What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Early Childhood Education June 2009[1]

Doors to Discovery™

Program Description[2]

Doors to Discovery™, an early childhood curriculum, focuses on the development of children's vocabulary and expressive and receptive language through a learning process called "shared literacy," where adults and children work together to develop literacy-related skills. Literacy activities, organized into thematic units, encourage children's development in a number of areas identified by research as the foundation for early literacy success: oral language, phonological awareness, concepts of print, alphabet knowledge, writing, and comprehension. Each unit is available as a kit that includes various teacher resources.

Research[3]

One study of *Doors to Discovery*™ meets What Works Clearinghouse (WWC) evidence standards, and one study meets WWC evidence standards with reservations. The two studies included 33 preschool classrooms and 220 prekindergarten children from three to five years of age in two locations in the southwest United States.[4]

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

Effectiveness

Doors to Discovery™ was found to have potentially positive effects on oral language and print knowledge, and no discernible effects on phonological processing and math.

	Oral language	*Print knowledge*	*Phonological processing*	*Early reading and writing*	*Cognition*	*Math*
Rating of effectiveness	Potentially positive effects	Potentially positive effects	No discernible effects	na	na	No discernible effects
Improvement index[5]	Average: +9 percentile points	Average: +16 percentile points	Average: +7 percentile points	na	na	Average: 0 percentile points
	Range: +6 to +12 percentile points	Range: +2 to +37 percentile points	na	na	na	Range: –5 to +5 percentile points

na = not applicable

1. This report has been updated to include reviews of three studies that were released since 2007, a review of one study that was released in 2005 but was not reviewed for the previous report, and a re-review of two studies that were included in the previous report. The findings described in the previous *Doors to Discovery*™ intervention report were based on a study by Assel et al. (2007). A re-review of that study for the present report revealed that the subcluster attrition rate of children exceeded standards, as specified in the Early Childhood Education protocol. Hence, results from the Assel et al. (2007) study were not considered when preparing the present intervention report.
2. The descriptive information for this program was obtained from publicly available sources: the program's website (https://www.wrightgroup.com/family.html?PHPSESSID=ae71226df93c0a0211ac7a57f5d22c66&gid=183&longCopy=Y, downloaded November 5, 2008) and the research literature (Assel et al., 2007; PCER Consortium, 2008). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
3. The studies in this report were reviewed using WWC Evidence Standards, Version 1.0 (see the WWC Standards).
4. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
5. These numbers show the average and range of student-level improvement indices for all findings across the studies.

Absence of conflict of interest

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

Additional program information

Developer and contact

Doors to Discovery™ was developed and is distributed by Wright Group/McGraw-Hill. Address: 220 East Danieldale Road, DeSoto, TX 75115. Web: www.wrightgroup.com, Telephone: (800) 648-2970. Fax: (800) 593-4418.

Scope of use

According to the developer, the curriculum is used in various early childhood settings, including Head Start, private child care, public schools, and Early Reading First Centers of Excellence. Information is not available on the number or demographics of children or centers using this program.

Teaching

Doors to Discovery™, an early childhood curriculum, uses thematic units of literacy activities to encourage children's development in a number of areas identified by research as the foundation for early literacy success: oral language, phonological awareness, concepts of print, alphabet knowledge, writing, and comprehension. *Doors to Discovery*™ includes eight thematic units: Backyard Detectives; Build it Big!; Discovery Street; Healthy Me!; New Places, New Faces; Our Water Wonderland; Tabby Tiger's Diner; and Vroom! Vroom!. Each unit is available as a kit that includes various teacher resources. Children are taught using specific teacher techniques (such as cloze techniques, student retelling, think aloud activities, and scaffolding to build oral language skills) within literacy-enriched learning centers. Family literacy activities are available to encourage partnerships between the school and the home. The focus of the curriculum is the development of children's vocabulary and expressive and receptive language through a learning process called "shared literacy" (where adults and children work together to develop literacy related skills). Teachers are trained during professional development activities and with other resources like the Discovery Guide (a built-in professional development resource).

Cost

The complete *Doors to Discovery*™ set is available to education professionals for $2,348.40. Alternatively, each theme kit can be purchased separately for $327.45. Teacher resources, such as alphabet posters and an assessment handbook, are also available for purchase. Additional pricing information for other materials (e.g., teacher resources and children's books) is available on the website. The prices listed on the website are for education professionals only. Information about the cost of professional development is not available.

