	0.10	0.04	ns[10]	+2

(continued)

Appendix A4.4 Summary of subgroup findings for the cognition domain *(continued)*

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	30/231[9]	50.03 (18.92)	47.37 (11.52)	–2.66[8]	–0.16	ns[10]	–6

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents subgroup findings for measures that fall in the cognition domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. For this outcome, the mean difference was calculated so that a positive effect was found when intervention group children took less time to complete the task than comparison group children (comparison group mean minus the intervention group mean).
9. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
10. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A5.1 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	30/486	54.04 (7.97)	52.63 (9.68)	1.41	0.16	ns[8]	+6
Domain average[9] for oral language						0.16	ns	+6

ns = not statistically significant
PLS-IV = Preschool Language Scales-IV

1. This appendix presents a summary of study findings for measures that fall in the oral language domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors, or where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.2 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	30/486	18.77 (9.16)	16.19 (9.21)	2.58	0.28	Statistically significant[8]	+11
Domain average[9] for print knowledge						0.28	Statistically significant	+11

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents a summary of study findings for measures that fall in the print knowledge domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors, or where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.3 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	*Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	Effect size[4]	Statistical significance[5] (at $\alpha = 0.05$)	Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Blending subtest	3–5 year olds	30/486	14.14 (4.54)	14.23 (4.34)	–0.09	–0.02	ns[8]	–1
P-CTOPPP Elision subtest	3–5 year olds	30/486	8.86 (4.03)	8.91 (3.79)	–0.05	–0.01	ns[8]	–1
Domain average[9] for print knowledge						–0.02	ns	–1

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents a summary of study findings for measures that fall in the phonological processing domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors, or where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.4 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Non-word Repetition subtest	3–5 year olds	30/486	8.89 (4.29)	9.29 (3.96)	–0.40	–0.10	ns[8]	–4
P-CTOPPP Word Span subtest	3–5 year olds	30/486	8.88 (2.66)	8.84 (2.42)	0.04	0.02	ns[8]	+1
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	30/486	47.87 (14.34)	48.51 (20.36)	0.64[9]	0.04	ns[8]	+1
Domain average[9] for cognition						0.06	ns	+2

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents a summary of study findings for measures that fall in the cognition domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. For this outcome, the mean difference was calculated so that a positive effect was found when mentoring group children took less time to complete the task than workshop group children (workshop group mean minus the mentoring group mean).
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A6.1 *Literacy Express* 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 language, the WWC rated *Literacy Express* as having potentially positive effects. It did not meet the criteria for positive effects because only one study had statistically significant positive findings. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received

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

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

  Met. One of the two studies showed statistically significant positive effects.

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.

  Met. The studies did not show statistically significant negative, substantively important negative, or indeterminate effects.

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. Two studies examined effects on oral language and both studies had a strong design, but only one study showed statistically significant positive effects.

AND

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

  Met. The studies did not show statistically significant or substantively important negative effects.

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

Appendix A6.2 *Literacy Express* 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 *Literacy Express* as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received
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.
Met. Two studies examined effects on print knowledge; both studies had a strong design and showed statistically significant positive effects.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. The studies did not show statistically significant or substantively important negative effects.

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

Appendix A6.3 *Literacy Express* 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 *Literacy Express* as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received
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.
Met. Two studies examined effects on phonological processing; both studies had a strong design and showed statistically significant positive effects.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. The studies did not show statistically significant or substantively important negative effects.

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

Appendix A6.4 *Literacy Express* rating for the cognition 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 cognition, the WWC rated *Literacy Express* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

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. One study examined effects on cognition and it did not show statistically significant or substantively important effects, either positive or negative; both studies showed indeterminate effects.

Other ratings considered

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

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

  Not met. Only one study examined effects on cognition.

AND

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

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

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 study did not show statistically significant or substantively important positive effects.

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 study did not show statistically significant or substantively important effects, either positive or negative; the study showed indeterminate effects.

(continued)

Appendix A6.4 *Literacy Express* rating for the cognition domain *(continued)*

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.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

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 study did not show statistically significant or substantively important effects, either positive or negative.

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 study did not show statistically significant or substantively important negative effects.

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 study did not show statistically significant or substantively important positive effects.

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. Only one study examined effects on cognition.

AND

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

  Met. The study did not show statistically significant or substantively important positive effects.

