What Works Clearinghouse

Doors to Discovery™

Program description[1]

Doors to Discovery™ is an early childhood curriculum that uses thematic units to engage young children and support them as they build an understanding of their world. *Doors to Discovery*™ literacy activities are used 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.

Research

One study of *Doors to Discovery*™ met the What Works Clearinghouse (WWC) evidence standards.[2] This study included 76 classrooms in universal pre-kindergarten, Head Start, or Title I programs in the Houston, Texas, metropolitan area and examined intervention effects on children's oral language, print knowledge, and phonological processing. This report focuses on immediate posttest findings to determine the effectiveness of the intervention.[3] The WWC considers the extent of evidence for *Doors to Discovery*™ to be small for oral language, print knowledge, and phonological processing. No studies that met WWC evidence standards with or without reservations addressed early reading/writing, cognition, or math.

Effectiveness

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

	Oral language	*Print knowledge*	*Phonological processing*	*Early reading/writing*	*Cognition*	*Math*
Rating of effectiveness	No discernible effects	No discernible effects	No discernible effects	na	na	na
Improvement index[4]	Average: –8 percentile points Range: –11 to –6 percentile points	Average: +3 percentile points	Average: +5 percentile points Range: +4 to +5 percentile points	na	na	na

na = not applicable

1. The descriptive information for this program was obtained from publicly available sources: the program's web site (http://www.wrightgroup.com/index.php/programsummary?isbn=0076036243, downloaded March 16, 2007) and the research literature (Assel, Landry, Swank, & Gunnewig, 2006). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
2. To be eligible for the WWC's review, the Early Childhood Education (ECE) intervention had to be implemented in English in center-based settings with children ages 3 to 5 or in preschool.
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available. *Doors to Discovery*™ is being studied under the Preschool Curriculum Evaluation Research (PCER) Grants administered through the U.S. Department of Education's Institute of Education Sciences. The final PCER reports were not released in time to be reviewed for this report.
4. These numbers show the average and range of student-level improvement indices for all findings across the study.

Additional program information[1]

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™ 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 of which provides opportunities for children to explore. 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 major 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). The study reviewed also provided other details about the program including extended discussion after storybook reading, the use of monthly themes, and the emphasis on small-group activities.

Cost

The complete *Doors to Discovery*™ set is available to education professionals for $2,130. Alternatively, each theme kit can be purchased separately for $297. 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 web site. The prices listed on the web site are for education professionals only. Information about the cost of professional development is not available.

Research

Two studies reviewed by the WWC investigated the effects of *Doors to Discovery*™ in center-based settings. One study (Assel, Landry, Swank, & Gunnewig, 2006) was a randomized controlled trial that met WWC evidence standards. The remaining study did not meet WWC evidence screens.

Assel et al. (2006) included 76 classrooms from universal pre-kindergarten, Head Start, and Title I programs in the Houston, Texas, metropolitan area. Within these three program types, Assel et al. randomly assigned school sites to one of three conditions (*Doors to Discovery*™*, Let's Begin with the Letter People*®*,* or a business-as-usual comparison condition).[5] Schools in each of the two intervention conditions were further assigned to mentoring and no-mentoring conditions. The WWC is interested in the overall effectiveness of *Doors to Discovery*™. Variations in intervention effects by implementation (with or without mentoring) or program type (universal pre-kindergarten, Head Start, or Title I) are outside the scope of this review. Therefore, the WWC combined the *Doors to Discovery*™ mentoring and *Doors to Discovery*™ no-mentoring groups across program type. The rating of effectiveness is

5. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Doors to Discovery*™ intervention group to the business-as-usual comparison group; however, results for the comparison between the curricula are included in a separate section of this report and Appendices A6.1–A6.3. The WWC includes the *Let's Begin with the Letter People*® versus business-as-usual comparison in a separate WWC *Let's Begin with the Letter People*® intervention report.

Research *(continued)*

based on the comparison of oral language, print knowledge, and phonological processing outcomes of the combined group with the business-as-usual comparison group.[6]

Extent of evidence

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

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

Effectiveness

Findings

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

Oral language. Assel et al. (2006) analyzed the differences between the *Doors to Discovery*™ and business-as-usual comparison groups within program type and by mentoring condition for two measures in this outcome domain [the Preschool Language Scale-IV (PLS-IV) Auditory Comprehension subscale and the Expressive Vocabulary Test (EVT)]. The differences between the intervention and business-as-usual comparison groups combined across program type and mentoring condition were not statistically significant for either outcome as calculated by the WWC, and the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25).

Print knowledge. Assel et al. (2006) analyzed the differences between the *Doors to Discovery*™ and business-as-usual comparison groups within program type and by mentoring condition for one measure in this outcome domain, the Woodcock-Johnson III (W-J III) Letter Word Identification subtest. The difference between the intervention and business-as-usual comparison groups combined across program type and mentoring condition was not statistically significant as calculated by the WWC and it was not large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25).

