

Appendix

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

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. <i>Reading and Writing</i> . 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 (<i>Let's Begin with the Letter People</i> ®, <i>Doors to Discovery</i> ™, 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 <i>Let's Begin with the Letter People</i> ® mentoring and <i>Let's Begin with the Letter People</i> ® 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 <i>Let's Begin with the Letter People</i> ® curriculum, which includes 26 thematic units focusing 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. In addition to providing on-site professional development for teachers in the 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 the curriculum was being implemented at high levels. 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 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 *Let's Begin with the Letter People*® 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 *Doors to Discovery*™ versus business-as-usual comparison in a separate WWC *Doors to Discovery*™ 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 *Let's Begin with the Letter People*®. 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 *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.
4. For further details about the outcomes included in the Early Childhood Education topic review, please see the [Early Childhood Education Protocol](#).

Appendix A1.2 Study characteristics: Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller (in press) (randomized controlled trial)

Characteristic	Description
Study citation	Fischel, J. E., Bracken, S. S., Fuchs-Eisenberg, A., Spira, E. G., Katz, S., & Shaller, G. (in press). Evaluation of curricular approaches to enhance preschool early literacy skills. <i>Journal of Literacy Research</i> .
Participants	Twenty-seven classrooms were randomly assigned to one of three groups (<i>Let's Begin with the Letter People</i> ®, <i>Waterford Early Reading Level One</i> ™, or a business-as-usual comparison group) across the three years of the study. ¹ In year one of the study, six classrooms were assigned to the <i>Let's Begin with the Letter People</i> ® or business-as-usual comparison groups (three <i>Letter People</i> classes and three business-as-usual comparison classes). In year two of the study, eight new classrooms were assigned to these groups (three <i>Letter People</i> classes and five business-as-usual comparison classes) and two randomly selected <i>Letter People</i> classrooms from year one participated again. In year three of the study, five new classrooms were assigned to these groups (two <i>Letter People</i> classes and three business-as-usual comparison classes) and two randomly selected <i>Letter People</i> classrooms from year one participated again. ² The total study sample across all three groups and all three study years included preschool children with a mean age of 4 years, 4 months at the time of pretest. The children were 42% African-American, 41% Hispanic, 8% multi-racial, 7% Caucasian, and 2% were some other race/ethnicity. About 14% of the total sample was Spanish-language dominant at Head Start entry.
Setting	The study took place in 27 unique classrooms across conditions in six Head Start centers (four in year one, one additional center in year two, and one additional center in year three) in southeastern New York. All centers were part of the same Head Start grantee. In each year of the study, children attended full-day preschool, five days a week.
Intervention	Intervention group classrooms used the <i>Let's Begin with the Letter People</i> ® curriculum, which was overlaid on the existing <i>High/Scope</i> curriculum. <i>Let's Begin with the Letter People</i> ® addresses a broad array of language and literacy skills, as well as numeracy, art, music, science, social and motor development through 26 curriculum units organized around five main themes. No information was provided about the implementation of the intervention; however, fidelity was measured by the trainer during each classroom visit and was determined to be accurate. ³
Comparison	The business-as-usual comparison group classrooms used the standard classroom curriculum (<i>High/Scope</i>), which prescribes a daily routine (planning time, work time, cleanup time, time for recall, large-group time, small-group time, and outdoor play) and aligns well with Head Start's performance standards, focusing on language, literacy, and other school readiness skills such as numeracy, reasoning, problem-solving, and decision-making.
Primary outcomes and measurement⁴	The primary outcome domains assessed were children's oral language and print knowledge. Oral language was assessed with a standardized measure [the Peabody Picture Vocabulary Test-III (PPVT-III)] and a non-standardized measure (Comprehension). Print knowledge was assessed with six measures: Get Ready to Read! Screen (a non-standardized measure), Letters Known (a non-standardized measure), the Letter Word Identification and Dictation subtests from the Woodcock Johnson-Revised (WJ-R; a standardized measure), Book Knowledge (a non-standardized measure), and Print Conventions (a non-standardized measure) (see Appendices A2.1–2.2 for more detailed descriptions of outcome measures).
Teacher training	Teachers and teacher assistants in the <i>Let's Begin with the Letter People</i> ® group participated in a three-day curriculum training each August conducted by a professional trainer from Abrams and Company (the developer and distributor of this curriculum). The trainer visited each classroom in the <i>Let's Begin with the Letter People</i> ® condition in the fall and spring of each intervention year and provided individual feedback to teachers. Fischel et al. (in press) reported that additional training was offered by the trainer; however, details of the frequency, content, or degree of participation in these trainings were not provided. Teachers and assistants in the <i>Let's Begin with the Letter People</i> ® group and the business-as-usual comparison group participated in a week-long in-service <i>High/Scope</i> curriculum training at the beginning of the school year. Support was provided in the classroom by educational and child development specialists throughout the school year.

1. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Let's Begin with the Letter People*® group to the business-as-usual comparison group; however, results for the comparison between the curricula are included in Appendices A7.1–7.2. The WWC includes the *Waterford Early Reading Level One*™ versus business-as-usual

(continued)

Appendix A1.2 Study characteristics: Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller (in press) (randomized controlled trial) *(continued)*

comparison in a separate [WWC Waterford Early Reading Level One™ intervention report](#). Both intervention groups used the studied intervention in conjunction with the *High/Scope* curriculum, which was the standard curriculum used by the classrooms prior to the study.

2. This same process yielded three *Waterford Early Reading Level One™* classrooms in year one, five *Waterford Early Reading Level One™* classrooms (three new classrooms and two repeat classrooms) in year two, and four *Waterford Early Reading Level One™* classrooms (two new classrooms and two repeat classrooms) in year three. The WWC includes the data from children participating in classrooms that had not participated in previous waves (that is, children from unique classrooms) because including all instances of classrooms involved a confound of past study involvement with assignment. The possible effects of this confound could not be tested because no business-as-usual comparison classrooms were studied for a second year.
3. Children in the other intervention group used the *Waterford Early Reading Level One™* curriculum, which was overlaid on the existing *High/Scope* curriculum. Each child participated in the computerized instruction for 15 minutes a day and the related books and videos were incorporated into small- and large-group time within the *High/Scope* framework.
4. At pretest the “Spanish-dominant” children were assessed with Spanish versions of the PPVT-III, the WJ-R Letter Word Identification subtest, and the WJ-R Dictation subtest and English versions of the PPVT-III and the WJ-R Letter Word Identification subtest. For other measures, the instructions were translated into Spanish, but the measure was administered in English. The book used for the Book Knowledge, Print Conventions, and Comprehension measures was also translated into Spanish. Posttest measures were administered in English only and the results reported by the study authors include only the English language version of the measures. Because the Dictation subtest was administered to Spanish-dominant children in Spanish only, the scores reported for Dictation by Fischel et al. (in press) exclude Spanish-dominant children.

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).
Peabody Picture Vocabulary Test-III (PPVT-III)	A standardized measure of children's receptive vocabulary that requires children to identify pictures that correspond to words spoken aloud by the assessor (as cited in Fischel et al., in press).
Comprehension	A measure developed for the Family and Child Experiences Survey (FACES) used in each of the Head Start Quality Research Centers, where a child is handed the <i>Where's My Teddy</i> storybook and asked a series of questions designed to assess story comprehension (e.g., how a character feels) (as cited in Fischel et al., in press).

Appendix A2.2 Outcome measures 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).
Get Ready to Read! Screen¹	A non-standardized measure of readiness for reading instruction focusing on three core domains (print knowledge, emergent writing skills, and linguistic awareness) across 20 items to which children indicate their response by pointing (as cited in Fischel et al., in press).
Letters Known	A measure—developed for FACES and used in each of the Head Start Quality Research Centers—designed to assess children's letter knowledge by asking children to identify as many letters as possible from three incrementally difficult letter groupings. Once children are finished naming letters in a group, the assessor asks the child if he/she recognizes any of the other letters (as cited in Fischel et al., in press).
Woodcock Johnson-Revised (WJ-R) Letter Word Identification subtest	A subtest from a standardized measure of children's ability to name printed letters and words (as cited in Fischel et al., in press).
WJ-R Dictation subtest	A subtest from a standardized measure of children's prewriting skills such as drawing lines, copying letters, writing letters, writing phrases, punctuation, and capitalization (as cited in Fischel et al., in press).
Book Knowledge	A measure—developed for FACES and used in each of the Head Start Quality Research Centers—where a child is handed the <i>Where's My Teddy</i> storybook inverted and backwards and asked a series of questions about book knowledge (e.g., where is the front of the book and where do you start reading) (as cited in Fischel et al., in press).
Print Conventions	A measure—developed for FACES and used in each of the Head Start Quality Research Centers—where a child is handed the <i>Where's My Teddy</i> storybook inverted and backwards and asked a series of questions about print conventions such as reading left-to-right and top-to-bottom (as cited in Fischel et al., in press).

