What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

Early Childhood Education August 2010

Ladders to Literacy

Program Description[1]

Ladders to Literacy is a supplemental early literacy curriculum composed of more than 70 activities designed to develop children's print/book awareness, metalinguistic awareness, and oral language skills. The curriculum, published in the book *Ladders to Literacy: A Preschool Activity Book, Second Edition*, can be used in a variety of early childhood settings and provides guidance on how teachers can adapt the activities for children with special needs. The activities are intended as models or suggestions that teachers can adopt within an existing curriculum. Although a *Ladders to Literacy* curriculum is also available for kindergarten students (*Ladders to Literacy: A Kindergarten Activity Book*), this intervention report focuses on the preschool curriculum.

Research[2]

One study of *Ladders to Literacy* that falls within the scope of the Early Childhood Education review protocol meets What Works Clearinghouse (WWC) evidence standards, and one study meets WWC evidence standards with reservations. The two studies included 139 preschool children from 26 preschool classrooms in southern New Hampshire.[3]

Based on these two studies, the WWC considers the extent of evidence for *Ladders to Literacy* on preschool children to be medium to large for oral language and small for print knowledge, phonological processing, and math. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Ladders to Literacy* on preschool children in the early reading/writing or cognition domains.

1. The descriptive information for this program was obtained from a publicly available source: the program's website (http://www.brookespublishing.com/store/books/notari-69131/index.htm, downloaded February 2009). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by November 2008.
2. The studies in this report were reviewed using WWC Evidence Standards, Version 2.0 (see the WWC Procedures and Standards Handbook, Chapter III), as described in protocol Version 2.0.
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

Effectiveness

Ladders to Literacy was found to have potentially negative effects on oral language and no discernible effects on print knowledge, phonological processing, and math for preschool children.

	Oral language	*Print knowledge*	*Phonological processing*	*Early reading/ writing*	*Cognition*	*Math*
Rating of effectiveness	Potentially negative effects	No discernible effects	No discernible effects	na	na	No discernible effects
Improvement index[4]	Average: –7 percentile points	Average: –2 percentile points	–6 percentile points	na	na	Average: +1 percentile point
	Range: –15 to +2 percentile points	Range: –12 to +12 percentile points	na	na	na	Range: –6 to +7 percentile points

na = not applicable

Absence of conflict of interest

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

Additional program information

Developer and contact

Developed by Angela Notari-Syverson, Rollanda E. O'Connor, and Patricia F. Vadasy, *Ladders to Literacy* is distributed by Brookes Publishing Company. Address: P.O. Box 10624, Baltimore, MD 21285-0624. Email: custserv@brookespublishing.com. Web: http://www.brookespublishing.com. Telephone: (800) 638-3775. For professional development training, see http://www.brookespublishing.com/onlocation.

Scope of use

According to the developers, *Ladders to Literacy* has been field-tested in a variety of preschool settings with children from a range of cultural and socioeconomic backgrounds. These include sites that serve young children with disabilities in inclusive and special education settings.

Teaching

Ladders to Literacy is a supplemental early childhood curriculum that is published in the book *Ladders to Literacy: A Preschool Activity Book, Second Edition*, which focuses on developing early language and literacy skills. *Ladders to Literacy* addresses three components of literacy development: print/book awareness, metalinguistic awareness, and oral language skills. The print/book awareness section includes activities such as drawing pictures, pretending to write, and creating graphic representations. Metalinguistic awareness activities focus on the identification of sounds, phonemes, and rhymes through lessons such as Clap the Syllables and First Sound Song. The oral language component includes activities designed to enhance children's vocabulary development and to engage children in conversations in which they respond to open-ended questions and reconstruct

4. These numbers show the average and range of student-level improvement indices for all findings across the studies.

Additional program information *(continued)*

past experiences. There are approximately 20 activities in each of the three sections. For each lesson, the authors describe the activity; list the necessary materials; and explain the rationale for the activity, the lesson's overall goal, and the targeted skills. The book provides guidance on how to individualize the activity for children with varying skill levels and how to adapt the lesson for children with special needs. The book also provides an overview of the theoretical framework underlying the curriculum. An appendix provides a variety of activities that parents and children can do together at home to reinforce skills being taught in the classroom. The curriculum also includes a literacy checklist to help teachers monitor children's progress.

