

What Works Clearinghouse



Waterford Early Reading Level One™

Program description¹

Waterford Early Reading Level One™ is an emergent literacy curriculum that uses computer-based technology to prepare children for reading. It begins with a tutorial to familiarize the child with the computer and mouse and a reading placement evaluation to assess and determine whether a child should

work on Level One objectives: capital letters, lowercase letters, or beginning decoding skills. The computerized instruction is supplemented by activities for phonological and phonemic awareness, letter recognition, knowledge of story and print concepts, and general readiness skills.

Research

One study of *Waterford Early Reading Level One™* met the What Works Clearinghouse (WWC) evidence standards.² The study included 27 classrooms in six Head Start centers in southeastern New York. This report focuses on immediate posttest findings to determine the effectiveness of the intervention.³ The WWC

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

Effectiveness

Waterford Early Reading Level One™ was found to have no discernible effects on oral language or on print knowledge.

	Oral language	Print knowledge	Phonological processing	Early reading/writing	Cognition	Math
Rating of effectiveness	No discernible effects	No discernible effects	na	na	na	na
Improvement index⁴	Average: 0 percentile points Range: -3 to +3 percentile points	Average: +7 percentile points Range: -4 to +13 percentile points	na	na	na	na

na = not applicable

- The descriptive information for this program was obtained from publicly available sources: the program's web site (<http://www.waterford.org>, downloaded April 18, 2007) and the research literature (Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller, in press). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
- To be eligible for the WWC's review, the Early Childhood Education (ECE) intervention had to be implemented in English in center-based settings with children aged three to five or in preschool.
- The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
- These numbers show the average and range of student-level improvement indices for all findings across the study.

Additional program information¹

Developer and contact

Developed by Waterford Research Institute, *Waterford Early Reading Level One™* is distributed by Pearson Digital Learning. Address: 6710 East Camelback Road, Scottsdale, Arizona 85251. Email: pdinfo@pearson.com. Web: <http://www.pearsondigital.com/waterford/>. Telephone: (888) 977-7900.

Scope of use

According to the developer, more than 500,000 children across the United States use the various programs provided by *Waterford*.

Teaching

Waterford Early Reading Level One™ provides individualized, year-long instruction in daily 15-minute sessions. Teachers are advised to review both class and individual reports at least once

Research

Three studies reviewed by the WWC investigated the effects of *Waterford Early Reading Level One™* in center-based settings. One study (Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller, in press) was a randomized controlled trial that met WWC evidence standards. The remaining two studies did not meet WWC evidence screens.

Fischel et al. (in press) included 27 full-day Head Start classrooms over a three-year period in southeastern New York and compared oral language and print knowledge outcomes for children participating in a *Waterford Early Reading Level One™* intervention group, a *Let's Begin with the Letter People®* intervention group, or a business-as-usual comparison group.⁵ Children in all three conditions received the *High/Scope* curriculum as their base condition. The *Waterford Early Reading Level One™* intervention group used the studied intervention in conjunction with the *High/Scope* curriculum, which was the standard curriculum used by the classrooms prior to the study. The WWC includes the data from children participating in classrooms that had not participated in previous waves (that is,

a month to monitor progress and guide classroom instruction. Based on the students' performance, the teacher can reassign activities to ensure mastery of objectives. The curriculum includes the *Waterford* software, assessment materials, books, and videos. These are used in conjunction with take-home student books, CDs, and handouts. On-site training and online training webinars are available for initial training in addition to a detailed teacher guide. On-site teacher training could include a mid-year visit to review class progress using data from the computerized program.

Cost

For program costs, contact the Pearson representative in your area (see <http://www.pearsondigital.com/waterford/>). Information about the cost of professional development is not available.

children from unique classrooms) because including all instances of classrooms involved a confound of past study involvement with assignment and the possible effects of this confound could not be tested because no business-as-usual comparison classrooms were studied for a second year.

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate to large (see the [What Works Clearinghouse Extent of Evidence Categorization Scheme](#)). The extent of evidence takes into account the number of studies and the total sample size across the studies that met WWC evidence standards with or without reservations.⁶

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

5. For the rating of effectiveness in this WWC intervention report, the WWC includes only the results comparing the *Waterford Early Reading Level One™* intervention group to the business-as-usual comparison group; however, results for the comparison between the curricula are included in a separate section of this report and Appendices A4.1–A4.2. The WWC includes the *Let's Begin with the Letter People®* versus business-as-usual comparison in a separate [WWC Let's Begin with the Letter People® intervention report](#).
6. The Extent of Evidence Categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as the students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.

Effectiveness Findings

The WWC review of interventions for early childhood education addresses children's outcomes in six domains: oral language, print knowledge, phonological processing, early reading/writing, cognition, and math. Fischel et al. (in press) addressed outcomes in the oral language and print knowledge domains. The findings below present the authors' and the WWC-calculated estimates of the size and statistical significance of the effects of *Waterford Early Reading Level One™* on children's performance.⁷

Oral language. Fischel et al. (in press) analyzed the differences between the *Waterford Early Reading Level One™* and business-as-usual comparison groups for two measures in this outcome domain [the Peabody Picture Vocabulary Test-III (PPVT-III) and Comprehension] and found no significant effects; the WWC confirmed this. Furthermore, the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25).

Print knowledge. Fischel et al. (in press) analyzed the differences between the *Waterford Early Reading Level One™* and business-as-usual comparison groups for six measures in this

outcome domain [Get Ready to Read! Screen⁸, Letters Known, Woodcock Johnson-Revised (WJ-R) Letter Word Identification subtest, the WJ-R Dictation subtest, Book Knowledge, and Print Conventions] and found significant differences favoring *Waterford Early Reading Level One™* on one measure, Get Ready to Read! Screen. The WWC could not confirm statistically significant findings for any outcomes in this domain. Furthermore, the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25).

