

# Appendix

## Appendix A1.1 Study characteristics: Farver, Lonigan, & Eppe, 2009

Characteristic	Description
<b>Study citation</b>	Farver, J. M., Lonigan, C. J., & Eppe, S. (2009). Effective early literacy skill development for young Spanish-speaking English language learners: An experimental study of two methods. <i>Child Development, 80</i> (3), 703–719.
<b>Participants</b>	Ninety-six Spanish-speaking English language learners in a Head Start program were randomly assigned, balancing for gender, to one of three conditions: (1) the <i>High/Scope</i> curriculum supplemented with small groups using <i>Literacy Express</i> in English only, (2) the <i>High/Scope</i> curriculum supplemented with small groups using <i>Literacy Express</i> beginning in Spanish and transitioning to English, and (3) the <i>High/Scope</i> curriculum only. Children were assigned to conditions within 10 classrooms. During the course of the year, two children moved, resulting in a sample of 94 children (31 in English-only <i>Literacy Express</i> , 31 in the transitional <i>Literacy Express</i> , 32 in the control group). All children were born in the United States and lived in households in which Spanish was the primary language. Children receiving resource help for speech and language delays were not eligible for the study. The children in the sample were age 54.5 months, on average, and 46% were female.
<b>Setting</b>	The study was conducted in a Head Start program in inner-city Los Angeles, California.
<b>Intervention</b>	The intervention consisted of activities in dialogic reading, phonological awareness, and print knowledge. Dialogic reading activities included scaffolding techniques, such as asking “Wh-” and open-ended questions and using expansions and repetitions to encourage children to talk about the book. Phonological awareness involved word games that used picture puzzles to teach children that words were made of smaller sound units. Print knowledge activities taught children about the alphabet, including recognizing letters and their associated sounds. The intervention was delivered to small groups of four to five children in separate classrooms adjacent to the regular classrooms. The groups met for 20 minutes, four times a week, from mid-November to mid-May and were led by trained bilingual graduate students. In the Spanish-transition condition, instruction was in Spanish for the first eight weeks, followed by three to four weeks of transition. All lessons starting around week 14 were delivered in English. Children in the English-only condition received the full 21 weeks of lessons in English. <sup>1</sup>
<b>Comparison</b>	The comparison group received the <i>High/Scope</i> curriculum, which was typically offered in the center.
<b>Primary outcomes and measurement</b>	The primary outcome domains assessed were oral language, print knowledge, and phonological processing, all of which were assessed with standardized measures. Oral language was assessed with the Receptive Vocabulary and Definitional Vocabulary subtests from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP). Print knowledge was assessed with the Print Knowledge subtest from the Pre-CTOPPP. Phonological processing was assessed with the Blending and Elision subtests from the Pre-CTOPPP. Pretesting was done in fall of the preschool year, and posttesting was done in spring of the preschool year. Assessments were administered by trained research staff who were not involved in the delivery of the intervention and were blind to the children’s treatment status. Assessments were conducted with all children in English and Spanish; only the English assessments are used in the rating of the intervention. Outcomes for the Spanish assessments are included in Appendix 4.1–A4.3, A4.5, A4.7, and A4.9. For a more detailed description of the outcome measures, see Appendices A2.1–A2.5.
<b>Staff/teacher training</b>	Four bilingual graduate students were taught to deliver the intervention by one of the study’s authors, who also supervised the intervention. No other information on training is provided.

1. The study authors analyzed the English-only *Literacy Express* and transitional *Literacy Express* separately. The WWC uses the combined data to determine the rating of effectiveness for this WWC report, but provides the separate comparisons in Appendices A4.2–A4.9.

## Appendix A1.2 Study characteristics: Lonigan, Farver, Clancy-Menchetti, & Phillips, 2005

Characteristic	Description
<b>Study citation</b>	Lonigan, C. J., Farver, J. M., Clancy-Menchetti, J., & Phillips, B. M. (2005, June). <i>Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models</i> . Paper presented at the 12th annual meeting of the Society for the Scientific Study of Reading, Toronto, Ontario, Canada.
<b>Participants</b>	Eighteen preschools from Tallahassee, Florida, and 30 preschools from Los Angeles, California, were randomly assigned to a <i>Literacy Express</i> workshop group, a <i>Literacy Express</i> workshop plus mentoring group, or a control condition. This resulted in 15 preschools in each of the two <i>Literacy Express</i> groups and 18 preschools in the control 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 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.
<b>Setting</b>	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.
<b>Intervention</b>	Preschools in the intervention group participated in a <i>Literacy Express</i> plus professional development via workshops group ("workshop group") or a <i>Literacy Express</i> plus professional development via workshops and mentoring group ("mentoring group"). <i>Literacy Express</i> is a preschool literacy-focused curriculum that is intended to promote children's emergent literacy skills. The curriculum used in this study (an earlier version) was structured around 11 thematic units (with games and activities in each unit) that were sequenced in order of complexity. In the earlier and the current versions of this curriculum, each unit includes children's books that address theme-relevant vocabulary for small- and large-group reading activities. Each thematic unit includes small-group activities that provide children with the opportunity to attend to and practice skills related to oral language, phonological sensitivity, and print awareness and to receive individual feedback. Small-group activities are conducted three or four times a week. The curriculum provides guidance to teachers on grouping children who are progressing at similar rates. Large-group and extension activities provide opportunities for children to use new skills. The workshop group teachers participated in two-day workshops at the beginning of the school year and three half-day workshops during the school year. The mentoring group teachers participated in the same workshops and received regular classroom visits by a trained project mentor. <sup>1</sup>
<b>Comparison</b>	Classrooms in the business-as-usual comparison group participated in the preschool's standard curriculum, which in most cases was <i>High/Scope</i> or <i>Creative Curriculum</i> <sup>®</sup> .
<b>Primary outcomes and measurement</b>	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 subtest 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 Processing (Pre-CTOPPP). Phonological processing was assessed with the Blending and Elision subtests from the Pre-CTOPPP. Cognition was assessed with three subtests from the Pre-CTOPPP: Non-Word Repetition, Word Span, and Rapid Object Naming. Pretesting was done in fall of the preschool year, and posttesting was done in spring of the preschool year. Trained research staff administered all assessments in a quasi-random order at pretest and posttest. All assessments were conducted with all children in English. Research staff also observed the study classrooms for three hours twice a year to determine implementation fidelity and to administer two general measures of classroom language and literacy, but these measures are not discussed further in this WWC intervention report. For a more detailed description of the outcome measures, see Appendices A2.1–A2.5.
<b>Staff/teacher training</b>	The research staff provided all materials and training for the <i>Literacy Express</i> intervention groups. Classroom 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 provide comparisons between the *Literacy Express* mentoring and workshop groups and the control group, between English speakers in the California sample and the control group, and between Spanish speakers in the California sample and the control group. The WWC includes the results from these analyses in Appendices A4.1–A4.4. The study authors also provide comparisons between the *Literacy Express* mentoring group and the *Literacy Express* workshop group. The WWC includes the results from these analyses in Appendices A4.5–A4.8.