Cost

The *Ladders to Literacy: A Preschool Activity Book, Second Edition*, costs $49.95. Professional development for *Ladders to Literacy* is available for an extra cost and consists of a one- or two-day onsite seminar on how to use the curriculum.

Research

Eight studies reviewed by the WWC investigated the effects of *Ladders to Literacy* on preschool children. One study (Russell, 2005) is a randomized controlled trial that meets WWC evidence standards. One study (PCER Consortium, 2008) is a randomized controlled trial that meets WWC evidence standards with reservations. The remaining six studies do not meet either WWC evidence standards or eligibility screens.

Meets evidence standards

Russell (2005) conducted a randomized controlled trial of 5-year-old children from 12 Head Start classrooms in southern New Hampshire. The 12 classrooms were recruited for the pilot year of the national evaluation conducted by the PCER Consortium (2008) study described below. Because of delays in study implementation, Russell (2005) used a posttest-only design to investigate effects on oral language skills during the 2002–03 school year (one year before the PCER Consortium, 2008, study). Thirty-four children participated in the study; 18 were in the treatment classrooms that received *Creative Curriculum*® supplemented with *Ladders to Literacy*, and 16 were in the comparison classrooms that received *Creative Curriculum*® only. At baseline, the children in the study averaged 4.7 years of age, and none of the children were identified as having a disability.

Meets evidence standards with reservations

A study by the PCER Consortium (2008) assessed the effectiveness of *Ladders to Literacy* as part of the Preschool Curriculum Evaluation Research (PCER) effort.[5] The PCER Consortium (2008) used a randomized controlled trial design in which 14 Head Start preschool classrooms from southern New Hampshire were randomly assigned to implement *Ladders to Literacy* as a supplement to the *Creative Curriculum*® or to a control group that implemented the *Creative Curriculum*® without the *Ladders to Literacy* supplement. Eleven of the 14 classrooms were randomly assigned the previous year for the pilot, described above (Russell, 2005); three other classrooms were added to the sample. Pretest and posttest data, collected in the fall and spring of the 2003–04 school year, were obtained for 105 children (54 *Ladders to Literacy* and 51 control). Baseline equivalence on pretests was established for the treatment and control children. The study investigated effects on oral language, print knowledge, phonological processing, and math. Outcomes were assessed at two time points: end of preschool and end of kindergarten. At baseline, children in the study averaged 4.6 years of age, and 25% were identified as having a disability. Although the WWC used only the results at the end of the preschool year to determine the intervention rating, information on the kindergarten findings can be found in Appendices A4.1–A4.4.

5. The PCER Consortium (2008) evaluated a total of 14 preschool curricula, including *Ladders to Literacy*, in comparison to respective control conditions.

Research *(continued)*

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.[6]

The WWC considers the extent of evidence for *Ladders to Literacy* to be medium to large for oral language and small for print knowledge, phonological processing, and math for preschool children. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Ladders to Literacy* in the early reading/writing or cognition domains for preschool children.

Effectiveness

Findings

The WWC review of interventions for Early Childhood Education addresses student outcomes in six domains: oral language, print knowledge, phonological processing, early reading/writing, cognition, and math. The studies included in this report cover four domains: oral language, print knowledge, phonological processing, and math. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Ladders to Literacy* on preschool children.[7]

Oral language. Russell (2005) analyzed the effectiveness of *Ladders to Literacy* on oral language using two measures: mean length of utterance (MLU) and type token ratio (TTR). The author did not find statistically significant effects of *Ladders to Literacy* on either measure, and the effects were not large enough to be considered substantively important according to WWC criteria (that is, an effect size of at least 0.25). According to WWC criteria, this study shows indeterminate effects on oral language.