1. The WWC placed this measure in the print knowledge domain because the majority of the items are about print knowledge and the measure correlates most highly with other measures of alphabet knowledge.

Appendix A2.3 Outcome measures in the phonological awareness 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 domain¹

Outcome measure	Study sample	Sample size (schools or classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁴ (<i>Let's Begin with the Letter People</i> [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			<i>Let's Begin with the Letter People</i> [®] group ³	Comparison group ³				
Assel et al., 2006 (randomized controlled trial)⁸								
PLS-IV	Preschool children	24/366 ⁹	84.69 (17.78)	83.96 (14.65)	0.73	0.04	ns	+2
EVT	Preschool children	24/364 ⁹	91.82 (19.70)	91.44 (14.19)	0.38	0.02	ns	+1
Average¹⁰ for oral language (Assel et al., 2006)						0.03	ns	+1
Fischel et al., in press (randomized controlled trial)¹¹								
PPVT-III	Preschool children	19/272 ¹²	86.59 (13.80)	85.72 (13.68)	0.87	0.06	ns	+3
Comprehension	Preschool children	19/277 ¹²	0.89 (0.77)	0.90 (0.74)	-0.01	-0.01	ns	-1
Average¹⁰ for oral language (Fischel et al., in press)						0.02	ns	+1
Domain average¹⁰ for oral language across all studies						0.03	na	+1

ns = not statistically significant

na = not applicable

PLS-IV = Preschool Language Scale-IV

EVT = Expressive Vocabulary Test

PPVT-III = Peabody Picture Vocabulary Test-III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. For Assel et al. (2006), the WWC combined the *Let's Begin with the Letter People*[®] 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 *Let's Begin with the Letter People*[®] and *Doors to Discovery*[™] are not included in these ratings, but are reported in Appendices A4.1, A5.1, and A6.1, respectively. For Fischel et al. (in press), additional findings for the head-to-head comparison of *Let's Begin with the Letter People*[®] and *Waterford Early Reading Level One*[™] are not included in these ratings, but are reported in Appendix A7.1. The WWC includes the data from children participating in classrooms that had not participated in previous waves (that is, children from unique classrooms) because including all instances of classrooms involved a confound of past study involvement with assignment. The possible effects of this confound could not be tested because no business-as-usual comparison classrooms were studied for a second year.
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. For Fischel et al. (in press), the standard deviations were provided by the study authors upon WWC request.
3. For Assel et al. (2006), the intervention group mean equals the comparison group mean plus the mean difference. For Fischel et al. (in press), the posttest means are covariate-adjusted means provided by the study authors upon WWC request.

(continued)

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

4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Assel et al. (2006), 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.
5. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition 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.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). 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.
9. 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.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.
11. In the case of Fischel et al. (in press), a correction for clustering was needed, so the significance levels may differ from those reported in the original study. Further, the WWC analysis of Fischel et al. (in press) focused on new teachers while the original study reported findings based on analysis of new and experienced teachers; this also may cause the significance levels reported to differ from those reported in the original study.
12. The child-level posttest sample sizes were provided by the study authors upon WWC request.

Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain¹

Outcome measure	Study sample	Sample size (schools or classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁴ (<i>Let's Begin with the Letter People</i> [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			<i>Let's Begin with the Letter People</i> [®] group ³	Comparison group ³				
Assel et al., 2006 (randomized controlled trial)⁸								
W-J III Letter Word Identification subtest	Preschool children	24/339 ⁹	15.15 (6.72)	13.56 (5.67)	1.59	0.26	ns	+10
Average¹⁰ for print knowledge (Assel et al., 2006)						0.26	ns	+10
Fischel et al., in press (randomized controlled trial)¹¹								
Get Ready to Read! Screen	Preschool children	19/281 ¹²	12.62 (3.70)	11.59 (3.83)	1.03	0.27	ns	+11
Letters Known	Preschool children	19/277 ¹²	17.80 (9.01)	15.86 (9.68)	1.94	0.21	ns	+8
WJ-R Letter Word Identification subtest	Preschool children	19/235 ¹²	98.08 (12.06)	96.69 (11.90)	1.39	0.12	ns	+5
WJ-R Dictation subtest	Preschool children	19/194 ¹²	93.48 (15.48)	88.93 (15.03)	4.55	0.30	ns	+12
Book Knowledge	Preschool children	19/277 ¹²	2.85 (1.37)	2.53 (1.27)	0.32	0.24	ns	+10
Print Conventions	Preschool children	19/277 ¹²	0.43 (0.74)	0.27 (0.60)	0.16	0.24	ns	+9
Average¹⁰ for print knowledge (Fischel et al., in press)						0.23	ns	+9
Domain average¹⁰ for print knowledge across all studies						0.24	na	+10

ns = not statistically significant

na = not applicable

W-J III = Woodcock-Johnson III

WJ-R = Woodcock Johnson-Revised

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. For Assel et al. (2006), the WWC combined the *Let's Begin with the Letter People*[®] 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 *Let's Begin with the Letter People*[®] and *Doors to Discovery*[™] 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. For Fischel et al. (in press), findings for the head-to-head comparison of *Let's Begin with the Letter People*[®] and *Waterford Early Reading Level One*[™] are not included in these ratings, but are reported in Appendix 7.2. The WWC includes the data from children participating

(continued)

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

- in classrooms that had not participated in previous waves (that is, children from unique classrooms) because including all instances of classrooms involved a confound of past study involvement with assignment. The possible effects of this confound could not be tested because no business-as-usual comparison classrooms were studied for a second year.
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. For Fischel et al. (in press), the standard deviations were provided by the study authors upon WWC request.
 3. For Assel et al. (2006), the intervention group mean equals the comparison group mean plus the mean difference. For Fischel et al. (in press), the posttest means are covariate-adjusted means provided by the study authors upon WWC request.
 4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Assel et al. (2006), 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.
 5. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
 6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
 7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition 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.
 8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). 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.
 9. 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.
 10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.
 11. In the case of Fischel et al. (in press), a correction for clustering was needed, so the significance levels may differ from those reported in the original study. Further, the WWC analysis of Fischel et al. (in press) focused on new teachers while the original study reported findings based on analysis of new and experienced teachers; this also may cause the significance levels reported to differ from those reported in the original study.
 12. The child-level posttest sample sizes were provided by the study authors upon WWC request.

Appendix A3.3 Summary of study findings included in the rating for the phonological processing domain¹

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> [®] – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> [®] group ⁴	Comparison group ⁴				
Assel et al., 2006 (randomized controlled trial)⁹								
DSC Auditory subscale	Preschool children	24/351	43.86 (13.25)	36.87 (11.62)	6.99	0.56	Statistically significant	+21
W-J III Rhyming	Preschool children	24/339	4.78 (5.59)	3.76 (4.38)	1.02	0.20	ns	+8
Domain average¹⁰ for phonological processing						0.38	ns	+15

ns = not statistically significant

DSC = Developing Skills Checklist

W-J III = Woodcock-Johnson III

- This appendix reports findings considered for the effectiveness rating and the average improvement indices. The WWC combined the *Let's Begin with the Letter People*[®] 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 *Let's Begin with the Letter People*[®] and *Doors to Discovery*[™] 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.
- 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.
- 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 per WWC policy. 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.
- The intervention group mean equals the comparison group mean plus the mean difference.
- 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.
- For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
- Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
- 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.
- 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), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
- 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 *Let's Begin with the Letter People*[®] collapsed across mentoring condition by program type for the oral language domain¹