## Appendix A1.3 Study characteristics: Preschool Curriculum Evaluation Research (PCER) Consortium, 2008

Characteristic	Description
<b>Study citation</b>	Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). <i>Literacy Express</i> and <i>DLM Early Childhood Express</i> supplemented with <i>Open Court Reading Pre-K</i> . Florida State University. In <i>Effects of preschool curriculum projects on school readiness</i> (pp. 117–130). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Research.
<b>Participants</b>	This study, conducted during the 2003/04 and 2004/05 school years, included three intervention groups: <i>Literacy Express</i> , <i>DLM Early Childhood Express</i> supplemented with <i>Open Court Reading</i> , and a control group. Eighteen schools were initially recruited to participate in the study. Sixteen of these were assigned a letter grade of A, B, C, or D as part of Florida’s school grading system (2 schools were not part of the grading system and were dropped from the study). School grades were used as a blocking variable, and schools within each grade were ranked on average number of years of teaching experience. One additional school joined the study late (for a total of 17 participating schools). Schools were then grouped into triplets and randomized into three conditions: <i>Literacy Express</i> (6 schools), <i>DLM</i> (5 schools), and control (6 schools). Although schools were randomized into three groups, this review is restricted to a comparison of <i>Literacy Express</i> with the control group. <sup>1</sup> School is the unit of assignment; if a school had multiple preschool classrooms, all of those classrooms were assigned to the same intervention. The study as reviewed included 6 <i>Literacy Express</i> and 6 control schools and 10 <i>Literacy Express</i> and 9 control classrooms and began with a total of 196 children (99 <i>Literacy Express</i> and 97 control). The parental consent rate was 95% for the combined treatment group and 93% for the control group. At baseline, children in the study averaged age 4.6; 54% were male; and 6% were Hispanic, 30% were Caucasian, and 59% were African American. The analysis sample included between 177 and 188 children, depending on the outcome measure. There was no attrition of schools. Depending on the outcome, child-level attrition ranged from 6% to 10% for <i>Literacy Express</i> and from 2% to 9% for the control group.
<b>Setting</b>	The <i>Literacy Express</i> study was conducted with children from 12 schools and 19 classrooms selected from public prekindergarten programs in Florida.
<b>Intervention</b>	<i>Literacy Express</i> is a preschool literacy-focused curriculum that is intended to promote children’s emergent literacy skills. The version of the curriculum used in this study was structured around 11 thematic units (with games and activities in each unit). The version used in this study and the current version of the curriculum sequence the units in order of complexity. Each unit includes children’s books that address theme-relevant vocabulary for small- and large-group reading activities. Each thematic unit includes small-group activities that provide children with the opportunity to attend to and practice skills related to oral language, phonological sensitivity, and print awareness and to receive individual feedback. Small-group activities are conducted three or four times a week. The curriculum provides guidance to teachers on grouping children who are progressing at similar rates. Large-group and extension activities provide opportunities for children to use new skills. Fidelity observations were conducted in treatment and control classrooms during February 2004 and April/May 2004 using the Early Language and Literacy Classroom Observation (ELLCO) tool and the Center for Improving the Readiness of Children for Learning and Education (CIRCLE) observation tools. Observations lasted 2.5 to 3 hours in each classroom. Researchers used site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High (3.0), Medium (2.0), Low (1.0), or Not at All (0.0). Researchers also provided a global rating for the control group. <i>Literacy Express</i> was rated in the High-Medium range (2.5) on the global implementation fidelity measure, whereas the control group was rated at the Medium level (2.0).
<b>Comparison</b>	Teachers of control group classrooms were trained to use the <i>High/Scope</i> curriculum. Training provided to teachers in control classrooms included a week-long summer institute conducted by <i>High/Scope</i> trainers prior to the start of the project, additional training sessions throughout the school year conducted by <i>High/Scope</i> and district personnel, and classroom visits by the <i>High/Scope</i> trainer.

(continued)

## Appendix A1.3 Study characteristics: Preschool Curriculum Evaluation Research (PCER) Consortium, 2008 (continued)

Characteristic	Description
<b>Primary outcomes and measurement</b>	The primary outcome domains assessed were children’s oral language, print knowledge, phonological processing, and math, all of which were assessed with standardized measures. Oral language was assessed with the Peabody Picture Vocabulary Test III (PPVT-III) and the Grammatic Understanding subtest from the Test of Language Development–Primary III (TOLD-P:3). Print knowledge was assessed with the Test of Early Reading Ability–III (TERA-3) and the Woodcock-Johnson III (WJ-III) Letter-Word Identification and Spelling subtests. Phonological processing was assessed with the Elision subtest from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP). <sup>2</sup> Math was assessed with the WJ-III Applied Problems subtest, the Composite Score from the Child Math Assessment–Abbreviated (CMA-A), and the Building Blocks Shape Composition test. Pretesting was done in fall of the preschool year, and posttesting was done in spring of the preschool year. <sup>3</sup> Trained research staff administered all assessments, which were conducted with all children in English. Research staff also observed the study classrooms for three hours twice a year using the ELLCO and CIRCLE observation measures, but these measures are not discussed further in this WWC intervention report. For a more detailed description of these outcome measures, see Appendices A2.1–A2.5.
<b>Staff/teacher training</b>	Teachers received curriculum training from the Florida research team for four days in July 2003, prior to the start of the 2003/04 school year. The first two days of the training were spent in a workshop setting, and the other two days were used for team planning. The workshop training session familiarized teachers and their aides with the curriculum materials and provided hands-on experience in leading curricular activities. Videotaped training was made available for teachers who could not attend in person. Teachers and aides attended a two-hour training session every other month during the school year.

1. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing *Literacy Express* with the control group. The WWC does not include the *DLM Express* plus *Open Court Pre-K* versus control results 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. The CTOPP Elision test was used in the kindergarten follow-up reported in Appendix A4.14.
3. The study authors also provide comparisons of the *Literacy Express* group with the control group based on follow-up data collected in spring of the kindergarten year. The WWC includes the results from these analyses in Appendices A4.9–A4.12.

