What Works Clearinghouse™

Doors to Discovery™

Program Description[1]

Doors to Discovery™ is a preschool literacy curriculum that uses eight thematic units of activities to help children build fundamental early literacy skills in oral language, phonological awareness, concepts of print, alphabet knowledge, writing, and comprehension. The eight thematic units cover topics such as nature, friendship, communities, society, and health. Each unit is available as a kit that includes various teacher resources.

Research[2]

The What Works Clearinghouse (WWC) identified three studies of *Doors to Discovery*™ that both fall within the scope of the Early Childhood Education topic area and meet WWC evidence standards.[3] One study meets standards without reservations and two studies meet WWC evidence standards with reservations. Together, these studies included 585 preschool children aged three to five years in three locations.

The WWC considers the extent of evidence for *Doors to Discovery*™ on the school readiness of preschool children to be medium to large for one outcome domain—print knowledge—and small for three outcome domains—oral language, phonological processing, and math. There were no studies that meet standards in two other domains, so we do not report on the effectiveness of *Doors to Discovery*™ for those domains in this intervention report. (See the Effectiveness Summary on p. 5 for a full list of all 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 for preschool children.

Report Contents

Table 1. Summary of findings[4]

Outcome domain	Rating of effectiveness	Improvement index (percentile points) Average	Improvement index (percentile points) Range	Number of studies	Number of children	Extent of evidence
Oral language	Potentially positive effects	+11	na	1	37	Small
Print knowledge	Potentially positive effects	+15	+3 to +34	2	402	Medium to large
Phonological processing	No discernible effects	+6	na	1	182	Small
Math	No discernible effects	0	–5 to +6	1	183	Small

na = not applicable

Program Information

Background

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

Program details

Doors to Discovery™ is a preschool curriculum that 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. The program 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 teaching techniques, such as cloze techniques (the teacher presents a short sentence or phrase that leaves out a key word for the children to say out loud), student retelling, think aloud activities, and scaffolding, to build oral language skills, all within literacy-enriched learning centers. Family literacy activities are available to encourage additional early literacy practice, as well as 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," by which 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 $2,654.25. Alternatively, each theme kit can be purchased separately for $371.37. 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 Summary

The WWC identified six studies that investigated the effects of *Doors to Discovery*™ on the school readiness of preschool children.

Table 2. Scope of reviewed research

Grade	PK
Delivery method	Whole class
Program type	Curriculum

The WWC reviewed four of those studies against group design evidence standards. One study (Christie, Roskos, Vukelich, & Han, 2003) is a randomized controlled trial that meets WWC evidence standards without reservations, and two studies (Assel, Landry, Swank, & Gunnewig, 2007, and Preschool Curriculum Evaluation Research [PCER] Consortium, 2008, Chapter 6) are randomized controlled trials or quasi-experimental designs that meet WWC evidence standards with reservations. Those three studies are summarized in this report. One study does not meet WWC evidence standards.

The remaining two studies do not meet WWC eligibility screens for review in this topic area. Citations for all six studies are in the References section, which begins on p. 8.

Summary of study meeting WWC evidence standards without reservations

Christie et al. (2003) conducted a randomized controlled trial in which four Head Start classrooms were randomly assigned to either an intervention group implementing *Doors to Discovery*™ or to a comparison group, which used materials based on *The Creative Curriculum*®.[5] The authors also placed a fifth classroom into the intervention group, but since they provided child outcome data at the classroom level, the review focused on the four randomly-assigned classrooms.[6] This study was conducted in a large metropolitan area in the southwest United States. Data were collected on 37 children (21 *Doors to Discovery*™ and 16 comparison). 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 measured effects on children's oral language and print knowledge.

Summary of studies meeting WWC evidence standards with reservations

Assel et al. (2007) conducted a randomized controlled trial with high attrition at the subcluster level. As part of the PCER Consortium (2008, Chapter 6) study, schools in Houston, Texas with Title I and non-Title I (universal) preschool classrooms and Head Start centers were randomly assigned (within each program type) either to the *Doors to Discovery*™ intervention group, the *Let's Begin with the Letter People*® intervention group, or a comparison group. In the second stage of random assignment, classrooms assigned to *Doors to Discovery*™ or *Let's Begin with the Letter People*® were randomly assigned to receive mentoring or not as part of implementation of the intervention, creating a total of four intervention groups and one comparison group. For Assel et al. (2007), data were collected in the first year of the study for 550 children (184 *Doors to Discovery*™, 182 *Let's Begin with the Letter People*®, and 184 comparison) in 79 classrooms (27 *Doors to Discovery*™, 25 *Let's Begin with the Letter People*®, and 27 comparison).[7] Pretest data were collected prior to the implementation of the curriculum (spring 2002), and posttest data were collected at the end of the school year (spring 2003). The authors examined effects on oral language, print knowledge, and phonological processing. This review focuses on the comparison between the *Doors to Discovery*™ intervention group—including both classrooms assigned to receive mentoring and classrooms not assigned to receive mentoring—and the comparison group on the print knowledge outcomes, which are the only findings that meet WWC evidence standards. The study demonstrated the baseline equivalence of the outcome measures in the print knowledge domain for the analytic sample of intervention and comparison group children at the end of the preschool year. The study did not demonstrate baseline equivalence of the outcomes in the phonological processing and oral language domains. The study also discusses differences in child outcomes between the