Phonological processing. Assel et al. (2006) analyzed the differences between the *Doors to Discovery*™ and business-as-

6. The WWC recognizes that this is a different use of the data than intended by the study authors. The study authors reported findings separately for each condition (*Doors to Discovery*™ combined with mentoring, *Doors to Discovery*™ without mentoring) and each program type (universal pre-K, Head Start, or Title I). The WWC could not confirm these findings because critical data (the number of clusters for each condition and program type) were not available. Further, combining the data across mentoring conditions and program types better addresses overall intervention effectiveness, which is the main task for the WWC. Therefore, the WWC analysis, which uses data from the study, differs from the analysis in the original study. The study authors' findings are not reported in the body of this report because the analysis is not comparable to the WWC analysis, but the subgroup analyses for program type and for the mentoring condition are reported in Appendices A4.1–A4.3 and A5.1–A5.3.
7. The Extent of Evidence Categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as the students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.
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, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of *Doors to Discovery*™, a correction for clustering was needed. Assel et al. (2006) described more detailed findings (intervention effects by mentoring condition and program type). The WWC focused on intervention effects combined across these conditions; therefore, the author's findings are not provided but are available in the original study.

Effectiveness *(continued)*

usual comparison groups within program type and by mentoring condition for two measures in this outcome domain [the Developing Skills Checklist (DSC) Auditory subscale and the Rhyming section of the W-J III Sound Awareness subtest]. The differences between the intervention and comparison groups combined across program type and mentoring condition were not statistically significant for either outcome as calculated by the WWC, and the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25).

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings,[8] the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Intervention Rating Scheme).

The WWC found *Doors to Discovery*™ to have no discernible effects on oral language, print knowledge, and phonological processing

Improvement index

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

The average improvement index for oral language is –8 percentile points for the one study, with a range of –11 to –6 percentile points across findings. The improvement index for print knowledge is +3 percentile points for the one outcome in the study. The average improvement index for phonological processing is +5 percentile points for the one study, with a range of +4 to +5 percentile points across findings.

Findings for comparisons between *Doors to Discovery*™ and *Let's Begin with the Letter People*®

The data for the comparison described below were included in the Assel et al. (2006) study, but they do not contribute to the overall rating of effectiveness because the WWC included the comparison of *Doors to Discovery*™ with the business-as-usual comparison group in the rating for the same study, which provides the most direct evidence of *Doors to Discovery's* effects. However, the WWC believes that the findings from this comparison provide useful information to practitioners who may be interested in comparing the effects of different curricula. The WWC reports the findings for comparisons of *Doors to Discovery*™ and *Let's Begin with the Letter People*® here and in Appendices A6.1–A6.3. The WWC analyzed the differences between the *Doors to Discovery*™ and *Let's Begin with the Letter People*® groups combined across program type and mentoring condition.

Oral language. Assel et al. (2006) included data for two measures in this outcome domain. The differences between the *Doors to Discovery*™ and *Let's Begin with the Letter People*® groups were not statistically significant for either measure as calculated by the WWC, and the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25). The average improvement index for oral language is –8 percentile points (*Doors to Discovery*™ is the intervention group and *Let's Begin with the Letter People*® is the comparison group), with a range of –7 to –10 percentile points across findings.

Print knowledge. Assel et al. (2006) included data for one measure in this outcome domain. The difference between the *Doors to Discovery*™ and *Let's Begin with the Letter People*®

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

groups was not statistically significant as calculated by the WWC, and the effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25). The improvement index is –7 percentile points (*Doors to Discovery*™ is the intervention group and *Let's Begin with the Letter People*® is the comparison group) for the one outcome in the study.

Phonological processing. Assel et al. (2006) included data for two measures in this outcome domain, and the WWC analysis indicated a statistically significant difference favoring the *Let's Begin with the Letter People*® group over the *Doors to Discovery*™ group for the DSC Auditory subscale. The finding for the other outcome measure was not statistically significant; however, the average effect size across both outcome measures was large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25). The average improvement index for phonological processing is –10 percentile points (*Doors to Discovery*™ is the intervention group and *Let's Begin with the Letter People*® is the comparison group), with a range of –17 to –3 percentile points across findings.

Summary

The WWC reviewed two studies on *Doors to Discovery*™. One of these studies met WWC evidence standards; the other study did not meet WWC evidence screens. Based on this study, the WWC found no discernible effects for oral language, print knowledge, and phonological processing. Additional findings that were not considered for the rating of effectiveness indicated that *Doors to Discovery*™ and *Let's Begin with the Letter People*® affect children's outcomes similarly in the domains of oral language and print knowledge, but that *Let's Begin with the Letter People*® may have a larger impact on children's phonological processing than *Doors to Discovery*™. The evidence presented in this report may change as new research emerges.