The PCER Consortium (2008) analyzed the effectiveness of *Ladders to Literacy* on oral language using the Peabody Picture Vocabulary Test–Third Edition (PPVT-III) and the Test of Oral Language Development–Primary III (TOLD-P:3) Grammatic Understanding subtest. The authors show that differences between the *Ladders to Literacy* group and the control group are not statistically significant on the TOLD-P:3 Grammatic Understanding subtest and are not large enough to be considered substantively important according to WWC criteria. However, there was a substantively important (but not statistically significant) negative effect of –0.38 on the PPVT-III. According to WWC criteria, this study shows potentially negative effects on oral language.

Print knowledge. The PCER Consortium (2008) analyzed the effectiveness of *Ladders to Literacy* on print knowledge using the Test of Early Reading Ability–III (TERA-3), the Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest, and the WJ-III Spelling subtest. The authors report that differences between the *Ladders to Literacy* group and the control group are not statistically significant on any of these measures, although, according to WWC criteria, there is a substantively important negative effect of –0.30 on the TERA-3 and a substantively important positive effect of +0.30 on the WJ-III Spelling subtest. According to WWC criteria, this study shows indeterminate effects on print knowledge.

6. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Ladders to Literacy* is in Appendix A6.
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 Russell (2005), a correction for clustering 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.

Effectiveness *(continued)*

Phonological processing. The PCER Consortium (2008) analyzed the effectiveness of *Ladders to Literacy* on phonological processing using the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest. The authors report that the difference between the *Ladders to Literacy* group and the control group is not statistically significant and, according to WWC criteria, is not large enough to be substantively important. According to WWC criteria, this study shows indeterminate effects on phonological processing.

Math. The PCER Consortium (2008) analyzed the effectiveness of *Ladders to Literacy* on math using the WJ-III Applied Problems subtest, the Child Math Assessment–Abbreviated (CMA-A), and the Building Blocks Shape Composition task. The authors report that differences between the *Ladders to Literacy* group and the control group are not statistically significant and, according to WWC criteria, are not large enough to be considered substantively important. According to WWC criteria, this study shows indeterminate effects on math.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Procedures and Standards Handbook, Appendix E).

The WWC found *Ladders to Literacy* to have potentially negative effects on oral language and no discernible effects on print knowledge, phonological processing, and math for preschool children

Improvement index

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

Based on two studies, the average improvement index for *Ladders to Literacy* on oral language is –7 percentile points with a range of –15 to +2 percentile points across findings. Based on one study, the average improvement index on print knowledge is –2 percentile points, with a range of –12 to +12 percentile points across findings; the improvement index on phonological processing is –6 percentile points for a single finding from one study; and the average improvement index on math is +1 percentile point with a range of –6 to +7 percentile points across findings.

Summary

The WWC reviewed eight studies on *Ladders to Literacy* for preschool children. One of these studies meets WWC evidence standards; one study meets WWC evidence standards with reservations; and the remaining six studies do not meet either WWC evidence standards or eligibility screens. Based on the two studies, the WWC found potentially negative effects on oral language and no discernible effects on print knowledge, phonological processing, and math for preschool children. The conclusions presented in this report may change as new research emerges.

References

Meets WWC evidence standards

Russell, J. (2005). *An investigation of preschool oral language improvements through* Ladders to Literacy. Unpublished master's thesis, University of New Hampshire, Durham. (62329791).

Meets WWC evidence standards with reservations

Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). *Creative Curriculum* with *Ladders to Literacy*: University of New Hampshire. In *Effects of preschool curriculum programs on school readiness* (pp. 65–73). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

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

Coston, J. H. (1997). *The effects of a comprehensive curriculum on literacy development*. Unpublished master's thesis, Valdosta State University, GA. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Good, J. L. (2003). Developing early literacy skills in young children with symptoms of inattention and hyperactivity (Doctoral dissertation, University of Minnesota). *Dissertation Abstracts International, 64*(06A), 106–1966. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

McKnight, C., Lee, S., & Schowengerdt, R. (2001). *Effects of specific strategy training on phonemic awareness and reading aloud with preschoolers: A comparison study*. Retrieved from ERIC database (ED452518). The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Notari-Syverson, A. (1999). *Supporting early literacy development in young children with disabilities: A comprehensive interactive emergent literacy curriculum for preschoolers* (Final report to the U.S. Department of Education). Seattle, WA: Washington Research Institute. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Additional source:

Notari-Syverson, A., O'Connor, R. E. & Vadasy, P. F. (1996). *Facilitating language and literacy development in preschool children: To each according to their needs*. Paper presented at the American Educational Research Association Meeting, New York. (ERIC Document ED395692).