Research

Six studies reviewed by the WWC investigated the effects of *Doors to Discovery*™. One study (PCER Consortium, 2008) is a randomized controlled trial that meets WWC evidence standards. One study (Christie, Roskos, Vukelich, & Han, 2003) is a randomized controlled trial that meets WWC evidence standards with reservations. The remaining four studies do not meet either WWC evidence standards or eligibility screens.

Meets evidence standards

One study reviewed by the WWC (PCER Consortium, 2008) assessed the effectiveness of *Doors to Discovery*™ as part of the Preschool Curriculum Evaluation Research (PCER) effort.[6] The PCER Consortium (2008) used a randomized controlled trial design in which 29 full day Head Start and public prekindergarten preschool classrooms in Texas were randomly assigned either

6. The PCER Consortium (2008) evaluated a total of 14 preschool curricula, including *Doors to Discovery*™, in comparison to respective control conditions.

Research *(continued)*

to implement *Doors to Discovery*™ or to a control group.[7] Data were collected on 183 children (94 *Doors to Discovery*™ and 89 control). Pretest data were collected in the fall, and posttest data were collected in the spring, of the preschool year. The study investigated effects on oral language, print knowledge, phonological processing, and math. The control condition varied across sites and included teacher-developed, nonspecific curricula.

Meets evidence standards with reservations

One study (Christie et al., 2003) was a randomized controlled trial with severe subcluster attrition and baseline equivalence of the analytic sample. In this study, four Head Start classrooms in a large metropolitan area in the southwest United States were randomly assigned[8] to implement either *Doors to Discovery*™ or the control group, which used materials based on *Creative Curriculum*®.[9] Data were collected on 37 children (21 *Doors to Discovery*™ and 16 control group). Pretest data were collected during November and December of the preschool year; the *Doors to Discovery*™ curriculum was implemented from January through early April, and posttest data were collected in late April and May. The study investigated effects on oral language and print knowledge.

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.[10]

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

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 and writing, cognition, and math. The studies included in this report cover four domains: oral language, print knowledge, phonological processing, and math. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Doors to Discovery*™ on students.[11]

Oral Language. The PCER Consortium (2008) analyzed the effectiveness of *Doors to Discovery*™ on oral language using the Peabody Picture Vocabulary Test-III (PPVT-III) and the Test of

7. The study indicated, and the authors' confirmed, that the unit of assignment was the classroom; however, all classrooms within a school were assigned to the same treatment condition.
8. A fifth classroom participated in the study and implemented the *Doors to Discovery*™ curriculum. Since this classroom was not randomly assigned, it was omitted from the WWC review.
9. According to Christie et al. (2003), the comparison group was "loosely based" on *Creative Curriculum*®, a curriculum designed to foster children's social-emotional, physical, cognitive, and language development, relying heavily on the use of play centers (Han et al. 2005).
10. 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 *Doors to Discovery*™ is in Appendix A6.
11. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. No correction for clustering was needed for the study by the PCER Consortium (2008) because their analysis corrected for clustering by using hierarchical linear modeling (HLM), but a correction for multiple comparisons was needed, so the significance levels in this report may differ from those reported in the original study. For the study by Christie et al. (2003), the WWC excluded the one non-randomly assigned classroom and corrected for clustering, so the significance levels in this report may differ from those reported in the original study.

Effectiveness *(continued)*

Language Development-Primary III (TOLD-P:3) Grammatic Understanding subtest. The authors report, and the WWC confirms, that differences between the *Doors to Discovery*™ group and the control group are not statistically significant or substantively important on any of these measures. According to WWC criteria, this study shows no discernible effects on oral language.

Christie et al. (2003) analyzed the effectiveness of *Doors to Discovery*™ on oral language using the PPVT-III. WWC analyses of the Christie et al. (2003) data show a substantively important, but not statistically significant, positive effect of 0.30 when the *Doors to Discovery*™ group was compared to the control group.[12]

Print Knowledge. The PCER Consortium (2008) analyzed the effectiveness of *Doors to Discovery*™ 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, and the WWC confirms, that differences between the *Doors to Discovery*™ group and the control group are not statistically significant or substantively important on any of these measures. According to WWC criteria, this study shows no discernible effects on print knowledge.