References

Met WWC evidence standards

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2006). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing*. Retrieved March 23, 2007, from http://www.springerlink.com/content/gx325u2h3612817r/fulltext.pdf

Did not meet WWC evidence screens

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.[9]

For more information about specific studies and WWC calculations, please see the WWC *Doors to Discovery*™ Technical Appendices.

9. Complete data are not reported: the WWC could not compute effect sizes because complete study details were not reported.

Appendix

Appendix A1.1 Study characteristics: Assel, Landry, Swank, & Gunnewig (2006) (randomized controlled trial)[1]

Characteristic	Description
Study citation	Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2006). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing.* Retrieved March 23, 2007, from http://www.springerlink.com/content/gx325u2h3612817r/fulltext.pdf
Participants	Within three program types (Head Start, Title I, and universal pre-kindergarten), 32 school sites were randomly assigned to one of three groups (*Doors to Discovery™, Let's Begin with the Letter People®,* or a business-as-usual comparison group).[1] Following assignment to group, school sites in each of the two intervention groups were randomly assigned to one of two groups: a group in which teachers would receive mentoring or a group in which teachers would not receive mentoring. The WWC combined the *Doors to Discovery™* mentoring and *Doors to Discovery™* no-mentoring groups across program type to determine the overall rating of effectiveness.[2] However, the WWC reports additional findings for program type and mentoring in Appendices A4.1–A4.3 and A5.1–A5.3, respectively. The total study sample across all three program types included preschool children with a mean age of 4.6 years at the midpoint of the study; 49% of the children were female; 21% were African-American, 42% were Hispanic, 29% were Caucasian, and 8% were some other race/ethnicity.
Setting	The study took place in 32 universal pre-kindergarten, Head Start, and Title I programs in the Houston, Texas, metropolitan area. Nineteen universal pre-kindergarten classrooms, 31 Head Start classrooms, and 26 Title I classrooms were included and classroom size ranged from 15 to 20 children.
Intervention	Intervention group classrooms used the *Doors to Discovery™* curriculum, which focuses on the development of vocabulary and receptive/expressive language. No information was provided about the implementation of the intervention. In addition to on-site professional development for teachers in a mentoring condition, the mentors observed all classrooms (including those in the no-mentoring condition) and completed a Curriculum Fidelity Checklist three times a year to determine fidelity of implementation and determined that curriculum implementation was good.[3]
Comparison	The business-as-usual comparison group classrooms did not have a specified curriculum. The study authors indicated that the Title I and universal pre-kindergarten classes used various classroom materials (e.g., children's literature from numerous publishers and district-developed materials) that adhered to state guidelines and included language and literacy content. The Head Start classes used a number of materials including pieces from different curricula, various worksheets, and center-developed materials.
Primary outcomes and measurement	The primary outcome domains assessed were children's oral language, print knowledge, and phonological processing. Oral language was assessed with two standardized measures: the Preschool Language Scale-IV (PLS-IV) Auditory Comprehension subscale and the Expressive Vocabulary Test (EVT). Print knowledge was assessed with parts of one standardized measure, the Woodcock-Johnson III (W-J III) Letter Word Identification subtest. Phonological processing was assessed with parts of two standardized measures: the Developing Skills Checklist (DSC) Auditory subscale and the Rhyming section of the W-J III Sound Awareness subtest (see Appendices A2.1–2.3 for more detailed descriptions of the outcome measures). The study authors also conducted observations on a randomly selected group of classrooms using the CIRCLE-Teacher Behavior Rating Scale. The results from these observations are not included in this WWC review.[4]
Teacher training	The teachers were trained at a four-day workshop by individuals from the publishing companies. All training was provided in a small-group format, was learner-centered, and was built on previously learned information. Teachers who were in the mentoring classes received ongoing mentoring from senior level trainers for about an hour and a half twice a month.

1. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Doors to Discovery*™ group to the business-as-usual comparison group; however, results for the comparison between the curricula are included in Appendices A6.1–A6.3. The WWC includes the *Let's Begin with the Letter People®* versus business-as-usual comparison in a separate WWC *Let's Begin with the Letter People®* intervention report.
2. The WWC recognizes that this is a different use of the data than intended by the study authors; however, the WWC is interested in the overall effectiveness of *Doors to Discovery*™. Variations in intervention effects by implementation (with or without mentoring) or program type (universal pre-kindergarten, Head Start, or Title I) are outside the scope of this review.
3. Children in the other intervention group used the *Let's Begin with the Letter People®* curriculum, which focuses on the development of language and literacy as well as science, math, art, music, social development, and motor skills. No information was provided about the implementation of the intervention.
4. For further details about the outcomes included in the Early Childhood Education topic review, please see the Early Childhood Education Protocol.