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

Appendix A6.5 *Literacy Express* 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 *Literacy Express* as having potentially positive effects. It did not meet the criteria for positive effects because only one study had statistically significant positive findings. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received

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

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

  Met. The study showed statistically significant positive effects.

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.

  Met. The study did not show statistically significant negative, substantively important negative, or indeterminate effects.

Other ratings considered

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

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

  Not met. Only one study examined effects on math.

AND

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

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

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

Appendix A7 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Children	
Oral language	2	59	932	Moderate to large
Print knowledge	2	59	932	Moderate to large
Phonological processing	2	59	932	Moderate to large
Early reading/writing	0	0	0	na
Cognition	1	48	722	Small
Math	1	11	210	Small

na = not applicable/not studied

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

Appendix

Appendix A1.1 Study characteristics: Lonigan, 2006 (randomized controlled trial)

Characteristic	Description
Study citation	Lonigan, C. J. (2006, July). *Impact of preschool literacy curricula: Results of a randomized evaluation in a public prekindergarten program.* Paper presented at the 13th annual meeting of the Society for the Scientific Study of Reading, Vancouver, British Columbia, Canada. *Additional source:* Lonigan, C. J. (2005, December). *Impact of preschool literacy curricula: Results of a randomized evaluation in a public prekindergarten program.* Paper presented at the annual meeting of the National Association for the Education of Young Children, Washington, DC.
Participants	Fifteen of 17 preschools were rank ordered according to state letter grade and teacher experience and placed in groups of three that were then randomly assigned within triad to two intervention groups (*Literacy Express* or *DLM Express plus Open Court Pre-K*) or to a business-as-usual comparison group.[1] The two other preschools were randomly assigned to condition without going through this process due to factors outside of the researcher's control. This procedure resulted in five preschools (13 classrooms) in the *Literacy Express* group and six preschools (nine classrooms) in the business-as-usual comparison group.[2] In the intervention preschools with at least two classrooms, classrooms were randomly assigned to receive standard professional development or enhanced professional development; however, the study author found no effects for these conditions and did not include them in further analyses. Across groups, the study began with 406 preschool children with ages ranging from 33 to 64 months (mean age = 52.06 months). At posttest, 350 children remained in the sample (119 in the *Literacy Express* group and 91 in the business-as-usual comparison group) and ranged in age from 33 to 64 months (mean age = 52.05 months). The final sample included 60.5% African-American, 30.5% Caucasian children, and 9% children of other races and ethnicities. Forty-one percent of the children were female[3] and all children scored in the low-average range on a pretest of receptive vocabulary skills.
Setting	The study took place in 17 public preschool centers (35 classrooms) from two local school districts in northern Florida.
Intervention	Classrooms in the *Literacy Express* intervention group used the *Literacy Express* curriculum, which was taught in 11 three- to five-week thematic units using individual and whole-group activities, as well as small-group activies such as dialogic reading, phonological awareness activities, and print knowledge activities.[4]
Comparison	The business-as-usual comparison group classrooms used *High/Scope,* the standard curriculum used in their regular preschool centers. *High/Scope* prescribes a daily routine of outdoor play, planning, work, cleanup, recall, large-group time, and small-group time.
Primary outcomes and measurement	The primary outcome domains assessed were children's oral language, print knowledge, phonological processing, and mathematical knowledge. Oral language was assessed with two standardized measures: the Peabody Picture Vocabulary Test-III (PPVT-III) and the Preschool Comprehensive Test of Phonological and Print Processes (P-CTOPPP) Definitional Vocabulary subtest. Print knowledge was assessed with subtests from three standardized measures: the P-CTOPPP Print Knowledge subtest; the Test of Early Reading Ability-3 (TERA-3) Alphabet, Meaning, and Print Conventions subtests; and the Woodcock-Johnson III (W-J III) Spelling subtest. Phonological processing was assessed with the Blending and Elision subtests from the P-CTOPPP. Mathematical knowledge was assessed with two standardized measures: the W-J III Applied Problems subtest and the Children's Math Assessment, Abbreviated Version (see Appendices A2.1–A2.5 for more detailed descriptions of outcome measures).[5]

(continued)

Appendix A1.1 Study characteristics: Lonigan, 2006 (randomized controlled trial) *(continued)*

Characteristic	Description
Teacher training	The research staff provided all materials and training for the *Literacy Express* group. Two weeks prior to the beginning of preschool classes, staff in the *Literacy Express* group participated in separate two-day professional development workshops. Throughout the school year, staff in this group also attended separate two-hour professional development meetings every other month to troubleshoot and discuss curriculum implementation. At mid-year, a one-day professional development session was conducted with each group to provide training on the curriculum activities that would be used in the second part of the school year. All classrooms in the "enhanced" professional development condition were visited by the project's teacher mentor every other week for the school year.[6]