## Appendix A2.1 Outcome measures for the oral language domain

Outcome measure	Description
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Receptive Vocabulary subtest</b>	A subtest from a standardized measure that assesses receptive vocabulary by asking a child to point to one of four pictures that best represents a word spoken by the assessor (as cited in Farver, Lonigan, & Eppe, 2009).
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Definitional Vocabulary subtest</b>	A subtest from a standardized measure that assesses children's single-word oral and expressive vocabulary. The child is asked to name what is shown in a picture and describe one important feature (as cited in Farver, Lonigan, & Eppe, 2009).
<b>Peabody Picture Vocabulary Test—III (PPVT-III)</b>	A standardized measure of children's receptive vocabulary in which children show understanding of a spoken word by pointing to a picture that best represents the meaning (as cited in PCER Consortium, 2008).
<b>Test of Language Development—Primary III (TOLD-P:3)—Grammatical Understanding subtest</b>	A standardized measure of children's ability to comprehend the meaning of sentences by selecting pictures that most accurately represent the sentence (as cited in PCER Consortium, 2008).
<b>Preschool Language Scales—IV (PLS-IV)—Expressive Communication subtest</b>	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 for the print knowledge domain

Outcome measure	Description
<b>Test of Early Reading Ability—III (TERA-3)—Total Score</b>	A standardized measure of children's developing reading skills with three subtests: Alphabet, Conventions, and Meaning (as cited in PCER Consortium, 2008). <sup>1</sup>
<b>Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest</b>	A standardized measure of identification of letters and reading of words (as cited in PCER Consortium, 2008).
<b>Woodcock-Johnson III (WJ-III) Spelling subtest</b>	A standardized measure that assesses children's prewriting skills, such as drawing lines, tracing, and writing letters (as cited in PCER Consortium, 2008).
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Print Knowledge subtest</b>	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 and Farver, Lonigan, & Eppe, 2009).

1. By name, this measure sounds as if it should be captured under the early reading and writing domain; however, the description of the measure identifies constructs that are pertinent to print knowledge, such as knowing the alphabet, understanding print conventions, and environmental print.

### Appendix A2.3 Outcome measures for the phonological processing domain

Outcome measure	Description
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Elision subtest</b>	A measure of children's ability to identify and manipulate sounds in spoken words, using word prompts and picture plates for the first nine items and word prompts only for later items (as cited in PCER Consortium, 2008).
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Blending subtest</b>	A subtest from a standardized measure that requires children to combine words, syllables, or phonemes to make a new word (as cited in Farver, Lonigan, & Eppe, 2009).
<b>Comprehensive Test of Phonological Processing (CTOPP)—Elision subtest</b>	The CTOPP Elision subtest assesses phonological awareness and is similar to the Pre-CTOPPP Elision subtest but does not include pictures in the administration format (as cited in PCER Consortium, 2008).

### Appendix A2.4 Outcome measures for the cognition domain

Outcome measure	Description
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Non-Word Repetition subtest</b>	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).
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Word Span subtest</b>	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).
<b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—Rapid Object Naming subtest</b>	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 for the math domain

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

## Appendix A3.1 Summary of study findings included in the rating for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>Farver, Lonigan, &amp; Eppe, 2009<sup>8</sup></b>								
Pre-CTOPPP Receptive Vocabulary subtest	Preschoolers	10/94	31.21 (4.99)	28.33 (5.63)	2.88	0.55	Statistically significant	+21
Pre-CTOPPP Definitional Vocabulary subtest	Preschoolers	10/94	49.87 (12.66)	41.23 (16.85)	8.64	0.60	Statistically significant	+23
<b>Average for oral language (Farver, Lonigan, &amp; Eppe, 2009)<sup>9</sup></b>						<b>0.57</b>	<b>Statistically significant</b>	<b>+22</b>
<b>Lonigan et al., 2005<sup>8</sup></b>								
PLS-IV Expressive Communication subtest	Preschoolers	48/722	53.35 (8.78)	50.66 (9.71)	2.69	0.30	Statistically significant	+12
<b>Average for oral language (Lonigan et al., 2005)<sup>9</sup></b>						<b>0.30</b>	<b>Statistically significant</b>	<b>+12</b>
<b>PCER Consortium, 2008<sup>8</sup></b>								
PPVT-III	Preschoolers	12/186	90.30 (15.43)	87.31 (13.77)	2.99	0.17	ns	+7
TOLD-P:3 Grammatic Understanding subtest	Preschoolers	12/181	8.11 (2.62)	8.33 (2.71)	-0.22	-0.04	ns	-2
<b>Average for oral language (PCER Consortium, 2008)<sup>9</sup></b>						<b>0.06</b>	<b>ns</b>	<b>+2</b>
<b>Domain average for oral language across all studies<sup>9</sup></b>						<b>0.31</b>	<b>na</b>	<b>+12</b>

ns = not statistically significant

na = not applicable

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

PLS-IV = Preschool Language Scale-IV

PPVT-III = Peabody Picture Vocabulary Test-III

TOLD-P:3 = Test of Language Development-Primary III

(continued)

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

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language domain. The findings of Farver, Lonigan, and Eppe (2009) presented in this table combine the *Literacy Express* English-only group with the *Literacy Express* transitional group and compare children in this combined treatment group with children in the comparison group. Findings for each of these groups are presented separately in Appendices A4.1 and A4.5. Subgroup findings from Lonigan et al. (2005) are not included in these ratings but are reported in Appendices A4.1 and A4.6. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.12.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest. In the case of PCER Consortium (2008), each intervention group mean is calculated as the sum of the unadjusted control mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), a correction for clustering was not needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study. In the case of Lonigan et al. (2005), no corrections for multiple comparisons were needed and no correction for clustering was needed because the analysis corrected for clustering by using hierarchical linear models (HLM). In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

## Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>Farver, Lonigan, &amp; Eppe, 2009<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest	Preschoolers	10/94	22.01 (8.47)	16.61 (7.96)	5.40	0.64	Statistically significant	+24
<b>Average for print knowledge (Farver, Lonigan, &amp; Eppe, 2009)<sup>9</sup></b>						<b>0.64</b>	<b>Statistically significant</b>	<b>+24</b>
<b>Lonigan et al., 2005<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest	Preschoolers	48/722	17.55 (9.30)	14.70 (7.83)	2.85	0.32	Statistically significant	+13
<b>Average for print knowledge (Lonigan et al., 2005)<sup>9</sup></b>						<b>0.32</b>	<b>Statistically significant</b>	<b>+13</b>
<b>PCER Consortium, 2008<sup>8</sup></b>								
TERA-3	Preschoolers	12/183	82.36 (12.45)	81.10 (11.00)	1.26	0.17	ns	+7
WJ-III Letter-Word Identification subtest	Preschoolers	12/183	103.30 (14.10)	95.60 (12.38)	7.70	0.30	ns	+12
WJ-III Spelling subtest	Preschoolers	12/182	89.03 (14.34)	87.67 (11.75)	1.36	0.05	ns	+2
<b>Average for print knowledge (PCER Consortium, 2008)<sup>9</sup></b>						<b>0.17</b>	<b>ns</b>	<b>+7</b>
<b>Domain average for print knowledge across all studies<sup>9</sup></b>						<b>0.38</b>	<b>na</b>	<b>+15</b>

ns = not statistically significant

na = not applicable

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

TERA-3 = Test of Early Reading Ability–III

WJ-III = Woodcock-Johnson III

(continued)

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

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge domain. The findings of Farver, Lonigan, and Eppe (2009) presented in this table combine the *Literacy Express* English-only group and the *Literacy Express* transitional group and compare children in this combined group with children in the comparison group. Findings for each of these groups are presented separately in Appendices A4.2 and A4.7. Subgroup findings from Lonigan et al. (2005) are not included in these ratings but are reported in Appendix A4.2 and A4.8. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.13.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest. In the case of PCER Consortium (2008), each intervention group mean is calculated as the unadjusted control mean plus the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), a correction for clustering was not needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study. In the case of Lonigan et al. (2005), no correction for multiple comparisons was needed, and no correction for clustering was needed because the analysis corrected for clustering by using HLM. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

### Appendix A3.3 Summary of study findings included in the rating for the phonological processing domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>Farver, Lonigan, &amp; Eppe, 2009<sup>8</sup></b>								
Pre-CTOPPP Blending subtest	Preschoolers	10/94	14.37 (3.16)	12.69 (3.51)	1.68	0.51	Statistically significant	+19
Pre-CTOPPP Elision subtest	Preschoolers	10/94	8.00 (3.35)	6.37 (1.51)	1.63	0.56	Statistically significant	+21
<b>Average for phonological processing (Farver, Lonigan, &amp; Eppe, 2009)<sup>9</sup></b>						<b>0.54</b>	<b>Statistically significant</b>	<b>+20</b>
<b>Lonigan et al., 2005<sup>8</sup></b>								
Pre-CTOPPP Blending subtest	Preschoolers	48/722	14.15 (4.47)	13.47 (4.47)	0.68	0.15	ns	+6
Pre-CTOPPP Elision subtest	Preschoolers	48/722	8.86 (3.93)	7.43 (3.54)	1.43	0.38	Statistically significant	+15
<b>Average for phonological processing (Lonigan et al., 2005)<sup>9</sup></b>						<b>0.26</b>	<b>Statistically significant</b>	<b>+10</b>
<b>PCER Consortium, 2008<sup>8</sup></b>								
Pre-CTOPPP Elision subtest	Preschoolers	12/188	9.42 (4.54)	8.79 (4.37)	0.63	0.14	ns	+6
<b>Average for phonological processing (PCER Consortium, 2008)<sup>9</sup></b>						<b>0.14</b>	<b>ns</b>	<b>+6</b>
<b>Domain average for phonological processing across all studies<sup>9</sup></b>						<b>0.31</b>	<b>na</b>	<b>+12</b>

ns = not statistically significant

na = not applicable

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

(continued)

### Appendix A3.3 Summary of study findings included in the rating for the phonological processing domain *(continued)*

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the phonological processing domain. The findings of Farver, Lonigan, and Eppe (2009) presented in this table combine the *Literacy Express* English-only group and the *Literacy Express* transitional group and compare children in this combined group with children in the comparison group. Findings for each of these groups are presented separately in Appendices A4.3 and A4.9. Subgroup findings from Lonigan et al. (2005) are not included in these ratings but are reported in Appendices A4.3 and A4.10. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.14.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest. In the case of PCER Consortium (2008), each intervention group mean is calculated as the unadjusted control mean plus the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), a correction for clustering was not needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study. In the case of Lonigan et al. (2005), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

## Appendix A3.4 Summary of study findings included in the rating for the cognition domain<sup>1</sup>

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

ns = not statistically significant

na = not applicable

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

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the cognition domain. Subgroup findings from Lonigan et al. (2005) are not included in these ratings but are reported in Appendix A4.5 and A4.11.
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 WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Lonigan et al. (2005), no correction for clustering was needed and no correction for multiple comparisons was needed because no impacts were 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 than comparison group children to complete the task (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<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>PCER Consortium, 2008<sup>8</sup></b>								
WJ-III Applied Problems subtest	Preschoolers	12/177	89.12 (15.37)	87.86 (13.77)	1.26	0.05	ns	+2
CMA-A Composite	Preschoolers	12/185	0.51 (0.24)	0.52 (0.21)	-0.01	-0.02	ns	-1
Shape Composition	Preschoolers	12/185	1.54 (1.00)	1.55 (0.94)	-0.01	-0.01	ns	0
<b>Domain average for math<sup>9</sup></b>						<b>0.01</b>	<b>na</b>	<b>0</b>