mentoring and non-mentoring groups, but since the estimated differences are not presented in the paper, they are not included in the supplemental analyses in this report.

The PCER Consortium (2008, Chapter 6) also assessed the effectiveness of *Doors to Discovery*™ as part of the second year of the PCER initiative (2003–04 school year). Study authors randomly selected a subset of 45 of the original 79 full-day Head Start and public preschool classrooms in Houston, Texas to participate in the PCER evaluation in the year following the Assel et al. (2007) study. One teacher (and her classroom) from the randomly selected subset chose not to participate, and the final sample included 44 classrooms. During the pilot year, school sites had been randomly assigned; each classroom within a school was assigned to the same condition.

Although the study used a randomized controlled trial design to assign schools to intervention or comparison conditions in the pilot year, the PCER Consortium (2008, Chapter 6) study analyzed data from the second year of implementation, when children who had been in the classrooms at random assignment had moved on to kindergarten and a new class of children had replaced them. Thus, the study had high attrition at the child level and, under WWC standards, must demonstrate baseline equivalence between the intervention and comparison group sample of children used in the analyses of outcomes.

The authors investigated effects on oral language, print knowledge, phonological processing, and math. Findings for the math and phonological processing domains meet WWC evidence standards with reservations. The WWC based its effectiveness ratings on findings from comparisons of 94 children who received *Doors to Discovery*™ and 89 comparison children who received a variety of curricula, including teacher-developed, nonspecific curricula (children who received the *Let's Begin with the Letter People*® curriculum were not included in the comparisons). Children in the sample were 4.6 years old on average; just over half were male (54.6%); 12% were reported to have a disability; 43% were Hispanic, 30% were Caucasian, and 13% were African American. The study demonstrated the baseline equivalence of the outcome measures in the math and phonological processing domains for the analytic sample of intervention and comparison group children at the end of the preschool year. The study did not demonstrate baseline equivalence of the intervention and comparison groups on outcomes in the print knowledge and oral language domains for the analytic sample of children at the end of the preschool year. The authors reported on the effects of *Doors to Discovery*™ in the spring of the preschool year and again at the end of kindergarten. The kindergarten findings are not reported here because information about the baseline equivalence of the intervention and comparison groups on outcome measures for the kindergarten sample was not provided in the report. The authors also reported findings on the Social Skills Rating Scale; however, these findings are not reported here because the current Early Childhood Education topic area protocol does not include sociobehavioral outcomes.

Effectiveness Summary

The WWC review of *Doors to Discovery*™ for the Early Childhood Education topic area includes child outcomes in six domains: oral language, print knowledge, phonological processing, early reading and writing, cognition, and math. The three studies of *Doors to Discovery*™ that meet WWC evidence standards reported findings in four of the six domains: (a) oral language, (b) print knowledge, (c) phonological processing, and (d) math. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Doors to Discovery*™ on preschool children. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 20.

Summary of effectiveness for the oral language domain

One study that meets WWC standards without reservations reported findings in the oral language domain.

Christie et al. (2003) analyzed the effectiveness of *Doors to Discovery*™ on oral language outcomes using the Peabody Picture Vocabulary Test–Third Edition (PPVT-III). WWC analyses of the Christie et al. (2003) data show that the effect for the PPVT-III is not statistically significant, but large enough (0.27) to be considered substantively important according to WWC criteria (that is, at least 0.25). The WWC characterizes this study finding as a substantively important positive effect.

Thus, for the oral language domain, one study showed substantively important positive effects. This results in a rating of potentially positive effects, with a small extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the oral language domain

Rating of effectiveness	Criteria met
Potentially positive effects *Evidence of a positive effect with no overriding contrary evidence.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *oral language* domain showed substantively important positive effects.
Extent of evidence	**Criteria met**
Small	One study that included 37 children in four classrooms reported evidence of effectiveness in the *oral language* domain.