Notari-Syverson, A. (2005). Ladders to Literacy *Outreach Project. Final grant performance report*. (Submitted to the U.S. Department of Education No. H324R000008). Seattle, WA: Washington Research Institute. The study is ineligible for review because it does not use a comparison group.

Notari-Syverson, A. (2007). *Model demonstration project for children with disabilities: Final grant performance report* (Submitted to the U.S. Department of Education No. H324M020084). Seattle, WA: Washington Research Institute. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Appendix

Appendix A1.1 Study characteristics: Russell, 2005

Characteristic	Description
Study citation	Russell, J. (2005). *An investigation of preschool oral language improvements through* Ladders to Literacy. Unpublished master's thesis, University of New Hampshire, Durham. (62329791).
Participants[1]	This study was a posttest-only design (no pretest was possible due to delays in study implementation) conducted with children from 12 Head Start classrooms, some of which were included the PCER Consortium (2008) study that is described below. The classrooms were selected in 2002 from a list of prospective study participants and randomly assigned. The researchers first identified four urban full-day classrooms and randomly assigned two to the treatment group and two to the control group. Also, they selected (a) two urban half-day classrooms with high numbers of Spanish-speaking children, (b) two additional urban half-day classrooms, (c) two suburban/rural classrooms from towns with a kindergarten program, and (d) two classrooms from towns with no kindergarten program; from each group, one classroom was randomly assigned to treatment and one to control. The study was conducted in the 2002–03 school year. Study participants were limited to children speaking English as their primary language and not enrolled in a special education program. Although 60 children were eligible for the study, only the 34 children whose parents signed consent forms were included in the sample (18 in the treatment classrooms, 16 in the control classrooms). At baseline, children in the study averaged 4.7 years of age; 65% were male; 12% were Hispanic, 71% were Caucasian, and 6% were African-American; and none of the children were identified as having a disability.
Setting	The study took place in 12 Head Start classrooms in southern New Hampshire.
Intervention	*Ladders to Literacy* was implemented as a supplementary curriculum to the *Creative Curriculum*®. Teachers were trained to implement 18 language and literacy activities (of 50 that were available) across three domains (print/book awareness, metalinguistic awareness, and oral language). Fidelity of implementation was assessed twice during the study year: first in January/February 2003, and again in March/April 2003. For both treatment (*Ladders to Literacy* plus *Creative Curriculum*®) and control (*Creative Curriculum*® alone) classrooms, fidelity for the *Creative Curriculum*® was assessed using a checklist published by the *Creative Curriculum*® publishers. For the treatment group, technical assistance was available, if needed. A checklist for *Ladders to Literacy* was prepared by state Department of Education staff. Across both groups of classrooms, implementation of both curricula was low to moderate, averaging near 50% of the scheduled activities.
Comparison	Control classrooms implemented the *Creative Curriculum*®. *Creative Curriculum*® classrooms are designed to encourage children's choices and hands-on learning. Classrooms are set up with "interest areas," sections of the classrooms with different foci—for example, library area, house corner, and art center—which include relevant toys and objects. Children are encouraged to interact and play in the various interest areas (Russell, 2005).
Primary outcomes and measurement	For posttests, oral language is measured through analysis of samples of child speech and calculation of mean length of utterance (MLU) and type token ratio (TTR). For a more detailed description of these outcome measures, see Appendix A2.1.
Staff/teacher training	Both treatment and control teachers received at least one day of training in the *Creative Curriculum*®. Treatment group teachers received an additional two days of training on *Ladders to Literacy* activities in early fall 2002.

1. This study was conducted during the pilot year of the PCER Consortium (2008) study of *Ladders to Literacy*. Attrition at both the classroom and student levels was low enough to meet WWC standards for acceptable levels of bias under conservative assumptions.