ns = not statistically significant

na = not applicable

WJ-III = Woodcock-Johnson III

CMA-A = Child Math Assessment–Abbreviated

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.15.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For PCER Consortium (2008), each intervention group mean is calculated as the unadjusted control mean plus the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A4.1 Summary of subgroup findings for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>Farver, Lonigan, &amp; Eppe, 2009 (English-only group)<sup>8</sup></b>								
Pre-CTOPPP Receptive Vocabulary subtest (English)	Preschoolers	10/63	30.62 (5.85)	28.33 (5.63)	2.29	0.39	ns	+15
Pre-CTOPPP Definitional Vocabulary subtest (English)	Preschoolers	10/63	47.45 (12.96)	41.23 (16.85)	6.22	0.41	ns	+16
Pre-CTOPPP Receptive Vocabulary subtest (Spanish)	Preschoolers	10/63	24.58 (4.07)	23.79 (4.03)	0.79	0.19	ns	+8
Pre-CTOPPP Definitional Vocabulary subtest (Spanish)	Preschoolers	10/63	25.90 (19.30)	25.74 (15.97)	0.16	0.01	ns	0
<b>Farver, Lonigan, &amp; Eppe, 2009 (transitional group)<sup>8</sup></b>								
Pre-CTOPPP Receptive Vocabulary subtest (English)	Preschoolers	10/63	31.79 (3.95)	28.33 (5.63)	3.46	0.70	Statistically significant	+26
Pre-CTOPPP Definitional Vocabulary subtest (English)	Preschoolers	10/63	52.28 (12.07)	41.23 (16.85)	11.05	0.74	Statistically significant	+27
Pre-CTOPPP Receptive Vocabulary subtest (Spanish)	Preschoolers	10/63	27.03 (5.74)	23.79 (4.03)	3.24	0.65	Statistically significant	+24
Pre-CTOPPP Definitional Vocabulary subtest (Spanish)	Preschoolers	10/63	32.66 (18.30)	25.74 (15.97)	6.92	0.40	ns	+15
<b>Lonigan et al., 2005 (workshop group)<sup>8</sup></b>								
PLS-IV Expressive Communication subtest	Preschoolers	33/442	52.63 (9.68)	50.66 (9.71)	1.97	0.20	ns	+8
<b>Lonigan et al., 2005 (mentoring group)<sup>8</sup></b>								
PLS-IV Expressive Communication subtest	Preschoolers	33/516	54.04 (7.97)	50.66 (9.71)	3.38	0.38	Statistically significant	+15
<b>Lonigan et al., 2005 (English speakers in the California sample)<sup>8</sup></b>								
PLS-IV Expressive Communication subtest	Preschoolers	30/198 <sup>9</sup>	55.12 (6.40)	51.02 (8.01)	4.10	0.59	Statistically significant	+22

(continued)

**Appendix A4.1 Summary of subgroup findings for the oral language domain (continued)**

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> (Literacy Express – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Literacy Express group <sup>3</sup>	Comparison group				
<b>Lonigan et al., 2005 (Spanish speakers in the California sample)<sup>8</sup></b>								
PLS-IV Expressive Communication subtest	Preschoolers	30/231 <sup>9</sup>	46.62 (10.11)	44.04 (10.76)	2.58	0.25	ns	+10

ns = not statistically significant

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

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. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest.
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 WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), a correction for clustering was not needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study. In the case of Lonigan et al. (2005), no correction for multiple comparisons was needed and no correction for clustering was needed because the analysis used HLM.
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.

## Appendix A4.2 Summary of subgroup findings for the print knowledge domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>Farver, Lonigan, &amp; Eppe, 2009 (English-only group)<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest (English)	Preschoolers	10/63	20.11 (9.01)	16.61 (7.96)	3.50	0.41	ns	+16
Pre-CTOPPP Print Knowledge subtest (Spanish)	Preschoolers	10/63	13.14 (6.62)	12.83 (6.28)	0.31	0.05	ns	+2
<b>Farver, Lonigan, &amp; Eppe, 2009 (transitional group)<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest (English)	Preschoolers	10/63	23.90 (7.56)	16.61 (7.96)	7.29	0.93	Statistically significant	+32
Pre-CTOPPP Print Knowledge subtest (Spanish)	Preschoolers	10/63	16.54 (8.90)	12.83 (6.28)	3.71	0.48	Statistically significant	+18
<b>Lonigan et al., 2005 (workshop group)<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest	Preschoolers	33/442	16.19 (9.21)	14.72 (7.83)	1.47	0.17	ns	+7
<b>Lonigan et al., 2005 (mentoring group)<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest	Preschoolers	33/516	18.77 (9.16)	14.72 (7.83)	4.05	0.47	Statistically significant	+18
<b>Lonigan et al., 2005 (English speakers in the California sample)<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest	Preschoolers	30/198 <sup>9</sup>	17.40 (9.51)	15.32 (8.30)	2.08	0.23	ns	+9
<b>Lonigan et al., 2005 (Spanish speakers in the California sample)<sup>8</sup></b>								
Pre-CTOPPP Print Knowledge subtest	Preschoolers	30/231 <sup>9</sup>	15.32 (8.73)	11.29 (5.21)	4.03	0.52	Statistically significant	+20

ns = not statistically significant

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

(continued)

## Appendix A4.2 Summary of subgroup findings for the print knowledge domain *(continued)*

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. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest.
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 WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), a correction for clustering was not needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study. In the case of Lonigan et al. (2005), no correction for multiple comparisons was needed, and no correction for clustering was needed because the analysis used HLM.
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.

### Appendix A4.3 Summary of subgroup findings for the phonological processing domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>Farver, Lonigan, &amp; Eppe, 2009 (English-only group)<sup>8</sup></b>								
Pre-CTOPPP Blending subtest (English)	Preschoolers	10/63	14.31 (3.33)	12.69 (3.51)	1.62	0.47	ns	+18
Pre-CTOPPP Elision subtest (English)	Preschoolers	10/63	7.96 (3.24)	6.37 (1.51)	1.59	0.62	Statistically significant	+23
Pre-CTOPPP Blending subtest (Spanish)	Preschoolers	10/63	11.13 (2.49)	10.59 (3.02)	0.54	0.19	ns	+8
Pre-CTOPPP Elision subtest (Spanish)	Preschoolers	10/63	5.94 (1.75)	5.52 (1.32)	0.42	0.27	ns	+11
<b>Farver, Lonigan, &amp; Eppe, 2009 (transitional group)<sup>8</sup></b>								
Pre-CTOPPP Blending subtest (English)	Preschoolers	10/63	14.43 (3.04)	12.69 (3.51)	1.74	0.52	Statistically significant	+20
Pre-CTOPPP Elision subtest (English)	Preschoolers	10/63	8.04 (3.51)	6.37 (1.51)	1.67	0.61	Statistically significant	+23
Pre-CTOPPP Blending subtest (Spanish)	Preschoolers	10/63	12.71 (4.06)	10.59 (3.02)	2.12	0.59	Statistically significant	+22
Pre-CTOPPP Elision subtest (Spanish)	Preschoolers	10/63	7.40 (2.95)	5.52 (1.32)	1.88	0.82	Statistically significant	+29
<b>Lonigan et al., 2005 (workshop group)<sup>8</sup></b>								
Pre-CTOPPP Blending subtest	Preschoolers	33/442	14.23 (4.34)	13.47 (4.47)	0.76	0.17	ns	+7
Pre-CTOPPP Elision subtest	Preschoolers	33/442	8.91 (3.79)	7.43 (3.54)	1.48	0.40	Statistically significant	+16
<b>Lonigan et al., 2005 (mentoring group)<sup>8</sup></b>								
Pre-CTOPPP Blending subtest	Preschoolers	33/516	14.14 (4.54)	13.47 (4.47)	0.67	0.15	ns	+6
Pre-CTOPPP Elision subtest	Preschoolers	33/516	8.86 (4.03)	7.43 (3.54)	1.43	0.37	Statistically significant	+15