Summary of effectiveness for the print knowledge domain

One study that meets WWC standards without reservations and one study that meets WWC standards with reservations reported findings in the print knowledge domain.

Christie et al. (2003) analyzed the effectiveness of *Doors to Discovery*™ on print knowledge using Get Ready to Read! and the Developing Skills Checklist–Concepts of Print subtest. WWC analyses of the Christie et al. (2003) data show that the difference between the *Doors to Discovery*™ group and the comparison group on Get Ready to Read! (1.01) and the Developing Skills Checklist–Concepts of Print subtest (0.37) is positive and substantively important, but not statistically significant according to WWC criteria. The mean effect for the print knowledge domain (0.70) was substantively important but not statistically significant. The WWC characterizes these study findings as a substantively important positive effect.

Assel et al. (2007) examined the effectiveness of *Doors to Discovery*™ on print knowledge using the Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest. The authors reported differences between *Doors to Discovery*™ and the comparison group within program type (Head Start, Title I, and universal prekindergarten) rather than across the combined program types. WWC analyses of the Assel et al. (2007) data show that the effect size for the

Doors to Discovery™ group is not statistically significant nor substantively important according to WWC criteria. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the print knowledge domain, one study showed substantively important positive effects and one study showed indeterminate effects. This results in a rating of potentially positive effects, with a medium to large extent of evidence.

Table 4. Rating of effectiveness and extent of evidence for the print knowledge domain

Rating of effectiveness	Criteria met
Potentially positive effects *Evidence of a positive effect with no overriding contrary evidence.*	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the *print knowledge* domain was positive and substantively important in one study, and was neither statistically significant nor large enough to be substantively important in the other study.
Extent of evidence	**Criteria met**
Medium to large	Two studies that included 402 children in 58 classrooms reported evidence of effectiveness in the *print knowledge* domain.

Summary of effectiveness for the phonological processing domain

One study that meets WWC standards with reservations reported findings in the phonological processing domain.

The PCER Consortium (2008, Chapter 6) analyzed the effectiveness of *Doors to Discovery*™ on child outcomes in phonological processing using the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest. The authors reported that differences between the *Doors to Discovery*™ group and the comparison group were not statistically significant and, according to WWC criteria, were not large enough to be considered substantively important. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the phonological processing domain, one study showed indeterminate effects. This results in a rating of no discernible effects, with a small extent of evidence.

Table 5. Rating of effectiveness and extent of evidence for the phonological processing domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *phonological processing* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 182 children in 29 classrooms reported evidence of effectiveness in the *phonological processing* domain.

Summary of effectiveness for the math domain

One study that meets WWC standards with reservations reported findings in the math domain.

The PCER Consortium (2008, Chapter 6) analyzed the effectiveness of *Doors to Discovery*™ on child outcomes in math using the WJ-III Applied Problems subtest, the Child Math Assessment–Abbreviated (CMA-A), and the Building Blocks Shape Composition task. The authors reported that differences between the *Doors to Discovery*™ group and the comparison group were not statistically significant and, according to WWC criteria, were not large enough to be considered substantively important. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the math domain, one study showed indeterminate effects. This results in a rating of no discernible effects, with a small extent of evidence.

Table 6. Rating of effectiveness and extent of evidence for the math domain

Rating of effectiveness	Criteria met
No discernible effects effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *math* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 183 children in 29 classrooms reported evidence of effectiveness in the *math* domain.

References

Study that meets WWC evidence standards without reservations

Christie, J., Roskos, K., Vukelich, C., & Han, M. (2003). 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. Proceedings of the Head Start National Research Conference* (pp. 447–448). New York: Mailman School of Public Health, Columbia University.

Additional source:

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), 333–345.

Studies that meet WWC evidence standards with reservations

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2007). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing, 20*(5), 463–494.

Preschool Curriculum Evaluation Research (PCER) Consortium. (2008, Chapter 6). 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.

Study that does not meet WWC evidence standards

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 it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Studies that are ineligible for review using the Early Childhood Education Evidence Review Protocol

Chambers, B., Cheung, A., Slavin, R. E., Smith, D., & Laurenzano, M. (2010). *Effective early childhood education programs: A systematic review.* Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education. Retrieved from http://www.bestevidence.org/word/early_child_ed_Sep_22_2010.pdf The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

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 design or a single-case design.

Appendix A.1: Research details for Christie et al. (2003)

Christie, J., Roskos, K., Vukelich, C., & Han, M. (2003). 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. Proceedings of the Head Start National Research Conference* (pp. 447–478). New York: Mailman School of Public Health, Columbia University.