1. The data included in the study are from the second year of the Preschool Curriculum Evaluation Research project. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Literacy Express* group with the business-as-usual comparison group. The WWC does not include the *DLM Express plus Open Court Pre-K* versus business-as-usual comparison in a separate WWC intervention report because the effects of *DLM Express* and *Open Court Pre-K* on children's outcomes cannot be disentangled. The WWC does not include the head-to-head comparison of *Literacy Express* and *DLM Express plus Open Court Pre-K,* but interested readers can examine that comparison using the data provided in the original article.
2. This same process resulted in six preschools (13 classrooms; 140 children) in the *DLM Express plus Open Court Pre-K* group. The data about the number of preschools for each group were provided by the study author upon WWC request.
3. In the article, the study author reported that there were 203 males and 141 females (N = 344). The study author provided corrected sample sizes by gender (207 males and 143 females) upon WWC request.
4. Classrooms in the other intervention group used a combination of *DLM Express* (a comprehensive child-centered program that focuses on the development of the whole child through carefully sequenced lessons focused on language acquisition and other early reading skills, math, general knowledge, and socio-emotional development) and *Open Court Pre-K* (a supplemental curriculum focusing on the development of phonological awareness, phonics, early decoding, and comprehension skills primarily through whole-group instruction and activities).
5. The study author administered the Grammatical Understanding subtest of the Test of Oral Language Development (TOLD-GU), but this measure was used as a covariate, not a posttest outcome measure. The study author also trained research staff to observe the classrooms about three times a year to determine implementation fidelity and administer two general measures of classroom language and literacy, but these measures are not discussed further in this WWC intervention report. For further details about the outcomes included in the ECE review please see the Early Childhood Education Protocol.
6. The school districts provided all materials and training for the *High/Scope* comparison group. At the beginning of the preschool year, classroom staff were visited by *High/Scope* personnel and participated in a week-long "*High/Scope* Institute." Supplemental training was provided throughout the year by *High/Scope* trainers and school district staff. Periodic classroom visits were made by *High/Scope* trainers for additional training and assistance.

(continued)

Appendix A1.2 Study characteristics: Lonigan, Farver, Clancy-Menchetti, and Phillips, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Lonigan, C. J., Farver, J. M., Clancy-Menchetti, J., & Phillips, B. M. (2005, June). *Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models.* Paper presented at the 12th annual meeting of the Society for the Scientific Study of Reading, Toronto, Ontario, Canada. *Additional source:* Lonigan, C. J., Farver, J. M., Clancy-Menchetti, J., & Phillips, B. M. (2005, April). *Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models.* Paper presented at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.
Participants	Forty-eight preschools were randomly assigned within state to a *Literacy Express* workshop group, a *Literacy Express* workshop plus mentoring group, or the business-as-usual comparison condition, which resulted in 15 preschools in each of the two *Literacy Express* groups and 18 preschools in the business-as-usual comparison group. The study began with 808 preschool children ranging in age from 36 to 69 months (mean age = 50.63 months). At posttest 722 children remained in the sample and ranged in age from 36 to 69 months (mean age = 50.71 months). The final sample included 55.7% African-American children, 35.2% Latino/Hispanic children, 7.9% Caucasian children, and 1.1% children of other races and ethnicities. Forty-nine percent of the children were female; 52% of the children in the California sites and 1% of the children in the Florida sites were Spanish-speaking English language learners. All children were considered at-risk for academic difficulties as determined by pretest scores on a measure of cognitive performance.
Setting	The study took place in 18 preschools in Tallahassee, Florida, and 30 preschools in Los Angeles, California. The majority of the preschools (77%) were Head Start programs.
Intervention	Preschools in the intervention group participated in a *Literacy Express* plus professional development via workshops group ("workshop group") or a *Literacy Express* plus professional development via workshops and mentoring group ("mentoring group"). *Literacy Express* was taught in 11 three- to five-week thematic units using individual and whole-group activities, as well as small-group activities such as dialogic reading, phonological awareness activities, and print knowledge activities. The workshop group participated in two-day workshops at the beginning of the school year and three half-day workshops during the school year. The mentoring group participated in the same workshops and received regular classroom visits by a trained project mentor.[1]
Comparison	Preschools in the business-as-usual comparison group participated in the preschool's standard curriculum, which in most cases was *High/Scope* or *Creative Curriculum*.
Primary outcomes and measurement[2]	The primary outcome domains assessed were children's oral language, print knowledge, phonological processing, and cognition, all of which were assessed with standardized measures. Oral language was assessed with the Expressive Communication subscale from the Preschool Language Scales-IV (PLS-IV). Print knowledge was assessed with the Print Knowledge subtest from the Preschool Comprehensive Test of Phonological and Print Processes (P-CTOPPP). Phonological processing was assessed with the Blending and Elision subtests from the P-CTOPPP. Cognition was assessed with three subtests from the P-CTOPPP: Non-Word Repetition, Word Span, and Rapid Object Naming (see Appendices A2.1–A2.5 for more detailed descriptions of outcome measures).
Teacher training	The research staff provided all materials and training for the *Literacy Express* intervention groups. Classrooms teachers and aides attended a two-day curriculum-specific professional development workshop at the start of the school year as well as three half-day curriculum-specific professional development workshops throughout the school year. In all workshops, staff participated in both teacher-directed and hands-on components. Classroom staff in the mentoring group received regular classroom visits throughout the school year from trained project teacher-mentors in addition to the professional development activities.