Appendix A2.1 Outcome measures in the oral language domain

Outcome measure	Description
Preschool Language Scale-IV (PLS-IV) Auditory Comprehension subscale	A subscale from a standardized measure of children's understanding of complex language forms, including structure, grammar, and syntax, as well as their receptive vocabulary (as cited in Assel et al., 2006).
Expressive Vocabulary Test (EVT)	A standardized measure of children's expressive vocabulary and word retrieval that requires children to label objects or to provide synonyms for words (as cited in Assel et al., 2006).

Appendix A2.2 Outcome measure in the print knowledge domain

Outcome measure	Description
Woodcock-Johnson III (W-J III) Letter Word Identification subtest	A subtest from a standardized measure that assesses children's ability to identify letters and words in varying formats (e.g., multiple choice or free response) (as cited in Assel et al., 2006).

Appendix A2.3 Outcome measures in the phonological processing domain

Outcome measure	Description
Developing Skills Checklist (DSC) Auditory subscale	A subscale from a standardized measure that assesses children's ability to recognize words that sound different, to rhyme, and to segment sentences and words (as cited in Assel et al., 2006).
Rhyming section of the W-J III Sound Awareness subtest	A section from a subtest of a standardized measure that assesses children's rhyming (as cited in Assel et al., 2006).

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

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial)[9]								
PLS-IV Auditory Comprehension subscale	Preschool children	24/366	81.46 (18.05)	83.96 (14.65)	–2.50	–0.15	ns	–6
EVT	Preschool children	24/364	87.51 (15.10)	91.44 (14.19)	–3.93	–0.27	ns	–11
Domain average[10] for oral language						–0.21	ns	–8

ns = not statistically significant
PLS-IV = Preschool Language Scale-IV
EVT = Expressive Vocabulary Test

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. The WWC combined the *Doors to Discovery™* mentoring and no-mentoring groups across program type for the rating of effectiveness. Findings from the same study for program type, mentoring, and the head-to-head comparison of *Doors to Discovery™* and *Let's Begin with the Letter People®* are not included in these ratings, but are reported in Appendices A4.1, A5.1, and A6.1, respectively.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. Because school sites and not classrooms were the unit of assignment, the WWC used school sites to correct for clustering. The school site sample sizes provided in this table and used in our analyses are estimates based upon the information provided in the article, which affects the accuracy of the calculation of the statistical significance of the effect size. Specifically, the article reports that there were 10 Head Start centers and 22 pre-K and Title I schools. Because these units cannot be evenly distributed among three conditions, the WWC took a liberal approach and assumed that four school sites were assigned to each condition within each program type. When statistical significance was found with this liberal approach, using a more conservative estimate did not change the statistical significance.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial)[9]								
W-J III Letter Word Identification subtest	Preschool children	24/391	14.01 (6.37)	13.56 (5.67)	0.45	0.07	ns	+3
Domain average[10] for print knowledge						0.07	ns	+3

ns = not statistically significant
W-J III = Woodcock-Johnson III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. The WWC combined the *Doors to Discovery™* mentoring and no-mentoring groups across program type for the rating of effectiveness. Findings from the same study for program type, mentoring, and the head-to-head comparison of *Doors to Discovery™* and *Let's Begin with the Letter People®* are not included in these ratings, but are reported in Appendices A4.2, A5.2, and A6.2, respectively. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. Because school sites and not classrooms were the unit of assignment, the WWC used school sites to correct for clustering. The school site sample sizes provided in this table and used in our analyses are estimates based upon the information provided in the article, which affects the accuracy of the calculation of the statistical significance of the effect size. Specifically, the article reports that there were 10 Head Start centers and 22 pre-K and Title I schools. Because these units cannot be evenly distributed among three conditions, the WWC took a liberal approach and assumed that four school sites were assigned to each condition within each program type. When statistical significance was found with this liberal approach, using a more conservative estimate did not change the statistical significance.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
8. 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. 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial)[9]								
DSC Auditory subscale	Preschool children	24/349	38.02 (12.42)	36.87 (11.62)	1.15	0.10	ns	+4
W-J III Rhyming	Preschool children	24/391	4.40 (5.32)	3.76 (4.38)	0.64	0.13	ns	+5
Domain average[10] for phonological processing						0.11	ns	+5