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> [®] – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> [®] group ⁴	Comparison group ⁴				
Assel et al., 2006 (randomized controlled trial; Head Start sites)⁹								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/156	82.10 (11.54)	79.00 (10.42)	3.10	0.28	nr	+11
EVT	Preschool children	nr/156	91.31 (22.45)	85.39 (15.49)	5.92	0.30	nr	+12
Assel et al., 2006 (randomized controlled trial; Title I sites)⁹								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/116	82.18 (15.30)	82.63 (14.12)	-0.45	-0.03	nr	-1
EVT	Preschool children	nr/116	92.42 (12.32)	92.74 (10.98)	-0.32	-0.03	nr	-1
Assel et al., 2006 (randomized controlled trial; universal pre-K sites)⁹								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/94	91.62 (12.42)	92.86 (16.73)	-1.24	-0.08	nr	-3
EVT	Preschool children	nr/92	91.18 (6.72)	99.34 (10.66)	-8.16	-0.89	nr	-31

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.

(continued)

Appendix A4.1 Summary of findings for *Let's Begin with the Letter People*[®] collapsed across mentoring condition by program type for the oral language domain¹ (continued)

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. 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 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.

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> [®] – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> [®] group ⁴	Comparison group ⁴				
Assel et al., 2006 (randomized controlled trial; Head Start sites)⁹								
W-J III Letter Word Identification subtest	Preschool children	nr/162	11.63 (4.94)	11.85 (5.21)	–0.22	–0.04	nr	–2
Assel et al., 2006 (randomized controlled trial; Title I sites)⁹								
W-J III Letter Word Identification subtest	Preschool children	nr/95	16.03 (4.79)	14.19 (5.11)	1.84	0.37	nr	+14
Assel et al., 2006 (randomized controlled trial; universal pre-K sites)⁹								
W-J III Letter Word Identification subtest	Preschool children	nr/82	20.56 (6.66)	17.39 (5.66)	3.17	0.50	nr	+19

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 *Let's Begin with the Letter People*® collapsed across mentoring condition by program type for the phonological processing domain¹

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> ® – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> ® group ⁴	Comparison group ⁴				
Assel et al., 2006 (randomized controlled trial; Head Start sites)⁹								
DSC Auditory subscale	Preschool children	nr/141	44.55 (12.05)	33.98 (12.21)	10.57	0.87	nr	+31
W-J III Rhyming	Preschool children	nr/162	1.74 (1.79)	2.18 (2.97)	-0.44	-0.17	nr	-7
Assel et al., 2006 (randomized controlled trial; Title I sites)⁹								
DSC Auditory subscale	Preschool children	nr/116	46.96 (13.72)	38.24 (11.22)	8.72	0.69	nr	+25
W-J III Rhyming	Preschool children	nr/95	5.16 (5.89)	3.96 (4.35)	1.20	0.23	nr	+9
Assel et al., 2006 (randomized controlled trial; universal pre-K sites)⁹								
DSC Auditory subscale	Preschool children	nr/94	37.60 (12.25)	39.10 (10.68)	-1.50	-0.13	nr	-5
W-J III Rhyming	Preschool children	nr/82	9.93 (4.49)	7.81 (5.06)	2.12	0.44	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 *Let's Begin with the Letter People*® collapsed across mentoring condition by program type for the phonological processing domain¹ (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.

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> [®] – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> [®] group ⁴	Comparison group ⁴				
Assel et al., 2006 (randomized controlled trial; mentoring condition)⁹								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/261	83.04 (17.80)	83.96 (14.65)	–0.92	–0.06	nr	–2
EVT	Preschool children	nr/260	90.45 (19.04)	91.44 (14.19)	–0.99	–0.06	nr	–2
Assel et al., 2006 (randomized controlled trial; no-mentoring condition)⁹								
PLS-IV Auditory Comprehension subscale	Preschool children	nr/287	85.94 (17.86)	83.96 (14.65)	1.98	0.12	nr	+5
EVT	Preschool children	nr/285	92.86 (20.28)	91.44 (14.19)	1.42	0.09	nr	+3

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 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 (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 *Let's Begin with the Letter People*[®] collapsed across program type by mentoring condition for the print knowledge domain¹

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> [®] – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> [®] group ⁴	Comparison group ⁴				
Assel et al., 2006 (randomized controlled trial; mentoring condition)⁹								
W-J III Letter Word Identification subtest	Preschool children	nr/257	14.84 (6.86)	13.56 (5.67)	1.28	0.21	nr	+8
Assel et al., 2006 (randomized controlled trial; no-mentoring condition)⁹								
W-J III Letter Word Identification subtest	Preschool children	nr/263	15.43 (6.62)	13.56 (5.67)	1.87	0.31	nr	+12

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 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.