1. The study authors combined the two intervention groups to examine the effectiveness of *Literacy Express,* and the WWC uses the combined data to determine the rating of effectiveness for this WWC intervention report. The study authors also provided the data for the comparisons between the individual intervention groups and the business-as-usual comparison group and the WWC includes the results from these analyses in Appendices A4.1–A4.4. The WWC also includes the comparison between the mentoring and workshop groups in Appendices A5.1–A5.4.
2. The study authors also trained research staff to observe the classrooms for three hours twice a year to determine implementation fidelity and administer two general measures of classroom language and literacy, but these measures are not discussed further in this WWC intervention report. For further details about the outcomes included in the ECE review, please see the Early Childhood Education Protocol.

Appendix A2.1 Outcome measures in the oral language domain

Outcome measure	Description
Preschool Comprehensive Test of Phonological and Print Processes (P-CTOPPP) Definitional Vocabulary subtest	A subtest from a standardized measure that assesses children's expressive vocabulary by requiring children to name the items shown in pictures and provide simple definitions for the expressive vocabulary items (as cited in Lonigan, 2006).
Peabody Picture Vocabulary Test-III (PPVT-III)	A standardized measure of children's single-word receptive vocabulary that requires children to identify pictures that correspond to spoken words (as cited in Lonigan, 2006).
Preschool Language Scales-IV (PLS-IV) Expressive Communication subscale	A subtest from a standardized measure that assesses children's expressive communication skills in multiple areas of language development (vocal development, social communication, semantics, structure, and integrative thinking) (as cited in Lonigan et al., 2005).

Appendix A2.2 Outcome measures in the print knowledge domain

Outcome measure	Description
Test of Early Reading Ability-3 (TERA-3) Alphabet subtest	A subtest from a standardized measure that assesses children's knowledge of the alphabet and its uses (as cited in Lonigan, 2006).
TERA-3 Meaning subtest	A subtest from a standardized measure that assesses children's construction of meaning from print (as cited in Lonigan, 2006).
TERA-3 Print Conventions subtest	A subtest from a standardized measure that assesses children's knowledge of the conventions of print (as cited in Lonigan, 2006).
Woodcock-Johnson III (W-J III) Spelling subtest	A subtest from a standardized measure that assesses children's ability to write letter forms, write letters, and complete simple spelling tasks (as cited in Lonigan, 2006).
P-CTOPPP Print Knowledge subtest	A subtest from a standardized measure that assesses children's early print concepts, alphabet recognition, letter-name knowledge, and letter-sound knowledge (as cited in Lonigan et al., 2005).

Appendix A2.3 Outcome measures in the phonological processing domain

Outcome measure	Description
P-CTOPPP Blending subtest	A subtest from a standardized measure that requires children to combine word sounds to make a new word and uses both recognition and expressive formats (as cited in Lonigan, 2006; Lonigan et al., 2005).
P-CTOPPP Elision subtest	A subtest from a standardized measure that requires children to take away a sound from a word to create a new word and uses both recognition and expressive formats (as cited in Lonigan, 2006; Lonigan et al., 2005).