Christie et al. (2003) analyzed the effectiveness of *Doors to Discovery*™ on print knowledge using Get Ready to Read! and the Concepts of Print. WWC analyses of the Christie et al. (2003) data show a substantively important, but not statistically significant, positive effect of 0.74 when *Doors to Discovery*™ was compared to the control group.

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

Math. The PCER Consortium (2008) analyzed the effectiveness of *Doors to Discovery*™ on math using the WJ-III Applied Problems subtest, the Child Math Assessment-Abbreviated (CMA-A), and the Shape Composition task. The authors report, and the WWC confirms, that differences between the *Doors to Discovery*™ group and the control group are not statistically significant or substantively important on any of these measures. According to WWC criteria, this study shows no discernible effects on math.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, 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 *Doors to Discovery*™ to have potentially positive effects on oral language and print knowledge, and no discernible effects on phonological processing and math

Improvement index

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

12. Christie et al. (2003) report a statistically significant difference for the PPVT-III, but the results are based on a sample of five classrooms. As noted in Appendix A1.2, one of these classrooms was not randomly assigned, and thus excluded from the review.

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

Based on two studies, the average improvement index for *Doors to Discovery*™ on two measures of oral language is +9 percentile points, with a range of +6 to +12 percentile points across findings, and the average improvement index on five measures of print knowledge is +16 percentile points, with a range of +2 to +37 percentile points. Based on one study, the average improvement index for *Doors to Discovery*™ on one measure of phonological processing is +7 percentile points, and the average improvement index on three measures of math is 0 percentile points, with a range of –5 to +5 percentile points.

Summary

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

References

Meets WWC evidence standards

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

Meets WWC evidence standards with reservations

Christie, J., Roskos, K., Vukelich, C., & Han, M. (2003, June). The effects of a well-designed literacy program on young children's language and literacy development. In F. Lamb-Parker, J. Hagen, R. Robinson, & H. Rhee (Eds.), *The first eight years—pathways to the future: Implications for research, policy, and practice* (pp. 447–448). Proceedings of the Head Start National Research Conference. New York: Mailman School of Public Health, Columbia University.

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

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

Han, M., Roskos, K., Christie, J., Mandzuk, S., & Vukelich, C. (2005). Learning words: Large group time as a vocabulary development opportunity. *Journal of Research in Childhood Education, 19*(4). The study does not meet WWC evidence standards because it only includes outcomes that are overaligned with the intervention or measured in a way that is inconsistent with the protocol.[13]

Roskos, K., Ergul, C., Bryan, T., Burstein, K., Christie, J., & Han, M. (2008). Who's learning what words and how fast? Preschoolers' vocabulary growth in an early literacy program. *Journal of Research in Childhood Education, 22*(3), 275–290. The study is ineligible for review because it does not use a comparison group.

Wahlstrom, K., Hornbacher, J., & Rader, S. (2007). *Bloomington/Richfield—Early Reading First Get Ready Centers of Excellence year II report.* Minneapolis, MN: Center for Applied Research and Educational Improvement. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

13. This source describes two studies; one that is also described in Christie et al. (2003) that meets standards with reservations, and another that does not meet evidence standards for the reason described above.

Appendix

Appendix A1.1 Study characteristics: PCER Consortium, 2008 (randomized controlled trial)