Appendix A1.2 Study characteristics: PCER Consortium, 2008

Characteristic	Description
Study citation	Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). *Creative Curriculum* with *Ladders to Literacy*: University of New Hampshire. In *Effects of preschool curriculum programs on school readiness* (pp. 65–73). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
Participants[1]	This was a pre-, post-, follow-up study of the effectiveness of the *Ladders to Literacy* curriculum conducted during the 2003–04 and 2004–05 school years. In 2002–03 (the study's pilot year), the researchers recruited 12 Head Start classrooms to participate in the study, blocked them in various ways, and randomly assigned them to treatment and control groups from the blocks (described above in Russell, 2005). In the study's evaluation year (2003–04), 11 of the pilot-year classrooms and nine of the teachers were retained. One control classroom was replaced with another classroom from the same center, and two additional classrooms were randomly assigned to the treatment and control groups. This resulted in a sample of 14 classrooms (seven treatment and seven control). For most of the classrooms, the treatment condition had been in place for a full year when the evaluation year started, and thus parents of children in the new cohort had the opportunity to learn about the treatments and select the one they liked best. After parental consent was obtained, the sample included 123 children at baseline; 105 children were included in the final sample (54 treatment, 51 control). Baseline equivalence between the treatment and control children was established. At baseline, children in the study averaged 4.6 years of age; 44% were male; 31% were Hispanic, 39% were Caucasian, and 11% were African-American.
Setting	The study was conducted in Head Start classrooms in New Hampshire.
Intervention	*Ladders to Literacy* was implemented as a supplementary curriculum to the *Creative Curriculum*®. In this study, the researchers selected 27 of the more than 50 *Ladders to Literacy* activities to be used in all study classrooms. Teachers were trained to implement 27 language and literacy activities that covered three domains (print/book awareness, metalinguistic awareness, and oral language). Teachers were expected to implement nine activities (three from each of the three major domains) in the months of November and December 2003. Teachers were to add three to six additional activities on a monthly basis from January to May 2004. Researchers used a global fidelity measure to rate the overall fidelity with which the curricula were implemented. On a four-point scale (0 = Not at All to 3 = High), the *Ladders to Literacy* curriculum was rated in the high-medium range (2.71), whereas the control group curriculum was rated at the medium level (2.0).
Comparison	The control group implemented the *Creative Curriculum*® without *Ladders to Literacy*. The *Creative Curriculum*® is a comprehensive curriculum for 3- to 5-year-old children. It addresses four areas of development: social/emotional, physical, cognitive, and language development. *Creative Curriculum*® requires the physical space of the classroom to be structured into 10 interest areas: blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, and computers. Time is also allotted for outdoor activities. The 10 interest areas are designed to address curriculum content such as literacy, math, science, social studies, the arts, and technology, as well as process skills such as observing, exploring, and problem solving. *Creative Curriculum*® includes a developmental checklist that teachers are asked to use in ongoing assessments of child progress.
Primary outcomes and measurement	The outcome domains of oral language, print knowledge, phonological processing, and math were assessed with standardized measures. Oral language was assessed with the Peabody Picture Vocabulary Test–Third Edition (PPVT-III) and the Grammatic Understanding subtest from the Test of Oral Language Development–Primary III (TOLD-P:3). Print knowledge was assessed with the Test of Early Reading Ability–Third Edition (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).[2] 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 the fall of the preschool year, and posttesting was done in the spring of the preschool year. Trained research staff administered all assessments, which were conducted in English. For a more detailed description of these outcome measures, see Appendices A2.1–A2.4.
Staff/teacher training	All teachers (both treatment and control) received at least one day of training in the *Creative Curriculum*® from a staff member at Teaching Strategies, Inc. Treatment group teachers received *Ladders to Literacy* training in September 2003 and ongoing training on a monthly basis throughout the 2003–04 school year.

1. Attrition at both the classroom and student levels was low enough to meet WWC standards for acceptable levels of bias under conservative assumptions.
2. The Comprehensive Test of Phonological Processing (CTOPP) was used in the kindergarten follow-up (reported in Appendix A4.3).