Appendix A2.4 Outcome measures in the cognition domain

Outcome measure	Description
P-CTOPPP Non-Word Repetition subtest	A subtest from a standardized measure that assesses children's auditory short-term memory by having children repeat non-words built from English phonology that grow increasingly longer throughout the assessment (as cited in Lonigan et al., 2005).
P-CTOPPP Word Span subtest	A subtest from a standardized measure that assesses children's auditory short-term memory by having children repeat one- to seven-word lists of common words (as cited in Lonigan et al., 2005).
P-CTOPPP Rapid Object Naming subtest	A subtest from a standardized measure that assesses children's lexical access by measuring the speed with which children can name pictures of five common objects that are arranged randomly within rows (as cited in Lonigan et al., 2005). To make effect size estimates consistent across measures, the WWC reversed the direction of the effect so that a higher score reflected a better outcome.

Appendix A2.5 Outcome measures in the math domain

Outcome measure	Description
W-J III Applied Problems subtest	A subtest from a standardized measure that assesses children's math skills by asking children to count small sets and to solve simple addition and subtraction questions using pictures (as cited in Lonigan, 2006).
Children's Math Assessment, Abbreviated Version	A standardized measure that assesses children's informal math knowledge in numerous skill areas, including knowledge of number, arithmetic, space and geometry, patterns, and nonstandard measurement (as cited in Lonigan, 2006).

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

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
P-CTOPPP Definitional Vocabulary subtest	3–4 year olds	11[9]/210	58.47 (15.79)	51.25 (15.86)	7.22	0.45	ns	+18
PPVT-III	3–4 year olds	11[9]/210	90.38 (15.45)	83.67 (17.91)	6.71	0.40	ns	+16
Average[10] for oral language (Lonigan, 2006)						0.43	ns	+17
Lonigan et al., 2005 (randomized controlled trial)[11]								
PLS-IV Expressive Communication subscale	3–5 year olds	48/722	53.35 (8.78)	50.66 (9.71)	2.69	0.30	Statistically significant	+12
Average[10] for oral language (Lonigan et al., 2005)						0.30	Statistically significant	+12
Domain average[10] for oral language across all studies						0.36	na	+14

ns = not statistically significant
na = not applicable
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes
PPVT-III = Peabody Picture Vocabulary Test-III
PLS-IV = Preschool Language Scales-IV

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.1 and A5.1.
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. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. 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 size.
11. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed.

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

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
P-CTOPPP Print Knowledge subtest	3–4 year olds	11[9]/210	23.54 (10.42)	19.37 (9.60)	4.17	0.41	ns	+16
TERA-3 Alphabet subtest	3–4 year olds	11[9]/210	9.29 (3.03)	7.63 (2.71)	1.66	0.57	Statistically significant	+22
TERA-3 Meaning subtest	3–4 year olds	11[9]/210	9.93 (2.18)	8.07 (2.33)	1.86	0.83	Statistically significant	+30
TERA-3 Print Conventions subtest	3–4 year olds	11[9]/210	8.22 (2.54)	7.31 (2.80)	0.91	0.34	ns	+13
W-J III Spelling subtest	3–4 year olds	11[9]/210	57.53 (11.74)	51.83 (10.95)	5.70	0.50	Statistically significant	+19
Average[10] for print knowledge (Lonigan, 2006)						0.53	Statistically significant	+20
Lonigan et al., 2005 (randomized controlled trial)[11]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	48/722	17.55 (9.30)	14.70 (7.83)	2.85	0.32	Statistically significant	+13
Average[10] for print knowledge (Lonigan et al., 2005)						0.32	Statistically significant	+13
Domain average[10] for print knowledge across all studies						0.43	na	+16

ns = not statistically significant
na = not applicable
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes
TERA-3 = Test of Early Reading Ability-3
W-J III = Woodcock-Johnson III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.2 and A5.2.
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. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.

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

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 the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), a correction for multiple comparisons was needed, so the significance levels may differ from those found by the study author.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. 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 size.
11. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed.