Characteristic	Description
Study citation	Preschool Curriculum Evaluation Research (PCER) Consortium, (2008). *Doors to Discovery* and *Let's Begin with the Letter People.* In *Effects of preschool curriculum projects on school readiness* (pp. 85–98). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
Participants	The study, conducted during the 2003–2004 and 2004–2005 school years, included three groups: *Doors to Discovery*™, *Let's Begin with the Letter People*®, and a control group. Nineteen full-day Head Start and public prekindergarten preschools were recruited for the study. From these 19 preschools, 95 teachers/classrooms were recruited, of which 76 were included in random assignment. The manuscript notes, and the authors confirmed, that the researchers randomly assigned the classrooms to three conditions (*Doors to Discovery*™, *Let's Begin with the Letter People*®, and control); however, all classrooms within a preschool were assigned to the same condition. The resulting sample of teachers/classrooms included 25 *Doors to Discovery*™ classrooms, 24 *Let's Begin with the Letter People*® classrooms, and 27 control classrooms. Forty-five of the 76 classrooms were then randomly selected to participate in the PCER study. One of the 45 classrooms dropped out, leaving 14 *Doors to Discovery*™ classrooms, 15 *Let's Begin with the Letter People*® classrooms, and 15 control classrooms. Seven children whose parents had provided consent to participate in the study were randomly selected from each classroom, for a total of 308 children.[1] The parental consent rate was 65% for the treatment group (combined *Doors to Discovery*™ and *Let's Begin with the Letter People*®) and 55% for the control group. The total number of participating children in the study at baseline was 297 (101 *Doors to Discovery*™, 100 *Let's Begin with the Letter People*®, and 96 control). At baseline, children in the study averaged 4.6 years of age; 55% were male; and 43% were Hispanic, 30% were Caucasian, and 13% were African-American. The analysis sample for the *Doors to Discovery*™ study included 183 children (94 *Doors to Discovery*™ and 89 control). Depending on the outcome, child-level attrition ranged from 7% to 10%.
Setting	The *Doors to Discovery*™ study was conducted with children from 29 full-day preschool classrooms (14 *Doors to Discovery*™ and 15 control) selected from Head Start and public prekindergarten programs in Texas.
Intervention	*Doors to Discovery*™ is a prekindergarten curriculum that promotes learning in five areas associated with early literacy success: oral language, phonological awareness, concepts of print, alphabet knowledge and writing, and comprehension. Eight thematic units cover topics such as nature, friendship, communities, society, and health. Activities include teacher-directed, large- and small-group, and independent practice through activities tied to the curriculum. Family learning activities are also available. In the PCER study, each classroom's fidelity to the curriculum was rated on a four-point scale, ranging from "not at all" (0) to "high" (3). The average score for the *Doors to Discovery*™ classrooms was 2.13 on the measure.
Comparison	Control teachers used teacher-developed nonspecific curricula. Their classrooms were rated with the same fidelity measure used in the *Doors to Discovery*™ classrooms, which ranged from 0 to 3. The average score for the control classrooms was 1.0.
Primary outcomes and measurement	The outcome domains assessed were children's oral language, print knowledge, phonological processing, and math. Oral language was assessed with the Peabody Picture Vocabulary Test-III (PPVT-III) and the Test of Language Development-Primary III (TOLD-P:3) Grammatic Understanding subtest. Print knowledge was assessed with the Test of Early Reading Ability-III (TERA-3), the Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest, and the WJ-III Spelling subtest. Phonological processing was assessed with the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest. Math was assessed with the WJ-III Applied Problems subtest, the Child Math Assessment-Abbreviated (CMA-A), and the Shape Composition task. For a more detailed description of these outcome measures, see Appendices A2.1–2.4.
Staff/teacher training	Teachers received curriculum training prior to the start of the 2003–2004 school year. This was the second year of implementation of the treatment, and most of the teachers had been trained prior to the start of the 2002–2003 school year. New teachers each received 12 hours of training, and returning teachers each received six hours of training. The research team collected site-specific curriculum fidelity data three times during the preschool year. All classrooms were observed using the Teacher Behavior Rating Scale in fall and spring of the preschool year.

1. PCER Consortium (2008, p. 88) reported that eight children were selected from each classroom. In response to a query, the study authors noted that eight children were randomly selected for the site-specific study; however, only seven children were randomly selected for the PCER Consortium study.

Appendix A1.2 Study characteristics: Christie et al., 2003 (randomized controlled trial)