Table A1. Summary of findings

Meets WWC evidence standards without reservations

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Oral language	4 classrooms/37 children	+11	No
Print knowledge	4 classrooms/37 children	+25	No

Setting

The study was conducted with children from five Head Start classrooms in a large metropolitan area in the southwest United States.

Study sample

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 comparison group. The four-classroom analysis sample included 21 children in the *Doors to Discovery*™ group and 16 children in the comparison group.

Intervention group

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 group

The comparison classrooms used the existing curriculum, which the study authors described as loosely based on *The Creative Curriculum*®.

Outcomes and measurement

The outcomes assessed were children's oral language and print knowledge. Oral language was assessed with the PPVT-III. Print knowledge was assessed with Get Ready to Read! and the Developing Skills Checklist—Concepts of Print subtest. All assessments were conducted in English (J. Christie, personal communication, January 23, 2009). For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

A professional development specialist employed by the distributer conducted a 5-hour training session on *Doors to Discovery*™ for teachers in the intervention group. During the training, teachers were provided with specific guidance on how to implement the first unit. Research assistants provided teachers with suggestions about how to implement the second and third units before teachers introduced these units in the classroom.

Appendix A.2: Research details for Assel et al. (2007)

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2007). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing, 20*(5), 463–494.

Table A2. Summary of findings

Meets WWC evidence standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Print knowledge	54 classrooms/365 children	+3	No

Setting

The study was conducted with children from 20 full-day preschools (54 classrooms: 27 *Doors to Discovery*™ and 27 comparison) selected from Head Start and public preschool programs in the greater Houston, Texas area.

Study sample

This randomized controlled study, conducted during the 2003–04 school year, included three groups: *Doors to Discovery*™, *Let's Begin with the Letter People*®, and a comparison group. Study authors recruited 32 Title I and non-Title I (universal) preschools and Head Start centers that included a total of 79 classrooms. Within each of the three school types (Title I, non-Title I, and Head Start), schools were randomly assigned to either the *Doors to Discovery*™ intervention group, the *Let's Begin with the Letter People*® intervention group, or a comparison group, with all classrooms within a preschool being assigned to the same intervention condition. Fourteen schools (27 classrooms) in the study were assigned to implement the *Doors to Discovery*™ curriculum, 12 schools (25 classrooms) implemented *Let's Begin with the Letter People*®, and six schools (27 classrooms) were assigned to the comparison condition. In a second stage of random assignment, those schools assigned to *Doors to Discovery*™ and *Let's Begin with the Letter People*® were randomly assigned (within each curriculum) to two groups: one that received mentoring, and one that did not receive mentoring. The number of children at random assignment is not provided in the article or in author queries. However, the authors indicated that the consent rate was 65% in the intervention classrooms and 55% in the comparison classrooms.[8] The authors also noted that at random assignment, 215 study children were in intervention group classrooms, and 203 study children were in comparison classrooms.[9] Using the consent rate and the number of study children, we estimate that the sample of children totaled 729: 324 in intervention classrooms and 405 in comparison classrooms. At baseline, 51% of children in the Head Start classrooms were Hispanic, and 38% were African American; in the Title I classrooms, 53% of children were Hispanic, and 27% were Caucasian; and in the non-Title I programs, 71% of children were Caucasian, and 15% were categorized as Other. Pretest data were collected prior to the implementation of the curriculum, and posttest data were collected at the end of the school year. The analysis sample for the *Doors to Discovery*™ study included 365 children (183 *Doors to Discovery*™ and 182 comparison).

Although the Assel et al. (2007) study used a randomized controlled trial design to assign schools to intervention or comparison conditions, the study had high attrition at the child level and must demonstrate baseline equivalence between the intervention and comparison group sample of children used in the analyses of outcomes. An author query was conducted to obtain the study data necessary to establish equivalence at baseline (i.e., unadjusted means and standard deviations of the outcome measures for the intervention and comparison groups). Baseline equivalence was established for outcomes in the print knowledge domain but not for the oral language or phonological processing domains. The study also discusses differences in child outcomes for the groups that received mentoring compared with those that did not, but since the estimated differences are not presented in the paper, we do not present these analyses in this intervention report.

Intervention group

Intervention group teachers implemented *Doors to Discovery*™. Eight thematic units cover topics such as nature, friendship, communities, society, and health and present rich information. The teacher guide offers open-ended statements and questions to promote discussion. Fidelity to the curriculum was measured three times during the school year. At the first evaluation, 29% of teachers scored at high levels for curriculum fidelity (4 or 5 on a 5-point scale). By mid-year, 57% of teachers received high scores for curriculum fidelity.

A second intervention group was assigned to the *Let's Begin with the Letter People*® curriculum; the effects of this intervention on the study sample are not discussed in this report.

