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]

			Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation[2])					
Outcome measure	Study sample	Sample size (schools/ children)[3]	*Doors to Discovery™* group[4]	Comparison group[4]	Mean difference[5] (*Doors to Discovery™* – comparison)	Effect size[6]	Statistical significance[7] (at $\alpha = 0.05$)	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]

			Authors' findings from the study					
			Mean outcome (standard deviation[2])		WWC calculations			
Outcome measure	**Study sample**	**Sample size (schools/ children)[3]**	***Doors to Discovery™* group[4]**	***Let's Begin with the Letter People®* group[4]**	**Mean difference[5] (*Doors to Discovery™* – *Let's Begin with the Letter People®*)**	**Effect size[6]**	**Statistical significance[7] (at α = 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."