Characteristic	Description
Study citation	Christie, J., Roskos, K., Vukelich, C., & Han, M. (2003, June). The effects of a well-designed literacy program on young children's language and literacy development. In F. Lamb-Parker, J. Hagen, R. Robinson, & H. Rhee. (Eds.), *The first eight years—pathways to the future: Implications for research, policy, and practice* (pp. 447–448). Proceedings of the Head Start National Research Conference. New York: Mailman School of Public Health, Columbia University.
Participants	In this study, four Head Start classrooms—two serving English-speaking children and two serving Spanish-speaking children—were blocked on primary language of the children and randomly assigned to implement either *Doors to Discovery*™ or the *Creative Curriculum*®. One additional classroom served a mixed-language group and was assigned to implement *Doors to Discovery*™. Since this classroom was not assigned at random, it was omitted from WWC analyses. At baseline, the four-classroom study included 35 children in the *Doors to Discovery*™ group and 28 children in the control group. The four-classroom analysis sample was substantially smaller, containing 21 children in the *Doors to Discovery*™ group and 16 children in the control group. This translates to a child-level attrition rate of 41%. Baseline differences between the treatment and control groups were large, but not statistically significant. For the analytic sample, the baseline difference was (in standard deviation units) 0.40 for the Peabody Picture Vocabulary Test (PPVT), 0.45 for Get Ready to Read!, and –0.29 for Concepts of Print.
Setting	The study was conducted with Head Start classrooms in a large metropolitan area in the southwest United States.
Intervention	Teachers in the intervention classrooms used three units from the *Doors to Discovery*™ curriculum: Vroom! Vroom!; Build It Big!; and Tabby Tiger's Diner. Each unit was taught for 4 weeks.
Comparison	The control classrooms used the existing curriculum, which the authors described as loosely based on the *Creative Curriculum*®.
Primary outcomes and measurement	The outcomes assessed were children's oral language and print knowledge. Oral language was assessed with the PPVT. Print knowledge was assessed with Get Ready to Read! and Concepts of Print. All assessments were conducted in English (J. Christie, personal communication, January 23, 2009). For a more detailed description of these outcome measures, see Appendices A2.1–2.2.
Staff/teacher training	No information on training was provided.

Appendix A2.1 Outcome measures for the oral language domain

Outcome measure	Description
Peabody Picture Vocabulary Test-3rd Edition (PPVT-III)	A standardized measure of children's receptive vocabulary where 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)	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 Letter-Word Identification subtest	A standardized measure of identification of letters and reading of words (as cited in PCER Consortium, 2008).
Woodcock-Johnson 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).
Concepts of Print	An eight-item measure of concepts of print, adapted from the Developing Skills Checklist, which assesses children's knowledge of book handling; the difference between print and pictures; the concepts of "letter", "word", and "number"; and several conventions of print, e.g., left-right sequence and capitalization (J. Christie, personal communication, January 23, 2009).
Get Ready to Read!	An early literacy screening tool that measures print recognition, concepts of print, book concepts, and phonemic awareness (J. Christie, personal communication, January 23, 2009).

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

Appendix A2.3 Outcome measures for the phonological processing domain

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

Appendix A2.4 Outcome measures for the math domain

Outcome measure	Description
Woodcock-Johnson III 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] — *Doors to Discovery*™ group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Doors to Discovery*™ – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
PPVT-III	Preschoolers	29/183	94.63 (18.20)	91.33 (18.12)	3.30	0.15	ns	+6
TOLD-P:3 Grammatic Understanding subtest	Preschoolers	29/183	10.19 (3.06)	9.33 (2.71)	0.86	0.17	ns	+7
Average for oral language (PCER Consortium, 2008)[9]						**0.16**	**na**	**+6**
Christie et al., 2003 (randomized controlled trial)[8]								
PPVT-III	Preschoolers	4/37	35.98 (19.32)	30.25 (17.09)	5.73	0.30	ns	+12
Average for oral language (Christie et al., 2003)[9]						**0.30**	**na**	**+12**
Domain average for oral language across all studies[9]						**0.23**	**na**	**+9**

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

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. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted. For Christie et al. (2003), the treatment group means are the sum of the control group means and the mean difference, which is adjusted for pretest. The standard deviations were pooled across classrooms.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted. For the study by Christie et al. (2003), the WWC excluded one non-randomly assigned classroom, so the means, standard deviations, effect sizes, and significance levels in this report may differ from those reported in the original study.
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, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant. In the case of Christie et al. (2003), the WWC corrected for clustering.
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] — *Doors to Discovery*™ group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Doors to Discovery*™ – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
TERA-3	Preschoolers	29/182	93.4 (17.22)	92.76 (17.86)	0.64	0.06	ns	+2
WJ-III Letter-Word Identification subtest	Preschoolers	29/183	108.82 (14.56)	106.04 (13.82)	2.78	0.10	ns	+4
WJ-III Spelling subtest	Preschoolers	29/183	98.91 (12.56)	97.37 (12.63)	1.54	0.06	ns	+2
Average for print knowledge (PCER Consortium, 2008)[9]						**0.07**	**na**	**+3**
Christie et al., 2003 (randomized controlled trial)[8]								
Concepts of Print	Preschoolers	4/37	4.48 (1.51)	2.82 (1.49)	1.66	1.08	ns	+37
Get Ready to Read!	Preschoolers	4/37	8.62 (3.96)	7.06 (3.81)	1.56	0.39	ns	+16
Average for print knowledge (Christie et al., 2003)[9]						**0.74**	**na**	**+27**
Domain average for print knowledge across all studies[9]						**0.41**	**na**	**+16**