Comparison group

Comparison group teachers used nonspecific curricula, which included a variety of curriculum materials that followed state guidelines for public preschool programs. Head Start comparison classrooms did not use a curriculum with a specified scope or sequence.

Outcomes and measurement

Print knowledge was assessed with the WJ-III Letter-Word Identification subtest. For a more detailed description of this outcome measure, see Appendix B. In addition, the study authors assessed children in the oral language and phonological processing domains. Oral language was assessed with the Preschool Language Scale, Fourth Edition (PLS-4) Auditory Comprehension Subscale and the Expressive Vocabulary Test (EVT). Phonological processing was assessed with the Developing Skill Checklist (DSC) and the Rhyming section from the WJ-III Sound Awareness subtest. The authors did not establish baseline equivalence on the outcomes in these domains.

Support for implementation

Intervention teachers were trained on *Doors to Discovery*™ by the curriculum's publishing company during a 4-day workshop. Training took place in small groups and included instruction in all content areas. The mentors were senior-level trainers of the curriculum. Teachers in the mentoring condition received help from one of three senior-level trainers of the curriculum who served as mentors. Mentors met with teachers two times a month for about one and a half hours, providing assistance in areas of lesson planning, curriculum components, and fidelity, among other topics. Mentors also identified and discussed areas of improvement for individual teachers. All teachers, regardless of mentoring condition, received three feedback sessions over the course of the school year surrounding their implementation of the intervention.

Appendix A.3: Research details for PCER Consortium (2008, Chapter 6)

Preschool Curriculum Evaluation Research (PCER) Consortium. (2008, Chapter 6). 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.

Table A3. Summary of findings

Meets WWC evidence standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Phonological processing	29 classrooms/182 children	+6	No
Math	29 classrooms/183 children	0	No

Setting

The study was conducted with children from 29 full-day preschool classrooms (14 *Doors to Discovery*™ and 15 comparison) selected from Head Start and public preschool programs in the greater Houston, Texas area.

Study sample

This randomized controlled study, conducted during the 2003–04 and 2004–05 school years, included three groups: *Doors to Discovery*™, *Let's Begin with the Letter People*®, and a comparison group. Study authors recruited 32 Title I and non-Title I (universal) preschools and Head Start centers that included a total of 79 classrooms. Within each of the three school types (Title I, non-Title I, and Head Start), schools were randomly assigned to either the *Doors to Discovery*™ intervention group, the *Let's Begin with the Letter People*® intervention group, or a comparison group, with all classrooms within a preschool being assigned to the same intervention condition. Fourteen schools (27 classrooms) in the study were assigned to implement the *Doors to Discovery*™ curriculum, 12 schools (25 classrooms) implemented *Let's Begin with the Letter People*®, and six schools (27 classrooms) were assigned to the comparison condition.

Subsequent to randomization, teachers were provided with a description of the national PCER study; of those teachers that opted to participate in the national PCER study during the 2003–04 school year, 45 were randomly selected (15 from each group). All 79 classrooms participated in the local investigator's pilot study during the first year. Following the pilot year, and prior to starting the national PCER study, one teacher (and her classroom) dropped out of the study, leaving 14 *Doors to Discovery*™ classrooms,15 *Let's Begin with the Letter People*® classrooms, and 15 comparison classrooms.

The evaluation of *Doors to Discovery*™ included 29 of the 44 classrooms (14 *Doors to Discovery*™ classrooms and 15 comparison classrooms, while the remaining 15 classrooms were assigned to *Let's Begin with the Letter People*®). Seven children (whose parents had provided consent to participate in the study) were randomly selected from each classroom at baseline for inclusion in the study.[10] The number of children participating in the study at baseline was 196 (100 *Doors to Discovery*™ and 96 comparison). The parental consent rate was 65% for the intervention group and 55% for the comparison group.[11] At baseline, children in the study averaged 4.6 years of age; 55% were male; 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 comparison).

For the PCER Consortium (2008, Chapter 6) study, the *Doors to Discovery*™ intervention had been in place for a full (pilot) year when the evaluation year started. Although the PCER Consortium (2008, Chapter 6) study used a randomized controlled trial design to assign schools to intervention or comparison conditions in the pilot year—with all classrooms in a school assigned to the same condition—the study analyzed data from the second year of implementation, when children who had been in the classrooms at random assignment had moved to kindergarten and a new class of children had replaced them. Thus, the study had high attrition at the child level and must demonstrate baseline equivalence between the intervention and comparison group sample of children used in the analyses of outcomes. An author query was conducted to obtain the study data necessary to establish equivalence at baseline for one outcome measure in each domain (i.e., unadjusted means and standard deviations of the outcome measures for the intervention and comparison groups). The pretest data provided for each domain were used to establish baseline equivalence for the domain. Baseline equivalence was established from the data provided by the study authors. Baseline equivalence of the analytic sample of children in the two groups at the end of kindergarten was not available, so findings from the kindergarten follow-up are not reported.