Appendix A2.1 Outcome measures for the oral language domain

Outcome measure	Description
Mean length of utterance (MLU)	All utterances (at least 50 per child) were recorded for each child in the treatment and control groups. This required 15–30 minutes of recording per child. Utterances were transcribed verbatim by staff who were blind to the treatment status of the child. The median 50 utterances were selected from the resulting transcriptions and used to compute MLU. For a given child, MLU is calculated as [number of morphemes]/[number of utterances] based on the entire sample of the child's speech (as described in Russell, 2005).
Type token ratio (TTR)	All utterances (at least 50 per child) were recorded for each child in the treatment and control groups. This required 15–30 minutes of recording per child. Utterances were transcribed verbatim by staff who were blind to the treatment status of the child. The median 50 utterances were selected from the resulting transcriptions and used to compute TTR. For a given child, TTR is calculated as [number of different words in the sample]/[total number of words in the sample] based on the entire sample of the child's speech (as described in Russell, 2005).
Peabody Picture Vocabulary Test–Third Edition (PPVT-III)	A standardized measure of children's receptive vocabulary for which children show understanding of a spoken word by pointing to a picture that best represents the meaning (as cited in PCER Consortium, 2008).
Test of Language Development–Primary III (TOLD-P:3) Grammatic Understanding subtest	A standardized measure of children's ability to comprehend the meaning of sentences by selecting pictures that most accurately represent the sentence (as cited in PCER Consortium, 2008).

Appendix A2.2 Outcome measures for the print knowledge domain

Outcome measure	Description
Test of Early Reading Ability–III (TERA-3) Total Score	A standardized measure of children's developing reading skills with three subtests: Alphabet, Conventions, and Meaning (as cited in PCER Consortium, 2008).[1]
Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest	A standardized measure of identification of letters and reading of words (as cited in PCER Consortium, 2008).
WJ-III Spelling subtest	A standardized measure that assesses children's prewriting skills, such as drawing lines, tracing, and writing letters (as cited in PCER Consortium, 2008).

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

Appendix A2.3 Outcome measures for the phonological processing domain

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

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

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Russell, 2005[8]								
MLU	Preschoolers	12/34	3.36 (0.56)	3.45 (0.80)	–0.09	–0.13	ns	–5
TTR	Preschoolers	12/34	0.52 (0.49)	0.50 (0.06)	0.02	0.04	ns	+2
Average for oral language (Russell, 2005)[9]						**–0.04**	**ns**	**–1**
PCER Consortium, 2008[8]								
PPVT-III	Preschoolers	14/105	88.24 (18.03)	95.43 (14.88)	–7.19	–0.38	ns	–15
TOLD-P:3 Grammatic Understanding subtest	Preschoolers	14/104	8.38 (2.87)	9.45 (2.61)	–1.07	–0.22	ns	–9
Average for oral language (PCER Consortium, 2008)[9]						**–0.30**	**ns**	**–12**
Domain average for oral language across all studies[9]						**–0.17**	**na**	**–7**

ns = not statistically significant
na = not applicable
MLU = Mean length of utterance
TTR = Type token ratio
PPVT-III = Peabody Picture Vocabulary Test–III
TOLD-P:3 = Test of Oral Language Development–Primary III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. 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.

(continued)

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

5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. For Russell (2005), the effect size is based on WWC calculations, which use the pooled standard deviation and may differ from those reported by the study authors. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For 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 Russell (2005), a correction for clustering 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 using hierarchical linear modeling (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[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008[8]								
TERA-3	Preschoolers	14/105	82.90 (14.66)	85.27 (14.66)	–2.37	–0.30	ns	–12
WJ-III Letter-Word Identification subtest	Preschoolers	14/105	93.81 (13.20)	97.90 (13.56)	–4.09	–0.16	ns	–6
WJ-III Spelling subtest	Preschoolers	14/105	97.31 (12.13)	89.96 (15.12)	7.35	0.30	ns	+12
Domain average for print knowledge[9]						**–0.05**	**na**	**–2**