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> ® – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> ® group ⁴	Comparison group ⁴				
Assel et al., 2006 (randomized controlled trial; mentoring condition)⁹								
DSC Auditory subscale	Preschool children	nr/246	42.21 (13.24)	36.87 (11.62)	5.34	0.44	nr	+17
W-J III Rhyming	Preschool children	nr/257	4.77 (5.79)	3.76 (4.38)	1.01	0.21	nr	+8
Assel et al., 2006 (randomized controlled trial; no-mentoring condition)⁹								
DSC Auditory subscale	Preschool children	nr/275	45.06 (13.24)	36.87 (11.62)	8.19	0.67	nr	+25
W-J III Rhyming	Preschool children	nr/263	4.79 (5.42)	3.76 (4.38)	1.03	0.22	nr	+9

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 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 A6.1 Summary of findings for comparisons between *Let's Begin with the Letter People*® and *Doors to Discovery*™ for the oral language domain¹

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> ® – <i>Doors to Discovery</i> ™)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> ® group ⁴	<i>Doors to Discovery</i> ™ group ⁴				
Assel et al., 2006 (randomized controlled trial)⁹								
PLS-IV Auditory Comprehension subscale	Preschool children	24/368	92.53 (17.78)	89.30 (18.05)	3.23	0.18	ns	+7
EVT	Preschool children	24/366	96.91 (19.70)	92.61 (15.10)	4.30	0.24	ns	+10
Domain average¹⁰ 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 *Let's Begin with the Letter People*® and *Doors to Discovery*™ 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 *Let's Begin with the Letter People*® 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 *Let's Begin with the Letter People*® group mean equals the *Doors to Discovery*™ group mean plus the mean difference.
5. Positive differences and effect sizes favor the *Let's Begin with the Letter People*® group; negative differences and effect sizes favor the *Doors to Discovery*™ group. The mean differences were computed by the WWC and took into account pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effects when the *Let's Begin with the Letter People*® group had lower pretest scores than the *Doors to Discovery*™ group and underestimate the intervention's effects when the *Let's Begin with the Letter People*® group had higher pretest scores than the *Doors to Discovery*™ 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 *Let's Begin with the Letter People*® condition versus the percentile rank of the average student in the *Doors to Discovery*™ condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Let's Begin with the Letter People*® 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 *Let's Begin with the Letter People*® and *Doors to Discovery*™ for the print knowledge domain¹

Outcome measure	Study sample	Sample size (schools/ children) ³	Authors' findings from the study					WWC calculations	
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> ® – <i>Doors to Discovery</i> ™)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸	
			<i>Let's Begin with the Letter People</i> ® group ⁴	<i>Doors to Discovery</i> ™ group ⁴					
Assel et al., 2006 (randomized controlled trial)⁹									
W-J III Letter Word Identification subtest	Preschool children	24/368	15.43 (6.72)	14.28 (6.37)	1.15	0.17	ns	+7	
Domain average¹⁰ 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 *Let's Begin with the Letter People*® and *Doors to Discovery*™ 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 *Let's Begin with the Letter People*® 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 *Let's Begin with the Letter People*® group mean equals the *Doors to Discovery*™ group mean plus the mean difference.
5. Positive differences and effect sizes favor the *Let's Begin with the Letter People*® group; negative differences and effect sizes favor the *Doors to Discovery*™ group. The mean differences were computed by the WWC and took into account pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effects when the *Let's Begin with the Letter People*® group had lower pretest scores than the *Doors to Discovery*™ group and underestimate the intervention's effects when the *Let's Begin with the Letter People*® group had higher pretest scores than the *Doors to Discovery*™ 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 *Let's Begin with the Letter People*® condition versus the percentile rank of the average student in the *Doors to Discovery*™ condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Let's Begin with the Letter People*® 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.