Intervention group

Intervention group teachers implemented *Doors to Discovery*™. For this study, each classroom's fidelity to the curriculum was rated on a 4-point scale, ranging from "not at all" (0) to "high" (3). The average score for the *Doors to Discovery*™ classrooms was 2.13 on this measure.

A second intervention group was assigned to the *Let's Begin with the Letter People*® curriculum; the effects of this intervention on the study sample are not discussed in this report.

Comparison group

Comparison 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 comparison classrooms was 1.0.

Outcomes and measurement

The outcome domains assessed were children's oral language, print knowledge, phonological processing, and math. Only outcomes in the phonological processing and math domains met evidence standards with reservations. Phonological processing was assessed with the Pre-CTOPPP Elision subtest. Math was assessed with the WJ-III Applied Problems subtest, the CMA-A, and the Building Blocks Shape Composition task. For a more detailed description of these outcome measures, see Appendix B. Oral language was assessed with the 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 WJ-III Letter-Word Identification subtest, and the WJ-III Spelling subtest. Baseline equivalence was not established for outcomes in the oral language and print knowledge domains, and therefore, these findings are not reported.

Support for implementation

Teachers received curriculum training prior to the start of the 2003–04 school year. This was the second year of implementation of the intervention, and most of the teachers had been trained prior to the start of the 2002–03 school year. New teachers each received 12 hours of training, and returning teachers each received 6 hours of training. 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.

Appendix B: Outcome measures for each domain

Oral language	
Peabody Picture Vocabulary Test–Third Edition (PPVT-III)	A nationally-standardized, individually-administered assessment of children's receptive vocabulary in which children demonstrate understanding of a spoken word by pointing to a picture that best represents the meaning (as cited in Christie et al., 2003 and PCER Consortium, 2008, Chapter 6).
Test of Language Development–Primary III (TOLD-P:3) Grammatic Understanding subtest	A nationally-standardized, individually-administered assessment 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, Chapter 6).
Print knowledge	
Developing Skills Checklist–Concepts of Print subtest	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!	A 20-question, nonstandardized screening test designed to measure emergent writing skills (identifying clearest writing exemplars), print knowledge (differentiating print from pictures, letter naming, and identifying letter sounds), and phonological awareness (rhyming, segmenting words, and deletion of sounds) (as cited in Assel et al., 2007 and J. Christie, personal communication, January 23, 2009).
Test of Early Reading Ability–III (TERA-3)	A nationally-standardized, individually-administered assessment of children's developing reading skills with three subtests: alphabet, conventions, and meaning (as cited in PCER Consortium, 2008, Chapter 6).[12]
Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest	A nationally-standardized, individually-administered assessment of identification of letters and reading of words (as cited in Assel et al., 2007 and PCER Consortium, 2008, Chapter 6).
WJ-III Spelling subtest	A nationally-standardized, individually-administered assessment that assesses children's prewriting skills, such as drawing lines, tracing, and writing letters (as cited in PCER Consortium, 2008, Chapter 6).
Phonological processing	
Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest	An individually-administered assessment 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, Chapter 6).
Math	
Building Blocks Shape Composition task	An individually-administered assessment of early mathematics achievement, this measure was modified for PCER from the Early Maths Assessment, developed by Clements, Sarama, and Liu (2008).[14] 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, Chapter 6).
Child Math Assessment–Abbreviated (CMA-A) Composite Score	An individually-administered assessment of early mathematics achievement, this measure is the average of four subscales: (a) solving addition and subtraction problems using visible objects, (b) constructing a set of objects equal in number to a given set, (c) recognizing shapes, and (d) copying a pattern using objects that vary in color and identity from the model pattern. This assessment was adapted for PCER from a more comprehensive early mathematics assessment by Klein and Starkey (2002), who also developed the pre-K mathematics curriculum and participated in one of the research teams for PCER (as cited in PCER Consortium, 2008, Chapter 6).[13]
WJ-III Applied Problems subtest	A nationally-standardized, individually-administered assessment of children's ability to solve numerical and spatial problems, presented verbally with accompanying pictures of objects (as cited in PCER Consortium, 2008, Chapter 6).