ns = not statistically significant
DSC = Developing Skills Checklist
W-J III = Woodcock-Johnson III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. The WWC combined the *Doors to Discovery™* mentoring and no-mentoring groups across program type for the rating of effectiveness. Findings from the same study for program type, mentoring, and the head-to-head comparison of *Doors to Discovery™* and *Let's Begin with the Letter People®* are not included in these ratings, but are reported in Appendices A4.3, A5.3, and A6.3, respectively. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. Because school sites and not classrooms were the unit of assignment, the WWC used school sites to correct for clustering. The school site sample sizes provided in this table and used in our analyses are estimates based upon the information provided in the article, which affects the accuracy of the calculation of the statistical significance of the effect size. Specifically, the article reports that there were 10 Head Start centers and 22 pre-K and Title I schools. Because these units cannot be evenly distributed among three conditions, the WWC took a liberal approach and assumed that four school sites were assigned to each condition within each program type. When statistical significance was found with this liberal approach, using a more conservative estimate did not change the statistical significance.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of findings for *Doors to Discovery™* collapsed across mentoring condition by program type for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial; Head Start sites)[9]								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/132	76.74 (13.06)	79.00 (10.42)	–2.26	–0.19	nr	–8
EVT	Preschool children	nr/133	85.30 (16.09)	85.39 (15.49)	–0.09	–0.01	nr	0
Assel et al., 2006 (randomized controlled trial; Title I sites)[9]								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/122	79.83 (18.27)	82.63 (14.12)	–2.80	–0.17	nr	–7
EVT	Preschool children	nr/122	94.84 (13.39)	92.74 (10.98)	2.10	0.17	nr	+7
Assel et al., 2006 (randomized controlled trial; universal pre-K sites)[9]								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/112	91.59 (14.09)	92.86 (16.73)	–1.27	–0.08	nr	–3
EVT	Preschool children	nr/109	85.46 (13.04)	99.34 (10.66)	–13.88	–1.15	nr	–37

nr = not reported
PLS-IV = Preschool Language Scale-IV
EVT = Expressive Vocabulary Test

1. This appendix presents subgroup findings for program type collapsed across mentoring condition for measures that fall in the oral language domain. Total group scores (i.e., combined data across mentoring condition and program type) 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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. The WWC determined that sufficient information was provided to estimate the number of school sites by program type per condition at the total group level (i.e., combined data across mentoring condition and program type) but not for subgroups.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the

(continued)

Appendix A4.1 Summary of findings for *Doors to Discovery™* collapsed across mentoring condition by program type for the oral language domain[1] *(continued)*

intervention group had higher pretest scores than the comparison group. In the Head Start sites, the main effects are driven by the fact that the *Let's Begin with the Letter People®* group began ½ standard deviation lower than the *Doors to Discovery™* group and the comparison group on the PLS-IV measure.

6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), the statistical significance of the effect sizes could not be calculated because the WWC was unable to obtain the number of schools in each condition and program type.

Appendix A4.2 Summary of findings for *Doors to Discovery™* collapsed across mentoring condition by program type for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial; Head Start sites)[9]								
W-J III Letter Word Identification subtest	Preschool children	nr/181	11.75 (5.20)	11.85 (5.21)	–0.10	–0.02	nr	–1
Assel et al., 2006 (randomized controlled trial; Title I sites)[9]								
W-J III Letter Word Identification subtest	Preschool children	nr/111	13.92 (4.92)	14.19 (5.11)	–0.27	–0.05	nr	–2
Assel et al., 2006 (randomized controlled trial; universal pre-K sites)[9]								
W-J III Letter Word Identification subtest	Preschool children	nr/99	18.59 (6.61)	17.39 (5.66)	1.20	0.19	nr	+8

nr = not reported
W-J III = Woodcock-Johnson III

1. This appendix presents subgroup findings for program type collapsed across mentoring condition for measures that fall in the print knowledge domain. Total group scores (i.e., combined data across mentoring condition and program type) were used for rating purposes and are presented in Appendix A3.2. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. The WWC determined that sufficient information was provided to estimate the number of school sites by program type per condition at the total group level (i.e., combined data across mentoring condition and program type) but not for subgroups.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), the statistical significance of the effect sizes could not be calculated because the WWC was unable to obtain the number of schools in each condition and program type.

Appendix A4.3 Summary of findings for *Doors to Discovery™* collapsed across mentoring condition by program type for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial; Head Start sites)[9]								
DSC Auditory subscale	Preschool children	nr/115	37.97 (11.93)	33.98 (12.21)	3.99	0.33	nr	+13
W-J III Rhyming	Preschool children	nr/181	1.79 (2.32)	2.18 (2.97)	–0.39	–0.14	nr	–6
Assel et al., 2006 (randomized controlled trial; Title I sites)[9]								
DSC Auditory subscale	Preschool children	nr/122	40.19 (11.76)	38.24 (11.22)	1.95	0.17	nr	+7
W-J III Rhyming	Preschool children	nr/111	3.82 (3.92)	3.96 (4.35)	–0.14	–0.03	nr	–1
Assel et al., 2006 (randomized controlled trial; universal pre-K sites)[9]								
DSC Auditory subscale	Preschool children	nr/112	37.74 (11.12)	39.10 (10.68)	–1.36	–0.12	nr	–5
W-J III Rhyming	Preschool children	nr/99	9.83 (4.10)	7.81 (5.06)	2.02	0.45	nr	+17

nr = not reported
DSC = Developing Skills Checklist
W-J III = Woodcock-Johnson III

1. This appendix presents subgroup findings for program type collapsed across mentoring condition for measures that fall in the phonological processing domain. Total group scores (i.e., combined data across mentoring condition and program type) were used for rating purposes and are presented in Appendix A3.3. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. The WWC determined that sufficient information was provided to estimate the number of school sites by program type per condition at the total group level (i.e., combined data across mentoring condition and program type) but not for subgroups.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.