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

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. 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 (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For 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 using HLM, and no impacts were statistically significant.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008[8]								
Pre-CTOPPP Elision subtest	Preschoolers	14/105	8.55 (4.60)	9.10 (5.09)	–0.55	–0.16	ns	–6
Domain average for phonological processing[9]						**–0.16**	**na**	**–6**

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 phonological processing domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. 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 (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For 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 were needed because the analysis corrected for clustering using HLM.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008[8]								
WJ-III Applied Problems subtest	Preschoolers	14/105	93.09 (14.25)	96.10 (15.39)	–3.01	–0.14	ns	–6
CMA-A Composite Score	Preschoolers	14/105	0.60 (0.23)	0.56 (0.25)	0.04	0.18	ns	+7
Building Blocks Shape Composition	Preschoolers	14/104	1.77 (1.00)	1.75 (0.98)	0.02	0.02	ns	+1
Domain average for math[9]						**0.02**	**na**	**+1**

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 A4.4.
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. 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 (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For 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 using HLM, and no impacts were statistically significant.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of follow-up findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008[8]								
PPVT-III	Kindergarteners	14/80	94.59 (13.33)	100.23 (9.24)	–5.64	–0.30	ns	–15
TOLD-P:3 Grammatic Understanding subtest	Kindergarteners	14/81	9.45 (2.58)	9.74 (1.93)	–0.29	–0.06	ns	–2

ns = not statistically significant

PPVT-III = Peabody Picture Vocabulary Test–III

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

1. This appendix presents follow-up findings for measures that fall in the oral language domain. Outcomes from the preschool year 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 (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting 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. In the case of PCER Consortium (2008), no corrections for multiple comparisons or clustering were needed because the analysis corrected for clustering using HLM, and no impacts were statistically significant.

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008[8]								
TERA-3	Kindergarteners	14/81	87.80 (16.99)	92.03 (12.48)	–4.23	–0.54	ns	–21
WJ-III Letter-Word Identification subtest	Kindergarteners	14/81	93.81 (12.75)	100.60 (13.92)	–6.79	–0.27	ns	–11
WJ-III Spelling subtest	Kindergarteners	14/81	96.40 (16.18)	98.29 (12.92)	–1.89	–0.08	ns	–3

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. Outcomes from the preschool year 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 (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting 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. In the case of PCER Consortium (2008), no corrections for multiple comparisons or clustering were needed because the analysis corrected for clustering using HLM, and no impacts were statistically significant.

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008[8]								
CTOPP Elision subtest	Kindergarteners	14/81	4.19 (3.66)	4.60 (4.55)	–0.41	–0.10	ns	–4

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. Outcomes from the preschool year 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 (Cohen's *d* based on ANCOVA).
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. In the case of PCER Consortium (2008), no correction for clustering was needed because the analysis corrected for clustering using HLM.

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

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] *Ladders to Literacy*[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008[8]								
WJ-III Applied Problems subtest	Kindergarteners	14/80	94.26 (14.70)	101.26 (9.12)	–7.00	–0.33	ns	–13
CMA-A Composite	Kindergarteners	14/81	0.66 (0.20)	0.71 (0.13)	–0.05	–0.19	ns	–8
Building Blocks Shape Composition	Kindergarteners	14/81	2.54 (0.78)	2.63 (0.69)	–0.09	–0.10	ns	–4

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. Outcomes from the preschool year 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. 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 (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting 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. In the case of PCER Consortium (2008), no corrections for multiple comparisons or clustering were needed because the analysis corrected for clustering using HLM, and no impacts were statistically significant.

Appendix A5.1 *Ladders to Literacy* 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.[1]

For the outcome domain of oral language, the WWC rated *Ladders to Literacy* as having potentially negative effects for preschool children. The remaining rating (negative effects) was not considered, as *Ladders to Literacy* was assigned the highest applicable rating.

Rating received

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.

  Met. One of the two studies that measured oral language showed a substantively important negative effect.

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. Neither of the two studies that measured oral language showed statistically significant or substantively important positive effects.

