Appendix

Appendix A1.1 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. *Journal of Literacy Research.*
Participants	Twenty-seven classrooms were randomly assigned to one of three groups (*Waterford Early Reading Level One™*, *Let's Begin with the Letter People®*, 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 *Waterford Early Reading Level One™* or business-as-usual comparison groups (three *Waterford* classes and three business-as-usual comparison classes). In year two of the study, eight new classrooms were assigned to these groups (three *Waterford* classes and five business-as-usual comparison classes) and two randomly selected *Waterford* classrooms from year one participated again. In year three of the study, five new classrooms were assigned to these groups (two *Waterford* classes and three business-as-usual comparison classes) and two randomly selected *Waterford* 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 *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.[3]
Comparison	The business-as-usual comparison group classrooms used the standard classroom curriculum (*High/Scope*), 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 measurement[4]	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 *Waterford Early Reading Level One™* group participated in a one-day curriculum training each August conducted by a Pearson Digital Learning trainer. Each teacher could be supervised by the trainer while implementing the curriculum and were taught ways to incorporate the materials (videotapes and books) into the curriculum. The trainer visited each *Waterford Early Reading Level One™* classroom mid-year to review summary data to assess classroom progress and provide support and additional training. Teachers and assistants in the *Waterford Early Reading Level One™* group and the business-as-usual comparison group participated in a week-long in-service *High/Scope* 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 *Waterford Early Reading Level One*™ group to the business-as-usual comparison group; however, results for the comparison between the curricula are included in Appendices A4.1–4.2. The WWC includes the *Let's Begin with the Letter People®* versus business-as-usual

(continued)

comparison in a separate WWC *Let's Begin with the Letter People*® 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 *Let's Begin with the Letter People*® classrooms in year one, five *Let's Begin with the Letter People*® classrooms (three new classrooms and two repeat classrooms) in year two, and four *Let's Begin with the Letter People*® 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 *Let's Begin with the Letter People*® curriculum, which was overlaid on the existing *High/Scope* curriculum. *Let's Begin with the Letter People*® addressed 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.
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
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) and used in each of the Head Start Quality Research Centers—where a child is handed the *Where's My Teddy* 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
Get Ready to Read! Screen[1]	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 pre-writing 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 *Where's My Teddy* 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 *Where's My Teddy* 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 A3.1 Summary of study findings included in the rating for the oral language domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Waterford Early Reading Level One™* group[4]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[4]	WWC calculations: Mean difference[5] (*Waterford Early Reading Level One™* – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Fischel et al., in press (randomized controlled trial)[9]								
PPVT-III	Preschool children	19/263	86.92 (14.39)	85.72 (13.68)	1.20	0.09	ns	+3
Comprehension	Preschool children	19/270	0.85 (0.76)	0.90 (0.74)	–0.05	–0.07	ns	–3
Domain average[10] for oral language						0.01	ns	0

ns = not statistically significant
PPVT-III = Peabody Picture Vocabulary Test-III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Additional findings for the head-to-head comparison of *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* are not included in these ratings, but are reported in Appendix A4.1. 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 intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of 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 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.
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (classrooms/ children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Waterford Early Reading Level One™* group[4]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[4]	WWC calculations: Mean difference[5] (*Waterford Early Reading Level One™* – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Fischel et al., in press (randomized controlled trial)[9]								
Get Ready to Read! Screen	Preschool children	19/268	12.84 (3.87)	11.59 (3.83)	1.25	0.32	ns	+13
Letters Known	Preschool children	19/270	18.03 (8.81)	15.86 (9.68)	2.17	0.23	ns	+9
WJ-R Letter Word Identification subtest	Preschool children	19/231	98.69 (11.41)	96.69 (11.90)	2.00	0.17	ns	+7
WJ-R Dictation subtest	Preschool children	19/183	90.37 (14.28)	88.93 (15.03)	1.44	0.10	ns	+4
Book Knowledge	Preschool children	19/270	2.41 (1.37)	2.53 (1.27)	–0.12	-0.09	ns	–4
Print Conventions	Preschool children	19/270	0.44 (0.77)	0.27 (0.60)	0.17	0.25	ns	+10
Domain average[10] for print knowledge						0.16	ns	+7