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

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
P-CTOPPP Blending subtest	3–4 year olds	11[9]/210	16.82 (4.31)	13.29 (4.58)	3.53	0.79	Statistically significant	+29
P-CTOPPP Elision subtest	3–4 year olds	11[9]/210	10.58 (5.78)	8.15 (4.32)	2.43	0.47	Statistically significant	+18
Average[10] for phonological processing (Lonigan, 2006)						0.63	Statistically significant	+24
Lonigan et al., 2005 (randomized controlled trial)[11]								
P-CTOPPP Blending subtest	3–5 year olds	48/722	14.15 (4.47)	13.47 (4.47)	0.68	0.15	ns	+6
P-CTOPPP Elision subtest	3–5 year olds	48/722	8.86 (3.93)	7.43 (3.54)	1.43	0.38	Statistically significant	+15
Average[10] for phonological processing (Lonigan et al., 2005)						0.26	ns	+10
Domain average[10] for phonological processing across all studies						0.45	na	+17

ns = not statistically significant
na = not applicable
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.3 and A5.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), a correction for multiple comparisons was needed, so the significance levels may differ from those found by the study author.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. 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 size.
11. In the case of Lonigan et al. (2005), a correction for multiple comparisons was needed, so the significance levels may differ from those found by the study authors.

Appendix A3.4 Summary of study findings included in the rating for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	48/722	9.07 (4.16)	9.55 (4.36)	–0.48	–0.11	ns	–5
P-CTOPPP Word Span subtest	3–5 year olds	48/722	8.86 (2.56)	8.54 (2.54)	0.32	0.13	ns	+5
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	48/722	48.21 (17.12)	49.57 (16.93)	1.36[8]	0.08	ns	+3
Domain average[9] for cognition						0.03	ns	+1

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.4 and A5.4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings were not statistically significant.
8. For this outcome, the mean difference was calculated so that a positive effect was found when intervention group children took less time to complete the task than comparison group children (comparison group mean minus the intervention group mean).
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 A3.5 Summary of study findings included in the rating for the math domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Literacy Express* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Lonigan, 2006 (randomized controlled trial)[8]								
W-J III Applied Problems subtest	3–4 year olds	11[9]/210	53.65 (12.80)	49.43 (10.68)	4.22	0.35	ns	+14
CMA Abbreviated Version	3–4 year olds	11[9]/210	31.81 (8.10)	26.59 (9.34)	5.22	0.60	Statistically significant	+23
Domain average[10] for math						0.48	ns	+18

ns = not statistically significant
W-J III = Woodcock-Johnson III
CMA = Comprehensive Math Assessment

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
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. Lonigan (2006) reported that the pretest was administered approximately one and a half months after the intervention began at the start of the school year because of a new state law requiring clearance for the researchers. Because the fall assessment was not a true pretest (children had been exposed to 1.5 to 3 months of the curriculum in their respective condition), the study author conducted the analyses three ways and the WWC used the estimates from the no-covariate model. The intervention had a significant impact on several outcomes by the time the pretest was conducted, so the analyses controlling for pretest underestimate the true impact of the intervention.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan (2006), a correction multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
9. The number of preschools in each group at assignment was provided by the study author upon WWC request.
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of subgroup findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	33/442	52.63 (9.68)	50.66 (9.71)	1.97	0.20	ns	+8
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	33/516	54.04 (7.97)	50.66 (9.71)	3.38	0.38	Statistically significant	+15
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	30/198[8]	55.12 (6.40)	51.02 (8.01)	4.10	0.59	Statistically significant[9]	+22
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	30/231[8]	46.62 (10.11)	44.04 (10.76)	2.58	0.25	ns[9]	+10

ns = not statistically significant
PLS-IV = Preschool Language Scales-IV

1. This appendix presents subgroup findings for measures that fall in the oral language domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
9. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A4.2 Summary of subgroup findings for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	33/442	16.19 (9.21)	14.72 (7.83)	1.47	0.17	ns	+7
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	33/516	18.77 (9.16)	14.72 (7.83)	4.05	0.47	Statistically significant	+18
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	30/198[8]	17.40 (9.51)	15.32 (8.30)	2.08	0.23	ns[9]	+9
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	30/231[8]	15.32 (8.73)	11.29 (5.21)	4.03	0.52	Statistically significant[9]	+20