(continued)

Appendix A4.3 Summary of findings for *Doors to Discovery™* collapsed across mentoring condition by program type for the phonological processing domain[1] *(continued)*

6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), the statistical significance of the effect sizes could not be calculated because the WWC was unable to obtain the number of schools in each condition and program type.

Appendix A5.1 Summary of findings for *Doors to Discovery™* collapsed across program type by mentoring condition for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial; mentoring condition)[9]								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/275	80.80 (17.56)	83.96 (14.65)	–3.16	–0.20	nr	–8
EVT	Preschool children	nr/273	86.89 (13.10)	91.44 (14.19)	–4.55	–0.33	nr	–13
Assel et al., 2006 (randomized controlled trial; no mentoring condition)[9]								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/273	82.14 (18.33)	83.96 (14.65)	–1.82	–0.11	nr	–5
EVT	Preschool children	nr/272	88.15 (16.96)	91.44 (14.19)	–3.29	–0.22	nr	–9

nr = not reported
PLS-IV = Preschool Language Scale-IV
EVT = Expressive Vocabulary Test

1. This appendix presents subgroup findings for mentoring condition collapsed across program type for measures that fall in the oral language domain. Total group scores (i.e., combined data across mentoring condition and program type) 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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. The WWC determined that sufficient information was provided to estimate the number of school sites by program type per condition at the total group level (i.e., combined data across mentoring condition and program type) but not for subgroups.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
6. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), the statistical significance of the effect sizes could not be calculated because the WWC was unable to obtain the number of schools in each condition and program type.

Appendix A5.2 Summary of findings for *Doors to Discovery™* collapsed across program type by mentoring condition for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial; mentoring condition)[9]								
W-J III Letter Word Identification subtest	Preschool children	nr/298	13.08 (5.93)	13.56 (5.67)	–0.48	–0.08	nr	–3
Assel et al., 2006 (randomized controlled trial; no mentoring condition)[9]								
W-J III Letter Word Identification subtest	Preschool children	nr/274	15.17 (6.89)	13.56 (5.67)	1.61	0.26	nr	+10

nr = not reported

W-J III = Woodcock-Johnson III

1. This appendix presents subgroup findings for mentoring condition collapsed across program type for measures that fall in the print knowledge domain. Total group scores (i.e., combined data across mentoring condition and program type) were used for rating purposes and are presented in Appendix A3.2. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. The WWC determined that sufficient information was provided to estimate the number of school sites by program type per condition at the total group level (i.e., combined data across mentoring condition and program type) but not for subgroups.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
6. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), the statistical significance of the effect sizes could not be calculated because the WWC was unable to obtain the number of schools in each condition and program type.

Appendix A5.3 Summary of findings for *Doors to Discovery™* collapsed across program type by mentoring condition for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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[4]	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]
Assel et al., 2006 (randomized controlled trial; mentoring condition)[9]								
DSC Auditory subscale	Preschool children	nr/262	41.44 (12.55)	36.87 (11.62)	4.57	0.38	nr	+15
W-J III Rhyming	Preschool children	nr/298	4.40 (4.88)	3.76 (4.38)	0.64	0.14	nr	+5
Assel et al., 2006 (randomized controlled trial; no mentoring condition)[9]								
DSC Auditory subscale	Preschool children	nr/257	34.41 (12.35)	36.87 (11.62)	–2.46	–0.21	nr	–8
W-J III Rhyming	Preschool children	nr/274	4.42 (5.71)	3.76 (4.38)	0.66	0.13	nr	+5

nr = not reported
DSC = Developing Skills Checklist
W-J III = Woodcock-Johnson III

1. This appendix presents subgroup findings for mentoring condition collapsed across program type for measures that fall in the phonological processing domain. Total group scores (i.e., combined data across mentoring condition and program type) were used for rating purposes and are presented in Appendix A3.3. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. The WWC determined that sufficient information was provided to estimate the number of school sites by program type per condition at the total group level (i.e., combined data across mentoring condition and program type) but not for subgroups.
4. The intervention group mean equals the comparison group mean plus the mean difference.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
6. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), the statistical significance of the effect sizes could not be calculated because the WWC was unable to obtain the number of schools in each condition and program type.