Appendix C.1: Findings included in the rating for the oral language domain

Outcome measure	Study sample	Sample size	Mean (standard deviation) Intervention group	Mean (standard deviation) Comparison group	WWC calculations Mean difference	WWC calculations Effect size	WWC calculations Improvement index	*p*-value
Christie et al., 2003[a]								
Peabody Picture Vocabulary Test–Third Edition (PPVT-III)	Preschool children	4 classrooms/ 37 children	35.98 (22.55)	30.25 (17.30)	5.73	0.27	+11	> 0.05
Domain average for oral language (Christie et al., 2003)						**0.27**	**+11**	**Not statistically significant**
Domain average for oral language across all studies						**0.27**	**+11**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on child outcomes, representing the average change expected for all children who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average child's percentile rank that can be expected if the child is given the intervention. The WWC calculated the program-group mean using a difference-in-differences approach (see WWC Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. na = not applicable.

[a] For Christie et al. (2003), means, standard deviations, and effect sizes were computed by the WWC from child-level data provided by the authors. *p*-values were not reported in the original study. A correction for clustering was needed and resulted in a WWC-computed *p*-value of 0.63 for the *PPVT-III;* therefore, the WWC does not find the result to be statistically significant. This study is characterized as having a substantively important positive effect because the single effect within the domain is positive, substantively important, and not statistically significant.

Appendix C.2: Findings included in the rating for the print knowledge domain

Outcome measure	Study sample	Sample size	Mean (standard deviation) Intervention group	Mean (standard deviation) Comparison group	WWC calculations Mean difference	WWC calculations Effect size	WWC calculations Improvement index	*p*-value
Christie et al., 2003[a]								
Developing Skills Checklist–Concepts of Print subtest	Preschool children	4 classrooms/ 37 children	4.48 (1.56)	2.82 (1.68)	1.66	1.01	+34	nr
Get Ready to Read!	Preschool children	4 classrooms/ 37 children	8.62 (4.16)	7.06 (4.07)	1.56	0.37	+14	nr
Domain average for print knowledge (Christie et al., 2003)						**0.69**	**+25**	**Not statistically significant**
Assel et al., 2007[b]								
Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest	Preschool children	54 classrooms/ 365 children	11.01 (5.12)	9.99 (5.01)	1.02	0.08	+3	nr
Domain average for print knowledge (Assel et al., 2007)						**0.08**	**+3**	**Not statistically significant**
Domain average for print knowledge across all studies						**0.39**	**+15**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on child outcomes, representing the average change expected for all children who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average child's percentile rank that can be expected if the child is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. The WWC calculated the intervention group mean using a difference-in-differences approach (see WWC Handbook) by adding the impact of the intervention (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. nr = not reported. na = not applicable.

[a] For Christie et al. (2003), means, standard deviations, and effect sizes were computed by the WWC from child-level data provided by the authors. *p*-values were not reported in the original study. A correction for clustering was needed and resulted in a WWC-computed *p*-value of 0.09 for the *Developing Skills Checklist–Concepts of Print subtest* and a WWC-computed *p*-value of 0.52 for *Get Ready to Read!;* therefore, the WWC does not find the results to be statistically significant. This study is characterized as having a substantively important positive effect because the effect for at least one measure within the domain is positive and substantively important, and no effects are negative and statistically significant or substantively important.

[b] For Assel et al. (2007), the effect size presented here was calculated by the WWC using data provided by the authors, adjusting for the pretest. The study did not report group differences or effect sizes. A correction for clustering was needed and resulted in a WWC-computed *p*-value of 0.76 for the *WJ-III Letter-Word Identification subtest*; therefore, the WWC does not find the results to be statistically significant. This study is characterized as having an indeterminate effect because the mean effect is neither statistically significant nor substantively important.

Appendix C.3: Findings included in the rating for the phonological processing domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
PCER Consortium, 2008, Chapter 6[a]								
Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest	Preschool children	29 classrooms/ 182 children	nr	nr	nr	0.14	+6	> 0.05
Domain average for phonological processing (PCER Consortium, 2008, Chapter 6)						**0.14**	**+6**	**Not statistically significant**
Domain average for phonological processing across all studies						**0.14**	**+6**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on child outcomes, representing the average change expected for all children who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average child's percentile rank that can be expected if the child is given the intervention. nr = not reported. na = not applicable.

[a] For PCER Consortium (2008, Chapter 6), the effect sizes and *p*-values presented here were reported in the original study (in Table A-12, based on an alternative estimation approach, analysis of covariance [ANCOVA], that included the baseline pretest). The sample sizes reported here were presented in Table C-6a in the original study. A correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. Mean scores and differences are not reported in this table because the study-reported group means and differences were not adjusted for the baseline pretest scores. This study is characterized as having an indeterminate effect because the mean effect is neither statistically significant nor substantively important.