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents subgroup findings for measures that fall in the print knowledge domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
9. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A4.3 Summary of subgroup findings for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
P-CTOPPP Blending subtest	3–5 year olds	33/442	14.23 (4.34)	13.47 (4.47)	0.76	0.17	ns	+7
P-CTOPPP Elision subtest	3–5 year olds	33/442	8.91 (3.79)	7.43 (3.54)	1.48	0.40	Statistically significant	+16
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
P-CTOPPP Blending subtest	3–5 year olds	33/516	14.14 (4.54)	13.47 (4.47)	0.67	0.15	ns	+6
P-CTOPPP Elision subtest	3–5 year olds	33/516	8.86 (4.03)	7.43 (3.54)	1.43	0.37	Statistically significant	+15
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
P-CTOPPP Blending subtest	3–5 year olds	30/198[8]	14.70 (4.49)	14.10 (4.12)	0.60	0.14	ns[9]	+5
P-CTOPPP Elision subtest	3–5 year olds	30/198[8]	9.07 (3.43)	7.75 (3.32)	1.32	0.39	Statistically significant[9]	+15
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
P-CTOPPP Blending subtest	3–5 year olds	30/231[8]	12.94 (4.83)	12.18 (4.38)	0.76	0.16	ns[9]	+6
P-CTOPPP Elision subtest	3–5 year olds	30/231[8]	6.94 (3.40)	6.11 (2.09)	0.83	0.27	Statistically significant[9]	+11

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents subgroup findings for measures that fall in the phonological processing domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.

(continued)

Appendix A4.3 Summary of subgroup findings for the phonological processing domain *(continued)*

7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
9. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A4.4 Summary of subgroup findings for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial; workshop group)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	33/442	9.29 (3.96)	9.55 (4.36)	–0.26	–0.06	ns	–2
P-CTOPPP Word Span subtest	3–5 year olds	33/442	8.84 (2.42)	8.53 (2.54)	0.31	0.12	ns	+5
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	33/442	48.51 (20.36)	49.57 (16.93)	1.06[8]	0.06	ns	+2
Lonigan et al., 2005 (randomized controlled trial; mentoring group)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	33/516	8.89 (4.29)	9.55 (4.36)	–0.66	–0.15	ns	–6
P-CTOPPP Word Span subtest	3–5 year olds	33/516	8.88 (2.66)	8.53 (2.54)	0.35	0.13	ns	+5
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	33/516	47.87 (14.34)	49.57 (16.93)	1.70[8]	0.11	ns	+4
Lonigan et al., 2005 (randomized controlled trial; English speakers in the California sample)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	30/198[9]	10.62 (4.16)	9.55 (3.98)	1.07	0.26	ns[10]	+10
P-CTOPPP Word Span subtest	3–5 year olds	30/198[9]	9.30 (2.57)	8.57 (2.41)	0.73	0.29	ns[10]	+11
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	30/198[9]	44.89 (14.90)	46.24 (15.11)	1.35[8]	0.09	ns[10]	+4
Lonigan et al., 2005 (randomized controlled trial; Spanish speakers in the California sample)[7]								
P-CTOPPP Non-Word Repetition subtest	3–5 year olds	30/231[9]	8.52 (3.89)	8.38 (3.59)	0.14	0.04	ns[10]	+1
P-CTOPPP Word Span subtest	3–5 year olds	30/231[9]	7.47 (2.29)	7.37 (2.32)	0.10	0.04	ns[10]	+2

(continued)

Appendix A4.4 Summary of subgroup findings for the cognition domain *(continued)*

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Literacy Express* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	30/231[9]	50.03 (18.92)	47.37 (11.52)	–2.66[8]	–0.16	ns[10]	–6

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents subgroup findings for measures that fall in the cognition domain. Total group 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no correction for clustering was needed.
8. For this outcome, the mean difference was calculated so that a positive effect was found when intervention group children took less time to complete the task than comparison group children (comparison group mean minus the intervention group mean).
9. The number of preschools in each group at assignment and the number of children in each group at posttest was provided by the study authors upon WWC request.
10. The p-value for this contrast was provided by the study authors upon WWC request.

Appendix A5.1 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
PLS-IV Expressive Communication subscale	3–5 year olds	30/486	54.04 (7.97)	52.63 (9.68)	1.41	0.16	ns[8]	+6
Domain average[9] for oral language						0.16	ns	+6

ns = not statistically significant
PLS-IV = Preschool Language Scales-IV

1. This appendix presents a summary of study findings for measures that fall in the oral language domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors, or where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.2 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Print Knowledge subtest	3–5 year olds	30/486	18.77 (9.16)	16.19 (9.21)	2.58	0.28	Statistically significant[8]	+11
Domain average[9] for print knowledge						0.28	Statistically significant	+11

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents a summary of study findings for measures that fall in the print knowledge domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors, or where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.3 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	*Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	Effect size[4]	Statistical significance[5] (at $\alpha = 0.05$)	Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Blending subtest	3–5 year olds	30/486	14.14 (4.54)	14.23 (4.34)	–0.09	–0.02	ns[8]	–1
P-CTOPPP Elision subtest	3–5 year olds	30/486	8.86 (4.03)	8.91 (3.79)	–0.05	–0.01	ns[8]	–1
Domain average[9] for print knowledge						–0.02	ns	–1