Appendix A6.1 Summary of findings for comparisons between *Doors to Discovery™* and *Let's Begin with the Letter People®* for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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]) — *Let's Begin with the Letter People®* group[4]	WWC calculations: Mean difference[5] (*Doors to Discovery™ – Let's Begin with the Letter People®*)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Assel et al., 2006 (randomized controlled trial)[9]								
PLS-IV Auditory Comprehension subscale	Preschool children	24/368	89.30 (18.05)	92.53 (17.78)	–3.23	–0.18	ns	–7
EVT	Preschool children	24/366	92.61 (15.10)	96.91 (19.70)	–4.30	–0.24	ns	–10
Domain average[10] for oral language						–0.21	ns	–8

ns = not statistically significant
PLS-IV = Preschool Language Scale-IV
EVT = Expressive Vocabulary Test

1. This appendix presents findings for the head-to-head comparison of *Doors to Discovery™* and *Let's Begin with the Letter People®* for measures that fall in the oral language domain. For each intervention, the WWC combined mentoring and no-mentoring groups across program type. Comparisons of *Doors to Discovery™* and the business-as-usual comparison group 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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. Because school sites and not classrooms were the unit of assignment, the WWC used school sites to correct for clustering. The school site sample sizes provided in this table and used in our analyses are estimates based upon the information provided in the article, which affects the accuracy of the calculation of the statistical significance of the effect size. Specifically, the article reports that there were 10 Head Start centers and 22 pre-K and Title I schools. Because these units cannot be evenly distributed among three conditions, the WWC took a liberal approach and assumed that four school sites were assigned to each condition within each program type. When statistical significance was found with this liberal approach, using a more conservative estimate did not change the statistical significance.
4. The *Doors to Discovery™* group mean equals the *Let's Begin with the Letter People®* group mean plus the mean difference.
5. Positive differences and effect sizes favor the *Doors to Discovery™* group; negative differences and effect sizes favor the *Let's Begin with the Letter People®* group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the *Doors to Discovery™* group had lower pretest scores than the *Let's Begin with the Letter People®* group and underestimate the intervention's effects when the *Doors to Discovery™* group had higher pretest scores than the *Let's Begin with the Letter People®* group.
6. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Doors to Discovery™* condition versus the percentile rank of the average student in the *Let's Begin with the Letter People®* condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Doors to Discovery™* 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A6.2 Summary of findings for comparisons between *Doors to Discovery™* and *Let's Begin with the Letter People®* for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	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]) — *Let's Begin with the Letter People®* group[4]	WWC calculations: Mean difference[5] (*Doors to Discovery™ – Let's Begin with the Letter People®*)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Assel et al., 2006 (randomized controlled trial)[9]								
W-J III Letter Word Identification subtest	Preschool children	24/368	14.28 (6.37)	15.43 (6.72)	–1.15	–0.17	ns	–7
Domain average[10] for print knowledge						–0.17	ns	–7

ns = not statistically significant
W-J III = Woodcock-Johnson III

1. This appendix presents findings for the head-to-head comparison of *Doors to Discovery™* and *Let's Begin with the Letter People®* for measures that fall in the print knowledge domain. For each intervention, the WWC combined mentoring and no-mentoring groups across program type. Comparisons of *Doors to Discovery™* and the business-as-usual comparison group were used for rating purposes and are presented in Appendix A3.2. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. Because school sites and not classrooms were the unit of assignment, the WWC used school sites to correct for clustering. The school site sample sizes provided in this table and used in our analyses are estimates based upon the information provided in the article, which affects the accuracy of the calculation of the statistical significance of the effect size. Specifically, the article reports that there were 10 Head Start centers and 22 pre-K and Title I schools. Because these units cannot be evenly distributed among three conditions, the WWC took a liberal approach and assumed that four school sites were assigned to each condition within each program type. When statistical significance was found with this liberal approach, using a more conservative estimate did not change the statistical significance.
4. The *Doors to Discovery™* group mean equals the *Let's Begin with the Letter People®* group mean plus the mean difference.
5. Positive differences and effect sizes favor the *Doors to Discovery™* group; negative differences and effect sizes favor the *Let's Begin with the Letter People®* group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the *Doors to Discovery™* group had lower pretest scores than the *Let's Begin with the Letter People®* group and underestimate the intervention's effects when the *Doors to Discovery™* group had higher pretest scores than the *Let's Begin with the Letter People®* group.
6. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Doors to Discovery™* condition versus the percentile rank of the average student in the *Let's Begin with the Letter People®* condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Doors to Discovery™* 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A6.3 Summary of findings for comparisons between *Doors to Discovery™* and *Let's Begin with the Letter People®* for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (schools/ children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Doors to Discovery™* group[4]	*Let's Begin with the Letter People®* group[4]	WWC calculations: Mean difference[5] (*Doors to Discovery™ – Let's Begin with the Letter People®*)	Effect size[6]	Statistical significance[7] (at $\alpha = 0.05$)	Improvement index[8]
Assel et al., 2006 (randomized controlled trial)[9]								
DSC Auditory subscale	Preschool children	24/360	39.60 (12.42)	45.44 (13.25)	–5.84	-0.45	Statistically significant	–17
W-J III Rhyming	Preschool children	24/368	5.31 (5.32)	5.69 (5.59)	–0.38	-0.07	ns	–3
Domain average[10] for phonological processing						-0.26	ns	–10