ns = not statistically significant

WJ-R = Woodcock Johnson-Revised

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Additional findings for the head-to-head comparison of *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* are not included in these ratings, but are reported in Appendix A4.2. 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 intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of 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 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.
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of findings for comparisons between *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* for the oral language domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Waterford Early Reading Level One™* group[4]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Let's Begin with the Letter People®* group[4]	WWC calculations: Mean difference[5] (*Waterford Early Reading Level One™ – Let's Begin with the Letter People®*)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Fischel et al., in press (randomized controlled trial)[9]								
PPVT-III	Preschool children	16/241	86.92 (14.39)	86.59 (13.80)	0.33	0.02	ns	+1
Comprehension	Preschool children	16/247	0.85 (0.76)	0.89 (0.77)	–0.04	–0.05	ns	–2
Domain average[10] 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 *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®*. Comparisons of *Waterford Early Reading Level One™* 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 *Waterford Early Reading Level One™* group; negative differences and effect sizes favor the *Let's Begin with the Letter People®* 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 *Waterford Early Reading Level One™* condition versus the percentile rank of the average student in the *Let's Begin with the Letter People®* condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Waterford Early Reading Level One™* 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 A4.2 Summary of findings for comparisons between *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Waterford Early Reading Level One™* group[4]	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Let's Begin with the Letter People®* group[4]	WWC calculations: Mean difference[5] (*Waterford Early Reading Level One™ – Let's Begin with the Letter People®*)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Fischel et al., in press (randomized controlled trial)[9]								
Get Ready to Read! Screen	Preschool children	16/251	12.84 (3.87)	12.62 (3.70)	0.22	0.06	ns	+2
Letters Known	Preschool children	16/247	18.03 (8.81)	17.80 (9.01)	0.23	0.03	ns	+1
WJ-R Letter Word Identification subtest	Preschool children	16/208	98.69 (11.41)	98.08 (12.06)	0.61	0.05	ns	+2
WJ-R Dictation subtest	Preschool children	16/173	90.37 (14.28)	93.48 (15.48)	–3.11	–0.21	ns	–8
Book Knowledge	Preschool children	16/247	2.41 (1.37)	2.85 (1.37)	–0.44	–0.32	ns	–13
Print Conventions	Preschool children	16/247	0.44 (0.77)	0.43 (0.74)	0.01	0.01	ns	+1
Domain average[10] 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 *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®*. Comparisons of *Waterford Early Reading Level One™* 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 *Waterford Early Reading Level One™* group; negative differences and effect sizes favor the *Let's Begin with the Letter People®* 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 *Waterford Early Reading Level One™* condition versus the percentile rank of the average student in the *Let's Begin with the Letter People®* condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Waterford Early Reading Level One™* 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 A5.1 *Waterford Early Reading Level One*™ 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 *Waterford Early Reading Level One*™ as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

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

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

Other ratings considered

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

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

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

AND

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

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

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

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

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

AND

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

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

(continued)

Appendix A5.1 *Waterford Early Reading Level One*™ rating for the oral language domain *(continued)*

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

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

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

OR

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

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

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

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

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

AND

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

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

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

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

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

AND

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

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

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

Appendix A5.2 *Waterford Early Reading Level One*™ 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 *Waterford Early Reading Level One*™ as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

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

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

Other ratings considered

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

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

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

AND

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

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

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

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

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

AND

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

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

(continued)

Appendix A5.2 *Waterford Early Reading Level One*™ rating for the print knowledge domain *(continued)*

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

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

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

OR

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

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

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

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

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

AND

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

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

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

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

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

AND

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

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

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

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Centers	Classrooms/children	
Oral language	1	6	19/270	Small
Print knowledge	1	6	19/270	Small
Phonological processing	0	0	0	na
Early reading/writing	0	0	0	na
Cognition	0	0	0	na
Math	0	0	0	na

na = not applicable/not studied

1. 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."