Appendix C.4: Findings included in the rating for the math domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
PCER Consortium, 2008, Chapter 6[a]								
Building Blocks Shape Composition task	Preschool children	29 classrooms/ 183 children	nr	nr	nr	–0.13	–5	> 0.05
Child Math Assessment–Abbreviated (CMA-A) Composite	Preschool children	29 classrooms/ 183 children	nr	nr	nr	0.16	+6	> 0.05
Woodcock-Johnson III (WJ-III) Applied Problems subtest	Preschool children	29 classrooms/ 183 children	nr	nr	nr	0.00	0	> 0.05
Domain average for math (PCER Consortium, 2008, Chapter 6)						**0.01**	**0**	**Not statistically significant**
Domain average for math across all studies						**0.01**	**0**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on child outcomes, representing the average change expected for all children who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average child's percentile rank that can be expected if the child is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. nr = not reported. na = not applicable.

[a] For PCER Consortium (2008, Chapter 6), the effect sizes and *p*-values presented here were reported in the original study (in Table A-12, based on an alternative estimation approach, analysis of covariance [ANCOVA], that included the baseline pretest). The sample sizes reported here were presented in Table C-6a in the original study. A correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. Mean scores and differences are not reported in this table because the study-reported group means and differences were not adjusted for the baseline pretest scores. This study is characterized as having an indeterminate effect because the mean effect is neither statistically significant nor substantively important.

Endnotes

[1] The descriptive information for this program was obtained from publicly available sources: the program's website (https://www.mheonline.com/program/view/5/7/407/0076036243, downloaded February 2012) and the research literature (Assel et al., 2007; PCER Consortium, 2008, Chapter 6). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in March 2012; however, the WWC received no response. 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 December 2012.

[2] The previous report was released in June 2009. This report has been updated to include a review of one study that has been released since June 2009. That study was not within the scope of the review protocol for the Early Childhood Education topic area. A complete list and disposition of all studies reviewed are provided in the references. The report includes reviews of all previous studies that met WWC evidence standards with or without reservations. This resulted in a revised disposition of Christie et al. (2003), Assel et al. (2007), and PCER Consortium (2008, Chapter 6). Christie et al. (2003) is a randomized controlled trial with low attrition that meets standards without reservations. In the previous report, the study met standards with reservations. Assel et al. (2007) is a randomized controlled trial with severe attrition that demonstrated baseline equivalence for the analytic sample through response to an author query. Assel et al. (2007) meets standards with reservations in the current report; the study did not meet standards in the previous report. PCER Consortium (2008, Chapter 6) is a randomized controlled trial with severe attrition that demonstrated baseline equivalence for the analytic sample. This study meets standards with reservations; in the previous report, the study met standards without reservations. The revised dispositions are due to a change in the review protocol, particularly in baseline equivalence standards, as well as information received through author queries. The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1), along with those described in the Early Childhood Education review protocol (version 2.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] Absence of conflict of interest: The PCER Consortium (2008, Chapter 6) study summarized in this intervention report was prepared by staff of one of the WWC contractors. Because the principal investigator for the WWC review of early childhood education is also a staff member of that contractor, the study was rated by staff members from a different organization. The report was then reviewed by the principal investigator, a WWC Quality Assurance reviewer, and an external peer reviewer.

[4] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 20. These improvement index numbers show the average and range of child-level improvement indices for all findings across the studies.

[5] According to Christie et al. (2003), the comparison group was "loosely based" on *The 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).

[6] A related study (Han et al., 2005) used an outcome measure, the ratio of morphemes to utterances, which was measured during "circle time." This measure is overaligned with the intervention because the language-focused curriculum was delivered during circle time. Therefore, the Han et al. (2005) study did not meet WWC standards. However, the paper provided important details about the intervention and the sample used in Christie et al. (2003).

[7] The number of classrooms (79 overall) is from the author query response dated June 13, 2012. Assel et al. (2007) and PCER Consortium (2008, Chapter 6) state that there were 76 classrooms overall.

[8] This information was provided by the study authors, at the WWC's request.

[9] This information was provided by the study authors, at the WWC's request.

[10] PCER Consortium (2008, Chapter 6, 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.

[11] This information was provided by the study authors at the WWC's request.

[12] 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.

[13] Klein, A., & Starkey, P. (2002). *Child Math Assessment-Abbreviated*. Berkeley, CA: Author.

[14] Clements, D. H., Sarama, J., & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The research-based Early Maths Assessment. *Educational Psychology, 28*(4), 457–482.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, June). *Early Childhood Education intervention report: Doors to Discovery*™. Retrieved from http://whatworks.ed.gov

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows 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, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 20.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Rating of effectiveness The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 20.

Single-case design A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.