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. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted. For Christie et al. (2003), the treatment group means are the sum of the control group means and the mean difference, which is adjusted for pretest. The standard deviations were pooled across classrooms.

(continued)

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

4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted. For the study by Christie et al. (2003), the WWC excluded one non-randomly assigned classroom, so the means, standard deviations, effect sizes, and significance levels in this report may differ from those reported in the original study.
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, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections for clustering or multiple comparisons were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant. In the case of Christie et al. (2003), the WWC corrected for clustering.
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[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Doors to Discovery*™ group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Doors to Discovery*™ – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
Pre-CTOPPP Elision subtest	Preschoolers	29/182	10.78 (4.18)	10.11 (4.64)	0.67	0.18	ns	+7
Domain average for phonological processing (PCER Consortium, 2008)[9]						**0.18**	**na**	**+7**

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. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections 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 math domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Doors to Discovery*™ group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*Doors to Discovery*™ – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium, 2008 (randomized controlled trial)[8]								
WJ-III Applied Problems subtest	Preschoolers	29/183	99.53 (13.24)	99.28 (16.60)	0.25	0.01	ns	+0
CMA-A Composite	Preschoolers	29/183	0.68 (0.20)	0.65 (0.24)	0.03	0.13	ns	+5
Shape Composition	Preschoolers	29/183	1.61 (0.84)	1.72 (0.69)	–0.11	–0.13	ns	–5
Domain average for math (PCER Consortium, 2008)[9]						**0.00**	**na**	**+0**

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. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections 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 A4.1 Summary of follow-up findings for the oral language domain[1]

Outcome measure	Study sample	Sample size[3] (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Doors to Discovery*™ group[4]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[5] (*Doors to Discovery*™ – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
PCER Consortium, 2008 (randomized controlled trial)[9]								
PPVT-III	Kindergarteners	nr/152	98.13 (17.46)	94.00 (16.01)	4.13	0.18	ns	+7
TOLD-P:3 Grammatic Understanding subtest	Kindergarteners	nr/155	10.41 (3.19)	10.08 (2.80)	0.33	0.06	ns	+2

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

1. This appendix presents follow-up findings considered for measures that fall in the oral language domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The PCER Consortium (2008) study included 149 kindergarten classrooms across all three conditions in this study (*Doors to Discovery*™, control, and *Let's Begin with the Letter People*®). The number of classrooms for *Doors to Discovery*™ and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. For an explanation of the effect size calculation, see WWC Computations 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).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

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

Outcome measure	Study sample	Sample size[3] (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Doors to Discovery*™ group[4]	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[5] (*Doors to Discovery*™ – *comparison*)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
PCER Consortium, 2008 (randomized controlled trial)[9]								
TERA-3	Kindergarteners	nr/155	93.38 (18.88)	93.96 (16.47)	–0.58	–0.05	ns	–2
WJ-III Letter-Word Identification subtest	Kindergarteners	nr/155	106.99 (14.82)	109.53 (13.57)	–2.54	–0.09	ns	–4
WJ-III Spelling subtest	Kindergarteners	nr/155	100.51 (14.84)	103.46 (13.14)	–2.95	–0.12	ns	–5

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

1. This appendix presents follow-up findings for measures that fall in the print knowledge domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The PCER Consortium (2008) study included 149 kindergarten classrooms across all three conditions in this study (*Doors to Discovery*™, control, and *Let's Begin with the Letter People*®). The number of classrooms for *Doors to Discovery*™ and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. 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).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

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

Outcome measure	Study sample	Sample size[3] (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Doors to Discovery*™ group[4]	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[5] (*Doors to Discovery*™ - comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
PCER Consortium, 2008 (randomized controlled trial)[9]								
CTOPP Elision subtest	Kindergarteners	nr/155	4.68 (3.84)	5.04 (4.24)	–0.36	–0.09	ns	–4