(continued)

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> (Literacy Express – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Literacy Express group <sup>3</sup>	Comparison group				
<b>Lonigan et al., 2005 (English speakers in the California sample)<sup>8</sup></b>								
Pre-CTOPPP Blending subtest	Preschoolers	30/198 <sup>9</sup>	14.70 (4.49)	14.10 (4.12)	0.60	0.14	ns	+5
Pre-CTOPPP Elision subtest	Preschoolers	30/198 <sup>9</sup>	9.07 (3.43)	7.75 (3.32)	1.32	0.39	Statistically significant	+15
<b>Lonigan et al., 2005 (Spanish speakers in the California sample)<sup>8</sup></b>								
Pre-CTOPPP Blending subtest	Preschoolers	30/231 <sup>9</sup>	12.94 (4.83)	12.18 (4.38) <sup>10</sup>	0.76	0.16	ns	+6
Pre-CTOPPP Elision subtest	Preschoolers	30/231 <sup>9</sup>	6.94 (3.40)	6.11 (2.09)	0.83	0.27	Statistically significant	+11

ns = not statistically significant

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

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. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest.
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 WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), a correction for clustering was not needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study. In the case of Lonigan et al. (2005), a correction for 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 and the number of children in each group at posttest was provided by the study authors upon WWC request.
10. The correct standard deviation was provided by the study authors upon WWC request.

## Appendix A4.4 Summary of subgroup findings for the cognition domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Literacy Express</i> group	Comparison group				
<b>Lonigan et al., 2005 (workshop group)<sup>7</sup></b>								
Pre-CTOPPP Non-Word Repetition subtest	Preschoolers	33/442	9.29 (3.96)	9.55 (4.36)	-0.26	-0.06	ns	-2
Pre-CTOPPP Word Span subtest	Preschoolers	33/442	8.84 (2.42)	8.53 (2.54)	0.31	0.12	ns	+5
Pre-CTOPPP Rapid Object Naming subtest	Preschoolers	33/442	48.51 (20.36) <sup>8</sup>	49.57 (16.93)	1.06 <sup>9</sup>	0.06	ns	+2
<b>Lonigan et al., 2005 (mentoring group)<sup>7</sup></b>								
Pre-CTOPPP Non-Word Repetition subtest	Preschoolers	33/516	8.89 (4.29)	9.55 (4.36)	-0.66	-0.15	ns	-6
Pre-CTOPPP Word Span subtest	Preschoolers	33/516	8.88 (2.66)	8.53 (2.54)	0.35	0.13	ns	+5
Pre-CTOPPP Rapid Object Naming subtest	Preschoolers	33/516	47.87 (14.34)	49.57 (16.93)	1.70 <sup>9</sup>	0.11	ns	+4
<b>Lonigan et al., 2005 (English speakers in the California sample)<sup>7</sup></b>								
Pre-CTOPPP Non-Word Repetition subtest	Preschoolers	30/198 <sup>10</sup>	10.62 (4.16)	9.55 (3.98)	1.07	0.26	ns	+10
Pre-CTOPPP Word Span subtest	Preschoolers	30/198 <sup>10</sup>	9.30 (2.57)	8.57 (2.41)	0.73	0.29	ns	+11
Pre-CTOPPP Rapid Object Naming subtest	Preschoolers	30/198 <sup>10</sup>	44.89 (14.90)	46.24 (15.11)	1.35 <sup>9</sup>	0.09	ns	+4
<b>Lonigan et al., 2005 (Spanish speakers in the California sample)<sup>7</sup></b>								
Pre-CTOPPP Non-Word Repetition subtest	Preschoolers	30/231 <sup>10</sup>	8.52 (3.89)	8.38 (3.59)	0.14	0.04	ns	+2
Pre-CTOPPP Word Span subtest	Preschoolers	30/231 <sup>10</sup>	7.47 (2.29)	7.37 (2.32)	0.10	0.04	ns	+2
Pre-CTOPPP Rapid Object Naming subtest	Preschoolers	30/231 <sup>10</sup>	50.03 (18.92)	47.37 (11.52)	-2.66 <sup>9</sup>	-0.16	ns	-6

(continued)

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

ns = not statistically significant

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

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 WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Lonigan et al. (2005), no correction for clustering was needed and no correction for multiple comparisons was needed because no impacts were statistically significant.
8. The correct standard deviation 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 intervention group children took less time than comparison group children to complete the task (comparison group mean minus the intervention group mean).
10. 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.

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		WWC calculations			
			<i>Literacy Express</i> English-only group <sup>3</sup>	<i>Literacy Express</i> transitional group	Mean difference <sup>4</sup> ( <i>Literacy Express</i> English-only – <i>Literacy Express</i> transitional)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
<b>Farver, Lonigan, &amp; Eppe, 2009<sup>8</sup></b>								
Pre-CTOPPP Receptive Vocabulary subtest (English)	Preschoolers	10/62	30.62 (5.85)	31.79 (3.95)	-1.17	-0.23	ns	-9
Pre-CTOPPP Definitional Vocabulary subtest (English)	Preschoolers	10/62	47.45 (12.96)	52.28 (12.07)	-4.83	-0.38	ns	-15
Pre-CTOPPP Receptive Vocabulary subtest (Spanish)	Preschoolers	10/62	24.58 (4.07)	27.03 (5.74)	-2.45	-0.49	ns	-19
Pre-CTOPPP Definitional Vocabulary subtest (Spanish)	Preschoolers	10/62	25.90 (19.30)	32.66 (18.30)	-6.76	-0.35	ns	-14

ns = not statistically significant

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

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. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest.
4. Positive differences and effect sizes favor the first group (in this case, *Literacy Express* English-only group); negative differences and effect sizes favor the second group (in this case, *Literacy Express* transitional group).
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), a correction for clustering was not needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study.

## Appendix A4.6

## Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					WWC calculations	
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Literacy Express</i> mentoring – <i>Literacy Express</i> workshop)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>	
			<i>Literacy Express</i> mentoring group	<i>Literacy Express</i> workshop group					
<b>Lonigan et al., 2005<sup>7</sup></b>									
PLS-IV Expressive Communication subtest	Preschoolers	30/486	54.04 (7.97)	52.63 (9.68)	1.41	0.16	ns	+6	

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 first group (in this case, *Literacy Express* mentoring group); negative differences and effect sizes favor the second group (in this case, *Literacy Express* workshop group).
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Lonigan et al. (2005), no correction for multiple comparisons was needed and no correction for clustering was needed because the analysis used HLM.