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents a summary of study findings for measures that fall in the phonological processing domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors, or where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.4 Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* mentoring group	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Literacy Express* workshop group	WWC calculations: Mean difference[3] (*Literacy Express* mentoring – *Literacy Express* workshop)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Lonigan et al., 2005 (randomized controlled trial)[7]								
P-CTOPPP Non-word Repetition subtest	3–5 year olds	30/486	8.89 (4.29)	9.29 (3.96)	–0.40	–0.10	ns[8]	–4
P-CTOPPP Word Span subtest	3–5 year olds	30/486	8.88 (2.66)	8.84 (2.42)	0.04	0.02	ns[8]	+1
P-CTOPPP Rapid Object Naming subtest	3–5 year olds	30/486	47.87 (14.34)	48.51 (20.36)	0.64[9]	0.04	ns[8]	+1
Domain average[10] for cognition						0.06	ns	+2

ns = not statistically significant
P-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processes

1. This appendix presents a summary of study findings for measures that fall in the cognition domain for a comparison of two modes of implementation of *Literacy Express* that is not included in the overall effectiveness ratings.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the mentoring group; negative differences and effect sizes favor the workshop group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the mentoring condition and the percentile rank of the average student in the workshop condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the mentoring group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the study reported findings that were not statistically significant.
8. The p-value for this contrast was provided by the study authors upon WWC request.
9. For this outcome, the mean difference was calculated so that a positive effect was found when mentoring group children took less time to complete the task than workshop group children (workshop group mean minus the mentoring group mean).
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A6.1 *Literacy Express* 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 language, the WWC rated *Literacy Express* as having potentially positive effects. It did not meet the criteria for positive effects because only one study had statistically significant positive findings. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received

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

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

  Met. One of the two studies showed statistically significant positive effects.

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.

  Met. The studies did not show statistically significant negative, substantively important negative, or indeterminate effects.

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. Two studies examined effects on oral language and both studies had a strong design, but only one study showed statistically significant positive effects.

AND

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

  Met. The studies did not show statistically significant or substantively important negative effects.

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

Appendix A6.2 *Literacy Express* 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 *Literacy Express* as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received
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.
Met. Two studies examined effects on print knowledge; both studies had a strong design and showed statistically significant positive effects.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. The studies did not show statistically significant or substantively important negative effects.

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

Appendix A6.3 *Literacy Express* 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 *Literacy Express* as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received
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.
Met. Two studies examined effects on phonological processing; both studies had a strong design and showed statistically significant positive effects.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. The studies did not show statistically significant or substantively important negative effects.

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

Appendix A6.4 *Literacy Express* rating for the cognition 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 cognition, the WWC rated *Literacy Express* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

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. One study examined effects on cognition and it did not show statistically significant or substantively important effects, either positive or negative; both studies showed indeterminate effects.

Other ratings considered

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

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

  Not met. Only one study examined effects on cognition.

AND

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

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

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 study did not show statistically significant or substantively important positive effects.

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 study did not show statistically significant or substantively important effects, either positive or negative; the study showed indeterminate effects.

(continued)

Appendix A6.4 *Literacy Express* rating for the cognition domain *(continued)*

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.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

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 study did not show statistically significant or substantively important effects, either positive or negative.

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 study did not show statistically significant or substantively important negative effects.

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 study did not show statistically significant or substantively important positive effects.

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. Only one study examined effects on cognition.

AND

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

  Met. The study did not show statistically significant or substantively important positive effects.

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

Appendix A6.5 *Literacy Express* 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 *Literacy Express* as having potentially positive effects. It did not meet the criteria for positive effects because only one study had statistically significant positive findings. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Literacy Express* was assigned the highest applicable rating.

Rating received

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

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

  Met. The study showed statistically significant positive effects.

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.

  Met. The study did not show statistically significant negative, substantively important negative, or indeterminate effects.

Other ratings considered

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

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

  Not met. Only one study examined effects on math.

AND

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

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

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

Appendix A7 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Children	
Oral language	2	59	932	Moderate to large
Print knowledge	2	59	932	Moderate to large
Phonological processing	2	59	932	Moderate to large
Early reading/writing	0	0	0	na
Cognition	1	48	722	Small
Math	1	11	210	Small

na = not applicable/not studied

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