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings,⁷ the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the [WWC Intervention Rating Scheme](#)).

The WWC found *Waterford Early Reading Level One™* to have no discernible effects on oral language or on print knowledge

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see [Technical Details of WWC-Conducted Computations](#)). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement

index is based entirely on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analyses. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.

The average improvement index for oral language is 0 percentile points for the study, with a range of -3 to +3 percentile points across findings. The average improvement index for print knowledge is +7 percentile points for the study, with a range of -4 to +13 percentile points across findings.

7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate the statistical significance. In the case of *Waterford Early Reading Level One™*, a correction for clustering was needed. Fischel et al. (in press) included children from all classes in the analyses. The WWC focused on intervention effects for children in the unique classes only (i.e., those classes that had not previously participated in the study).
8. 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.

The WWC found *Waterford Early Reading Level One™* to have no discernible effects on oral language or on print knowledge (continued)

Findings for comparisons between *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®*

The data for the comparison described below were included in the Fischel et al. (in press) study, but they do not contribute to the overall rating of effectiveness because the WWC included the comparison of *Waterford Early Reading Level One™* with the business-as-usual comparison group in the rating for the same study, which provides the most direct evidence of *Waterford's* effects. However, the WWC believes that the findings from this comparison provide useful information to practitioners who may be interested in comparing the effects of different curricula. The WWC reports the findings for comparisons of *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* here and in Appendices A4.1 and A4.2.

Oral language. Fischel et al. (in press) included data for two measures in this outcome domain. The differences between the *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* groups were not statistically significant for either measure as calculated by the WWC, and the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25). The average improvement index for oral language is -1 percentile point (*Waterford Early Reading Level One™* is the intervention group and *Let's Begin with the Letter People®* is the comparison group), with a range of -2 to +1 percentile points across findings.

Print knowledge. Fischel et al. (in press) included data for six measures in this outcome domain. The difference between the *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* groups was not statistically significant for any of these measures as calculated by the WWC, and the average effect size was neither statistically significant nor large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25). The average improvement index for print knowledge is -3 percentile points (*Waterford Early Reading Level One™* is the intervention group and *Let's Begin with the Letter People®* is the comparison group), with a range of -13 to +2 percentile points across findings.

Summary

The WWC reviewed three studies on *Waterford Early Reading Level One™*. One of these studies met WWC evidence standards; the remaining studies did not meet WWC evidence screens. Based on this study, the WWC found no discernible effects on oral language and print knowledge. Additional findings that were not considered for the rating of effectiveness indicated that *Waterford Early Reading Level One™* and *Let's Begin with the Letter People®* affect children's outcomes similarly in the oral language and print knowledge domains. The evidence presented in this report may change as new research emerges.

References

Met WWC evidence standards

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*.

Did not meet WWC evidence screens

Cope, R., & Cummings, J. (2001). *Evaluation of the Waterford Early Reading Program in Madisonville Consolidated Independent School District*. Huntsville, TX: Sam Houston State University.⁹
Murray-Ward, M. (2000). *El Centrito interim grant report for the period of July 1, 1999 to December 31, 1999*. (Report No. 109). Thousand Oaks: California Lutheran University, Educational Research and Leadership Institute.¹⁰

For more information about specific studies and WWC calculations, please see the [WWC Waterford Early Reading Level One™ Technical Appendices](#).

9. Does not use a strong causal design: the study did not use a comparison group.

10. Incomparable groups: the intervention and comparison groups cannot be considered equivalent at baseline, even with the use of covariates in the analysis.

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. <i>Journal of Literacy Research</i> .
Participants	Twenty-seven classrooms were randomly assigned to one of three groups (<i>Waterford Early Reading Level One™</i> , <i>Let's Begin with the Letter People®</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>Waterford Early Reading Level One™</i> or business-as-usual comparison groups (three <i>Waterford</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>Waterford</i> classes and five business-as-usual comparison classes) and two randomly selected <i>Waterford</i> classrooms from year one participated again. In year three of the study, five new classrooms were assigned to these groups (two <i>Waterford</i> classes and three business-as-usual comparison classes) and two randomly selected <i>Waterford</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>Waterford Early Reading Level One™</i> curriculum, which was overlaid on the existing <i>High/Scope</i> 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 <i>High/Scope</i> framework. 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>Waterford Early Reading Level One™</i> 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 <i>Waterford Early Reading Level One™</i> classroom mid-year to review summary data to assess classroom progress and provide support and additional training. Teachers and assistants in the <i>Waterford Early Reading Level One™</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 *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)

Appendix A1.1 Study characteristics: Fischel, Bracken, Fuchs-Eisenberg, Spira, Katz, & Shaller (in press) (randomized controlled trial) (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 <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
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 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 <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 A3.1 Summary of study findings included in the rating 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 (Waterford Early Reading Level One™ – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			Waterford Early Reading Level One™ group ⁴	Comparison group ⁴				
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 average¹⁰ 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 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 (Waterford Early Reading Level One™ – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			Waterford Early Reading Level One™ group ⁴	Comparison group ⁴				
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 average¹⁰ 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 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>Waterford Early Reading Level One™</i> – <i>Let's Begin with the Letter People®</i>)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Waterford Early Reading Level One™</i> group ⁴	<i>Let's Begin with the Letter People®</i> group ⁴				
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 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 *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 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>Waterford Early Reading Level One™</i> – <i>Let's Begin with the Letter People®</i>)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
		<i>Waterford Early Reading Level One™</i> group ⁴	<i>Let's Begin with the Letter People®</i> group ⁴					
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 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 *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.”