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

1. This appendix presents follow-up findings for measures that fall in the phonological processing domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The PCER Consortium (2008) study included 149 kindergarten classrooms across all three conditions in this study (*Doors to Discovery*™, control, and *Let's Begin with the Letter People*®). The number of classrooms for *Doors to Discovery*™ and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. 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).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

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

Outcome measure	Study sample	Sample size[3] (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Doors to Discovery*™ group[4]	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[5] (*Doors to Discovery*™ – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
PCER Consortium, 2008 (randomized controlled trial)[9]								
WJ-III Applied Problems subtest	Kindergarteners	nr/155	101.84 (10.95)	102.40 (11.38)	–0.56	–0.02	ns	–1
CMA-A Composite	Kindergarteners	nr/155	0.68 (0.16)	0.72 (0.14)	–0.04	–0.16	ns	–6
Shape Composition	Kindergarteners	nr/155	2.40 (0.79)	2.51 (0.69)	–0.11	–0.12	ns	–5

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

1. This appendix presents follow-up findings for measures that fall in the math domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The PCER Consortium (2008) study included 149 kindergarten classrooms across all three conditions in this study (*Doors to Discovery*™, control, and *Let's Begin with the Letter People*®). The number of classrooms for *Doors to Discovery*™ and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. 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).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM, and no impacts were statistically significant.

Appendix A5.1 *Doors to Discovery*™ 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 *Doors to Discovery*™ as having potentially positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative, negative) were not considered, as *Doors to Discovery*™ was assigned the highest applicable rating.

Rating received

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

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

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

AND

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

  Met. Neither of the two studies that measured oral language showed a statistically significant or substantively important negative effect. One study showed a substantively important positive effect, and one study showed no effect.

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 a statistically significant positive effect.

AND

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

  Met. Neither of the two 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 *Doors to Discovery*™ 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 *Doors to Discovery*™ as having potentially positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative, negative) were not considered, as *Doors to Discovery*™ was assigned the highest applicable rating.

Rating received

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

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

  Met. One of the two studies that measured print knowledge showed a substantively important positive effect.

AND

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

  Met. Neither of the two studies that measured print knowledge showed a statistically significant or substantively important negative effect. One study showed a substantively important positive effect, and one study showed an effect that was not statistically significant or substantively important.

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 print knowledge showed a statistically significant positive effect.

AND

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

  Met. Neither of the two studies 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 *Doors to Discovery*™ 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 *Doors to Discovery*™ as having no discernible effects. The remaining ratings (potentially negative, negative) were not considered, as *Doors to Discovery*™ was assigned the highest applicable rating.

Rating received

No discernible effects: No affirmative evidence of effects.

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

  Met. The one study that measured phonological processing showed no statistically significant or substantively important effect.

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 effect.

AND

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

  Met. The one study that measured phonological processing 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 phonological processing showed no statistically significant or substantively important effect.

AND

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

  Not met. The one study that measured phonological processing showed no statistically significant or substantively important effect. No other studies measured phonological processing.

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

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

  Not met. The one study that measured phonological processing showed no statistically significant or substantively important effect. No other studies measured phonological processing.

(continued)

Appendix A5.3 *Doors to Discovery*™ rating for the phonological processing domain *(continued)*

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 effect. No other studies measured phonological processing.

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 *Doors to Discovery*™ 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 *Doors to Discovery*™ as having no discernible effects. The remaining ratings (potentially negative, negative) were not considered, as *Doors to Discovery*™ was assigned the highest applicable rating.

Rating received

No discernible effects: No affirmative evidence of effects.

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

  Met. The one study that measured math showed no statistically significant or substantively important effect.

Other ratings considered

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

- Criterion 1: Two or more studies showing statisticly 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 substantially important positive effect.

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 effect.

AND

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

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

OR

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

  Not met. The one study that measured math showed no statistically significant or substantively important effect. 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[1]
		Schools	Students	
Oral language	2	33	220	Medium to large
Print knowledge	2	33	220	Medium to large
Phonological processing	1	29	182	Small
Early reading and writing	0	na	na	na
Cognition	0	na	na	na
Math	1	29	183	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.