Other ratings considered

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

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

  Not met. Neither of the two studies that measured oral language showed statistically significant or substantively important positive effects.

AND

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

  Not met. One of the two studies that measured oral language showed a substantively important negative effect.

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

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

  Not met. Neither of the two studies that measured oral language showed 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. One of the two studies that measured oral language showed a substantively important negative effect.

(continued)

Appendix A5.1 *Ladders to Literacy* rating for the oral language domain *(continued)*

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

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

  Not met. Neither of the two studies that measured oral language showed statistically significant or substantively important positive effects.

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. One of the two studies that measured oral language showed a substantively important effect and one study showed an indeterminate effect.

No discernible effects: No affirmative evidence of effects.

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

  Not met. One of the two studies that measured oral language showed a substantively important negative effect.

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

Appendix A5.2 *Ladders to Literacy* 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.[1]

For the outcome domain of print knowledge, the WWC rated *Ladders to Literacy* 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 print knowledge showed no statistically significant or substantively important effects.

Other ratings considered

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

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

  Not met. The one study that measured print knowledge showed no statistically significant or substantively important effects.

AND

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

  Not met. The one study that measured print knowledge showed no statistically significant or substantively important effects.

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

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

  Not met. The one study that measured print knowledge showed no statistically significant or substantively important 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 print knowledge showed no statistically significant or substantively important effects. No study showed positive effects.

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 print knowledge showed no statistically significant or substantively important effects.

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 print knowledge showed no statistically significant or substantively important effects.

(continued)

Appendix A5.2 *Ladders to Literacy* rating for the print knowledge 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 print knowledge showed no statistically significant or substantively important effects.

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.

  Not met. The one study that measured print knowledge showed no statistically significant or substantively important effects.

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

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

  Not met. The one study that measured print knowledge showed no statistically significant or substantively important effects.

AND

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

  Met. The one study that measured print knowledge showed no statistically significant or substantively important 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 *Ladders to Literacy* 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.[1]

For the outcome domain of phonological processing, the WWC rated *Ladders to Literacy* 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 phonological processing showed no statistically significant or substantively important effects.

Other ratings considered

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

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

  Not met. The one study that measured phonological processing showed no statistically significant or substantively important effects.

AND

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

  Not met. The one study that measured phonological processing showed no statistically significant or substantively important effects.

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

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

  Not met. The one study that measured phonological processing showed no statistically significant or substantively important 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 phonological processing showed no statistically significant or substantively important effects. No study showed positive effects.

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 phonological processing showed no statistically significant or substantively important effects.

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 phonological processing showed no statistically significant or substantively important effects.

(continued)

Appendix A5.3 *Ladders to Literacy* rating for the phonological processing 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 phonological processing showed no statistically significant or substantively important effects.

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.

  Not met. The one study that measured phonological processing showed no statistically significant or substantively important effects.

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

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

  Not met. The one study that measured phonological processing showed no statistically significant or substantively important effects.

AND

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

  Met. The one study that measured phonological processing showed no statistically significant or substantively important 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 *Ladders to Literacy* 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.[1]

For the outcome domain of math, the WWC rated *Ladders to Literacy* 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 effects.

Other ratings considered

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

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

  Not met. The one study that measured math showed no statistically significant or substantively important effects.

AND

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

  Not met. The one study that measured math showed no statistically significant or substantively important effects.

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

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

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

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 effects.

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 effects.

(continued)

Appendix A5.4 *Ladders to Literacy* 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 effects.

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.

  Not met. The one study that measured math showed no statistically significant or substantively important effects.

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

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

  Not met. The one study that measured math showed no statistically significant or substantively important effects.

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 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 A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Preschool classrooms	Children	
Oral language	2	26[2]	139	Medium to large
Print knowledge	1	14	105	Small
Phonological processing	1	14	105	Small
Early reading/writing	na	na	na	na
Cognition	na	na	na	na
Math	1	14	105	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.
2. As noted in Appendices A1.1 and A1.2, some of the 12 classrooms in the Russell (2005) study are also included in the 14 classrooms in the PCER Consortium (2008) study, but the two studies are based on different cohorts of children.