ns = not statistically significant
DSC = Developing Skills Checklist
W-J III = Woodcock-Johnson III

1. This appendix presents findings for the head-to-head comparison of *Doors to Discovery™* and *Let's Begin with the Letter People®* for measures that fall in the phonological processing domain. For each intervention, the WWC combined mentoring and no-mentoring groups across program type. Comparisons of *Doors to Discovery™* and the business-as-usual comparison group were used for rating purposes and are presented in Appendix A3.3. The W-J III data separated by program type and mentoring condition were provided by the study authors upon WWC request.
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. Although the study authors provided the total number of school sites by program type in the study, they did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request. Because school sites and not classrooms were the unit of assignment, the WWC used school sites to correct for clustering. The school site sample sizes provided in this table and used in our analyses are estimates based upon the information provided in the article, which affects the accuracy of the calculation of the statistical significance of the effect size. Specifically, the article reports that there were 10 Head Start centers and 22 pre-K and Title I schools. Because these units cannot be evenly distributed among three conditions, the WWC took a liberal approach and assumed that four school sites were assigned to each condition within each program type. When statistical significance was found with this liberal approach, using a more conservative estimate did not change the statistical significance.
4. The *Doors to Discovery™* group mean equals the *Let's Begin with the Letter People®* group mean plus the mean difference.
5. Positive differences and effect sizes favor the *Doors to Discovery™* group; negative differences and effect sizes favor the *Let's Begin with the Letter People®* group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the *Doors to Discovery™* group had lower pretest scores than the *Let's Begin with the Letter People®* group and underestimate the intervention's effects when the *Doors to Discovery™* group had higher pretest scores than the *Let's Begin with the Letter People®* group.
6. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Doors to Discovery™* condition versus the percentile rank of the average student in the *Let's Begin with the Letter People®* condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Doors to Discovery™* 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Assel et al. (2006), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A7.1 *Doors to Discovery™* rating for the oral language domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of oral language, the WWC rated *Doors to Discovery*™ as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

Other ratings considered

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

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

  Not met. Only one study examined effects on oral language.

AND

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important negative effects.

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

AND

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative, but it did show indeterminate effects.

(continued)

Appendix A7.1 *Doors to Discovery™* rating for the oral language domain *(continued)*

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

OR

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important negative effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. The study single study reviewed in this domain not show statistically significant or substantively important positive effects.

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

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

  Not met. Only one study examined effects on oral language.

AND

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

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

Appendix A7.2 *Doors to Discovery™* rating for the print knowledge domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of print knowledge, the WWC rated *Doors to Discovery*™ as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

Other ratings considered

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

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

  Not met. Only one study examined effects on print knowledge.

AND

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important negative effects.

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

AND

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative, but it did show indeterminate effects.

(continued)

Appendix A7.2 *Doors to Discovery™* rating for the print knowledge domain *(continued)*

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

OR

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important negative effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

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

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

  Not met. Only one study examined effects on print knowledge.

AND

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

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

Appendix A7.3 *Doors to Discovery™* rating for the phonological processing domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of phonological processing, the WWC rated *Doors to Discovery*™ as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

Other ratings considered

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

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

  Not met. Only one study examined effects on phonological processing.

AND

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important negative effects.

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

AND

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative, but it did show indeterminate effects.

(continued)

Appendix A7.3 Doors to Discovery™ rating for the phonological processing domain *(continued)*

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

OR

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important effects, either positive or negative.

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

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

  Not met. The single study reviewed in this domain did not show statistically significant or substantively important negative effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

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

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

  Not met. Only one study examined effects on phonological processing.

AND

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

  Met. The single study reviewed in this domain did not show statistically significant or substantively important positive effects.

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

Appendix A8 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[2]
		Centers[1]	Classrooms/children	
Oral language	1	24	52/366	Small
Print knowledge	1	24	52/391	Small
Phonological processing	1	24	52/391	Small
Early reading/writing	0	0	0	na
Cognition	0	0	0	na
Math	0	0	0	na

na = not applicable/not studied

1. This is the estimated number of school sites because the study authors did not provide the number of school sites by program type assigned to each condition in the article or in response to the WWC request.
2. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."