**Appendix A4.7**

**Summary of findings for comparisons between *Literacy Express* English-only and *Literacy Express* transitional for the print knowledge domain<sup>1</sup>**

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					WWC calculations	
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> English-only – <i>Literacy Express</i> transitional)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>	
			<i>Literacy Express</i> English-only group <sup>3</sup>	<i>Literacy Express</i> transitional group					
<b>Farver, Lonigan, &amp; Eppe, 2009<sup>8</sup></b>									
Pre-CTOPPP Print Knowledge subtest (English)	Preschoolers	10/62	20.11 (9.01)	23.90 (7.56)	-3.79	-0.45	ns	-17	
Pre-CTOPPP Print Knowledge subtest (Spanish)	Preschoolers	10/62	13.14 (6.62)	16.54 (8.90)	-3.40	-0.43	ns	-17	

ns = not statistically significant

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

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. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest.
4. Positive differences and effect sizes favor the first group (in this case, *Literacy Express* English-only group); negative differences and effect sizes favor the second group (in this case, *Literacy Express* transitional group).
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), no correction for clustering was needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study.

**Appendix A4.8**

**Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the print knowledge domain<sup>1</sup>**

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Literacy Express</i> mentoring – <i>Literacy Express</i> workshop)	WWC calculations		
			<i>Literacy Express</i> mentoring group	<i>Literacy Express</i> workshop group		Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
<b>Lonigan et al., 2005<sup>7</sup></b>								
Pre-CTOPPP Print Knowledge subtest	Preschoolers	30/486	18.77 (9.16)	16.19 (9.21)	2.58	0.28	Statistically significant	+11

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

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 first group (in this case, *Literacy Express* mentoring group); negative differences and effect sizes favor the second group (in this case, *Literacy Express* workshop group).
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Lonigan et al. (2005), no correction for multiple comparisons was needed and no correction for clustering was needed because the analysis used HLM.

**Appendix A4.9**

**Summary of findings for comparisons between *Literacy Express* English-only and *Literacy Express* transitional for the phonological processing domain<sup>1</sup>**

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		WWC calculations			
			<i>Literacy Express</i> English-only group <sup>3</sup>	<i>Literacy Express</i> transitional group	Mean difference <sup>4</sup> ( <i>Literacy Express</i> English-only – <i>Literacy Express</i> transitional)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
<b>Farver, Lonigan, &amp; Eppe, 2009<sup>8</sup></b>								
Pre-CTOPPP Blending subtest (English)	Preschoolers	10/62	14.31 (3.33)	14.43 (3.04)	–0.12	–0.04	ns	–1
Pre-CTOPPP Elision subtest (English)	Preschoolers	10/62	7.96 (3.24)	8.04 (3.51)	–0.08	–0.02	ns	–1
Pre-CTOPPP Blending subtest (Spanish)	Preschoolers	10/62	11.13 (2.49)	12.71 (4.06)	–1.58	–0.46	ns	–18
Pre-CTOPPP Elision subtest (Spanish)	Preschoolers	10/62	5.94 (1.75)	7.40 (2.95)	–1.46	–0.59	ns	–22

ns = not statistically significant

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

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. In the case of Farver, Lonigan, and Eppe (2009), the means are adjusted for chronological age and pretest.
4. Positive differences and effect sizes favor the first group (in this case, *Literacy Express* English-only group); negative differences and effect sizes favor the second group (in this case, *Literacy Express* transitional group).
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Farver, Lonigan, and Eppe (2009), no correction for clustering was needed; however, a correction for multiple comparisons was needed, so the significance levels may differ from the original study.

**Appendix A4.10**

**Summary of findings for comparisons between *Literacy Express* mentoring and *Literacy Express* workshop for the phonological processing domain<sup>1</sup>**

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		WWC calculations			
			<i>Literacy Express</i> mentoring group	<i>Literacy Express</i> workshop group	Mean difference <sup>3</sup> ( <i>Literacy Express</i> mentoring – <i>Literacy Express</i> workshop)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
<b>Lonigan et al., 2005<sup>7</sup></b>								
Pre-CTOPPP Blending subtest	Preschoolers	30/486	14.14 (4.54)	14.23 (4.34)	-0.09	-0.02	ns	-1
Pre-CTOPPP Elision subtest	Preschoolers	30/486	8.86 (4.03)	8.91 (3.79)	-0.05	-0.01	ns	-1

ns = not statistically significant

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

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 first group (in this case, *Literacy Express* mentoring group); negative differences and effect sizes favor the second group (in this case, *Literacy Express* workshop group).
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Lonigan et al. (2005), no correction for clustering was needed because the analysis used HLM. A correction for multiple comparisons was needed, however, so the significance levels may differ from the original study.

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					WWC calculations	
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Literacy Express</i> mentoring – <i>Literacy Express</i> workshop)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>	
			<i>Literacy Express</i> mentoring group	<i>Literacy Express</i> workshop group					
<b>Lonigan et al., 2005<sup>7</sup></b>									
Pre-CTOPPP Non-Word Repetition subtest	Preschoolers	30/486	8.89 (4.29)	9.29 (3.96)	–0.40	–0.10	ns	–4	
Pre-CTOPPP Word Span subtest	Preschoolers	30/486	8.88 (2.66)	8.84 (2.42)	0.04	0.02	ns	+1	
Pre-CTOPPP Rapid Object Naming subtest	Preschoolers	30/486	47.87 (14.34)	48.51 (20.36) <sup>8</sup>	0.64 <sup>9</sup>	0.04	ns	+1	

ns = not statistically significant

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

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 first group (in this case, *Literacy Express* mentoring group); negative differences and effect sizes favor the second group (in this case, *Literacy Express* workshop group).
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Lonigan et al. (2005), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.
8. The correct standard deviation 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 intervention group children took less time than comparison group children to complete the task (comparison group mean minus the intervention group mean).