Summary of findings for comparisons between *Let's Begin with the Letter People*® and *Doors to Discovery*™ for the phonological processing domain¹

Outcome measure	Study sample	Sample size (schools/children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> ® – <i>Doors to Discovery</i> ™)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Let's Begin with the Letter People</i> ® group ⁴	<i>Doors to Discovery</i> ™ group ⁴				
Assel et al., 2006 (randomized controlled trial)⁹								
DSC Auditory subscale	Preschool children	24/360	45.44 (13.25)	39.60 (12.42)	5.84	0.45	Statistically significant	+17
W-J III Rhyming	Preschool children	24/368	5.69 (5.59)	5.31 (5.32)	0.38	0.07	ns	+3
Domain average¹⁰ for phonological processing						0.26	ns	+10

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

- This appendix presents findings for the head-to-head comparison of *Let's Begin with the Letter People*® and *Doors to Discovery*™ 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 *Let's Begin with the Letter People*® 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.
- 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.
- 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.
- The *Let's Begin with the Letter People*® group mean equals the *Doors to Discovery*™ group mean plus the mean difference.
- Positive differences and effect sizes favor the *Let's Begin with the Letter People*® group; negative differences and effect sizes favor the *Doors to Discovery*™ group. The mean differences were computed by the WWC and took into account pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effects when the *Let's Begin with the Letter People*® group had lower pretest scores than the *Doors to Discovery*™ group and underestimate the intervention's effects when the *Let's Begin with the Letter People*® group had higher pretest scores than the *Doors to Discovery*™ group.
- For an explanation of effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
- Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
- The improvement index represents the difference between the percentile rank of the average student in the *Let's Begin with the Letter People*® condition versus the percentile rank of the average student in the *Doors to Discovery*™ condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the *Let's Begin with the Letter People*® group.
- 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.
- 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 Summary of findings for comparisons between *Let's Begin with the Letter People*® and *Waterford Early Reading Level One*™ for the oral language domain¹

Outcome measure	Study sample	Sample size (classrooms/ children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> ® – <i>Waterford Early Reading Level One</i> ™)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
		<i>Let's Begin with the Letter People</i> ® group ⁴	<i>Waterford Early Reading Level One</i> ™ group ⁴					
Fischel et al., in press (randomized controlled trial)⁹								
PPVT-III	Preschool children	16/241	86.59 (13.80)	86.92 (14.39)	–0.33	–0.02	ns	–1
Comprehension	Preschool children	16/247	0.89 (0.77)	0.85 (0.76)	0.04	0.05	ns	+2
Domain average¹⁰ for oral language						0.01	ns	+1

ns = not statistically significant

PPVT-III = Peabody Picture Vocabulary Test-III

1. This appendix presents findings for the head-to-head comparison of *Let's Begin with the Letter People*® and *Waterford Early Reading Level One*™. Comparisons of *Let's Begin with the Letter People*® and the business-as-usual comparison group were used for rating purposes and are presented in Appendix A3.1. The WWC includes the data from children participating in classrooms that had not participated in previous waves (that is, children from unique classrooms) because including all instances of classrooms involved a confound of past study involvement with assignment. The possible effects of this confound could not be tested because no business-as-usual comparison classrooms were studied for a second year.
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. The standard deviations were provided by the study authors upon WWC request.
3. The child-level posttest sample sizes were provided by the study authors upon WWC request.
4. The posttest means are covariate-adjusted means provided by the study authors upon WWC request.
5. Positive differences and effect sizes favor the *Let's Begin with the Letter People*® group; negative differences and effect sizes favor the *Waterford Early Reading Level One*™ 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 *Let's Begin with the Letter People*® condition versus the percentile rank of the average student in the *Waterford Early Reading Level One*™ condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Let's Begin with the Letter People*® 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 Fischel et al. (in press), 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 A7.2 Summary of findings for comparisons between *Let's Begin with the Letter People*® and *Waterford Early Reading Level One*™ for the print knowledge domain¹

Outcome measure	Study sample	Sample size (classrooms/ children) ³	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁵ (<i>Let's Begin with the Letter People</i> ® – <i>Waterford Early Reading Level One</i> ™)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
		<i>Let's Begin with the Letter People</i> ® group ⁴	<i>Waterford Early Reading Level One</i> ™ group ⁴					
Fischel et al., in press (randomized controlled trial)⁹								
Get Ready to Read! Screen	Preschool children	16/251	12.62 (3.70)	12.84 (3.87)	-0.22	-0.06	ns	-2
Letters Known	Preschool children	16/247	17.80 (9.01)	18.03 (8.81)	-0.23	-0.03	ns	-1
WJ-R Letter Word Identification subtest	Preschool children	16/208	98.08 (12.06)	98.69 (11.41)	-0.61	-0.05	ns	-2
WJ-R Dictation subtest	Preschool children	16/173	93.48 (15.48)	90.37 (14.28)	3.11	0.21	ns	+8
Book Knowledge	Preschool children	16/247	2.85 (1.37)	2.41 (1.37)	0.44	0.32	ns	+13
Print Conventions	Preschool children	16/247	0.43 (0.74)	0.44 (0.77)	-0.01	-0.01	ns	-1
Domain average¹⁰ for print knowledge						0.06	ns	+3

ns = not statistically significant

WJ-R = Woodcock Johnson-Revised

1. This appendix presents findings for the head-to-head comparison of *Let's Begin with the Letter People*® and *Waterford Early Reading Level One*™. Comparisons of *Let's Begin with the Letter People*® and the business-as-usual comparison group were used for rating purposes and are presented in Appendix A3.2. The WWC includes the data from children participating in classrooms that had not participated in previous waves (that is, children from unique classrooms) because including all instances of classrooms involved a confound of past study involvement with assignment. The possible effects of this confound could not be tested because no business-as-usual comparison classrooms were studied for a second year.
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. The standard deviations were provided by the study authors upon WWC request.
3. The child-level posttest sample sizes were provided by the study authors upon WWC request.
4. The posttest means are covariate-adjusted means provided by the study authors upon WWC request.
5. Positive differences and effect sizes favor the *Let's Begin with the Letter People*® group; negative differences and effect sizes favor the *Waterford Early Reading Level One*™ group.
6. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).

(continued)

Appendix A7.2 Summary of findings for comparisons between *Let's Begin with the Letter People*® and *Waterford Early Reading Level One*™ for the print knowledge domain¹ (continued)

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 *Let's Begin with the Letter People* condition versus the percentile rank of the average student in the *Waterford Early Reading Level One*™ condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the *Let's Begin with the Letter People* 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 Fischel et al. (in press), 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 A8.1 *Let's Begin with the Letter People*[®] 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 *Let's Begin with the Letter People*[®] 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. Neither of the studies showed 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. Neither study showed statistically significant positive effects.

AND

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

Met. Neither study showed 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. Neither study showed statistically significant or substantively important positive effects.

AND

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

Not met. Neither study showed statistically significant or substantively important effects, either positive or negative; both studies showed indeterminate effects.

(continued)

Appendix A8.1 *Let's Begin with the Letter People*[®] 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. No studies showed 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. Neither study showed 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. Neither study showed 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. Neither study showed 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. Neither study showed statistically significant negative effects.

AND

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

Met. Neither study showed 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 A8.2 *Let's Begin with the Letter People*[®] 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 *Let's Begin with the Letter People*[®] as having potentially positive effects. It did not meet the criteria for positive effects as neither of the studies showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Let's Begin with the Letter People*[®] was assigned the highest applicable rating.

Rating received

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

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

Met. One study showed 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.

Met. Neither study showed statistically significant or substantively important negative effects; one study showed substantively important positive effects and the other study showed indeterminate effects.

Other ratings considered

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

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

Not met. Neither study showed statistically significant positive effects.

AND

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

Met. Neither study showed statistically significant or substantively important negative effects.

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

Appendix A8.3 *Let's Begin with the Letter People*[®] 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 *Let's Begin with the Letter People*[®] as having potentially positive effects. It did not meet the criteria for positive effects because only one study examined outcomes in this domain. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Let's Begin with the Letter People*[®] was assigned the highest applicable rating.

Rating received

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

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

Met. The single study reviewed in this domain showed statistically significant 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.

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

Other ratings considered

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

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

Not met. 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.

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 A9 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ²
		Centers ¹	Classrooms/children	
Oral language	2	30	70/643	Moderate to large
Print knowledge	2	30	70/620	Moderate to large
Phonological processing	1	24	51/351	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 Assel et al. (2006) did not provide the number of school sites by program type assigned to each condition in the article or in response to 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.”