## Appendix A4.12 Summary of follow-up findings for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>PCER Consortium, 2008<sup>8</sup></b>								
PPVT-III	Kindergarteners	12/151	92.09 (15.12)	89.23 (12.46)	2.86	0.16	ns	+6
TOLD-P:3 Grammatic Understanding subtest	Kindergarteners	12/154	8.96 (3.42)	8.44 (3.25)	0.52	0.10	ns	+4

ns = not statistically significant

PPVT-III = Peabody Picture Vocabulary Test–III

TOLD-P:3 = Test of Language Development–Primary III

1. This appendix presents follow-up findings for measures that fall in the oral language domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For PCER Consortium (2008), each intervention group mean is calculated as the unadjusted control mean plus the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

## Appendix A4.13 Summary of follow-up findings for the print knowledge domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		WWC calculations			
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group	Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
<b>PCER Consortium, 2008<sup>8</sup></b>								
TERA-3	Kindergarteners	12/153	81.21 (15.03)	82.06 (13.38)	–0.85	–0.11	ns	–4
WJ-III Letter-Word Identification subtest	Kindergarteners	12/151	101.90 (14.35)	99.74 (12.15)	2.16	0.08	ns	+3
WJ-III Spelling subtest	Kindergarteners	12/147	99.38 (17.67)	97.83 (13.15)	1.55	0.06	ns	+2

ns = not statistically significant

TERA-3 = Test of Early Reading Ability–III

WJ-III = Woodcock-Johnson III

1. This appendix presents follow-up findings for measures that fall in the print knowledge domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For PCER Consortium (2008), each intervention group mean is calculated as the unadjusted control mean plus the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

## Appendix A4.14 Summary of follow-up findings for the phonological processing domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>PCER Consortium, 2008<sup>8</sup></b>								
CTOPP Elision subtest	Kindergarteners	12/155	2.95 (3.43)	2.68 (2.98)	0.27	0.08	ns	+3

ns = not statistically significant

CTOPP = Comprehensive Test of Phonological Processing

1. This appendix presents follow-up findings for measures that fall in the phonological processing domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For PCER Consortium (2008), each intervention group mean is calculated as the unadjusted control mean plus the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed, because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

## Appendix A4.15 Summary of follow-up findings for the math domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> ( <i>Literacy Express</i> – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			<i>Literacy Express</i> group <sup>3</sup>	Comparison group				
<b>PCER Consortium, 2008<sup>8</sup></b>								
WJ-III Applied Problems subtest	Kindergarteners	12/147	90.10 (17.45)	90.54 (14.23)	–0.44	–0.02	ns	–1
CMA-A Composite	Kindergarteners	12/154	0.52 (0.22)	0.57 (0.21)	–0.05	–0.21	ns	–8
Shape Composition	Kindergarteners	12/153	1.98 (0.97)	2.11 (0.91)	–0.13	–0.14	ns	–6

ns = not statistically significant

WJ-III = Woodcock-Johnson III

CMA-A = Child Math Assessment–Abbreviated

1. This appendix presents follow-up findings for measures that fall in the math domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.5.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For PCER Consortium (2008), each intervention group mean is calculated as the unadjusted control mean plus the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

## Appendix A5.1 Literacy Express rating for the oral language domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of oral language, the WWC rated *Literacy Express* as having positive effects for preschool children. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *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 that measured oral language showed a statistically significant positive effect, and one study showed no effect. All of these studies met WWC evidence standards for a strong design.

### AND

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

**Met.** None of the three studies that measured oral language showed 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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A5.2 Literacy Express rating for the print knowledge domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of print knowledge, the WWC rated *Literacy Express* as having positive effects for preschool children. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *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 that measured print knowledge showed a statistically significant positive effect, and one study showed no effect. All of these studies met WWC evidence standards for a strong design.

### AND

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

**Met.** None of the three studies that measured print knowledge showed 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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

### Appendix A5.3 *Literacy Express* rating for the phonological processing domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of phonological processing, the WWC rated *Literacy Express* as having positive effects for preschool children. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *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 of the studies that measured phonological processing showed a statistically significant positive effect, and one study found no effect. All of the studies met WWC standards for a strong design.

**AND**

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

**Met.** None of the three studies that measured phonological processing showed 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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A5.4 Literacy Express rating for the cognition domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup> For the outcome domain of cognition, the WWC rated *Literacy Express* as having no discernible effects for preschool children.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** The one study that measured cognition showed no statistically significant or substantively important positive or negative effects. No other studies measured cognition.

### Other ratings considered

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

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

**Not met.** The one study that measured cognition showed no statistically significant positive effects.

#### AND

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

**Met.** The one study that measured cognition did not show a statistically significant or substantively important negative effect.

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

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

**Not met.** The one study that measured cognition showed no 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 one study that measured cognition showed no statistically significant or substantively important negative effects. No other studies measured cognition.

**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 one study that measured cognition showed no statistically significant or substantively important positive or negative effects. No other studies measured cognition.

#### OR

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

**Not met.** The one study that measured cognition showed no statistically significant or substantively important positive or negative effects. No other studies measured cognition.

(continued)

## Appendix A5.4 Literacy Express rating for the cognition domain (continued)

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

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important *positive* effect.

**Not met.** The one study that measured cognition showed no statistically significant or substantively important negative effects. No other studies measured cognition.

**OR**

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** The one study that measured cognition showed no statistically significant or substantively important positive effects. No other studies measured cognition.

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

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

**Not met.** The one study that measured cognition showed no statistically significant negative effects. No other studies measured cognition.

**AND**

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

**Met.** The one study that measured cognition showed no statistically significant or substantively important positive effects. No other studies measured cognition.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A5.5 Literacy Express rating for the math domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup> For the outcome domain of math, the WWC rated *Literacy Express* as having no discernible effects for preschool children.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** The one study that measured math showed no statistically significant or substantively important positive or negative effects. No other studies measured math.

### Other ratings considered

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

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

**Not met.** The one study that measured math showed no statistically significant positive effects.

#### AND

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

**Met.** The one study that measured math did not show a statistically significant or substantively important negative effect.

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

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

**Not met.** The one study that measured math showed no 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 one study that measured math showed no statistically significant or substantively important negative effects. No other studies measured math.

**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 one study that measured math showed no statistically significant or substantively important positive or negative effects. No other studies measured math.

#### OR

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

**Not met.** The one study that measured math showed no statistically significant or substantively important positive or negative effects. No other studies measured math.

(continued)

## Appendix A5.5 Literacy Express rating for the math domain (continued)

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

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important *positive* effect.

**Not met.** The one study that measured math showed no statistically significant or substantively important negative effects. No other studies measured math.

**OR**

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** The one study that measured math showed no statistically significant or substantively important positive effects. No other studies measured math.

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

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

**Not met.** The one study that measured math showed no statistically significant negative effects. No other studies measured math.

**AND**

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

**Met.** The one study that measured math showed no statistically significant or substantively important positive effects. No other studies measured math.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence <sup>1</sup>
		Preschool classrooms	Students	
Oral language	3	70	1,002	Medium to large
Print knowledge	3	70	999	Medium to large
Phonological processing	3	70	1,004	Medium to large
Early reading and writing	0	na	na	na
Cognition	1	48	722	Small
Math	1	12	185	Small

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.