

# Appendix

## Appendix A1.1 Study Characteristics: Macaruso, Hook, & McCabe, 2006

Characteristic	Description
<b>Study citation</b>	Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. <i>Journal of Research in Reading, 29</i> (2), 162–172.
<b>Participants</b>	Study participants were first-graders in 10 classrooms spread across five schools, with two classrooms in each school (one treatment classroom and one comparison classroom) participating in the study. The study initially included 92 intervention and 87 comparison students. Twelve students (9 intervention, 3 comparison) left the study when it was determined that they were eligible for special education services. The analysis sample contained 15 Title I students in each of the intervention and comparison groups (Title I students received an additional 30 minutes of academic instruction per day from a Title I staff member).
<b>Setting</b>	First-grade classrooms in a Massachusetts public school district.
<b>Intervention</b>	<i>Lexia Reading</i> is a computerized, supplementary reading software program designed for regular use, consisting of two to four weekly sessions of 20 to 30 minutes each, in a lab or classroom setting. In the study, intervention students were exposed to two <i>Lexia Reading</i> components: <i>Phonics Based Reading (PBR)</i> and <i>Strategies for Older Students (SOS)</i> . The <i>PBR</i> component has 3 levels, 17 skill activities, and 174 units covering basic phonics skills usually taught in grades 1 through 3. After finishing <i>PBR</i> activities, children were introduced to <i>SOS</i> activities, which consist of 5 levels, 24 skill activities, and 369 discrete units. Intervention classes used <i>Lexia Reading</i> software for approximately six months, with children completing an average of 64 sessions and 140 skill units. Most students worked on <i>PBR</i> activities only; 14 students (17%) in the intervention programs moved on to <i>SOS</i> activities, working mainly on early levels.
<b>Comparison</b>	Students in the comparison group classrooms received regular classroom instruction while intervention group classrooms were participating in the <i>Lexia Reading</i> program.
<b>Primary outcomes and measurement</b>	For both pre- and posttest, the authors used the Gates-MacGintie Reading Test, Level BR to assess reading performance. For a more detailed description of this outcome measure and its subtests, see Appendices A2.1, A2.3, and A2.4.
<b>Staff/teacher training</b>	Teachers in intervention classrooms had an average of 19 years of teaching experience, and teachers in comparison classrooms had an average of 18 years of teaching experience. Teachers in the intervention classrooms and computer lab staff received orientation and training sessions for implementing <i>Lexia Reading</i> software use.

## Appendix A1.2 Study Characteristics: Gale, 2006

Characteristic	Description
<b>Study citation</b>	Gale, D. (2006). <i>The effect of computer-delivered phonological awareness training on the early literacy skills of students identified as at-risk for reading failure</i> . Retrieved from the University of South Florida website: <a href="http://purl.fcla.edu/usf/dc/et/SFE0001531">http://purl.fcla.edu/usf/dc/et/SFE0001531</a> .
<b>Participants</b>	Kindergarten and first-grade students who were identified in the fall assessment period as needing intensive substantial intervention based on their performance on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) were recruited for this study. Forty-one kindergarten students and 38 first-grade students were randomly assigned to one of three groups: (1) <i>Lexia Early Reading</i> , (2) Earobics® Step 1, or (3) control. After attrition, the analysis sample contained 39 kindergarten and 37 first-grade students.
<b>Setting</b>	The elementary school in which this study occurred is located in a large school district in the southwest region of Florida serving approximately 114,500 pre-K to twelfth-grade students. The elementary school had a total kindergarten through fifth-grade student enrollment of 722. Students in the school represented the following ethnic groups: 60% Caucasian, 19% Hispanic, 8% Asian/Pacific Islander, 7% African-American, 5% multiracial, <1% American Indian/Alaskan Native. Approximately three-quarters of the students in this school were eligible for free or reduced-priced lunch.
<b>Intervention</b>	A rotation schedule was developed by the researcher based on teacher input. The two phonological awareness software programs were loaded on 14 numbered computers with headphones in the computer lab at the elementary school. Each student was assigned to a computer to use throughout the intervention period. Before the intervention period began, the researcher trained the participants in small groups of five on the relevant intervention software ( <i>Lexia Early Reading</i> or Earobics® Step 1) with regard to initiating and proceeding through the program and navigating the mouse. Students were required to pass at least five out of six areas on the training checklist as well as the task “use mouse to navigate activity” before beginning the intervention. The students were divided into four groups that alternated in the computer lab according to the rotation schedule. The researcher and a teacher assistant monitored the students each day during their training in the computer lab. Students used their respective computer programs in the school computer lab 20 minutes daily for 25 days, resulting in a total of 8 hours 20 minutes of exposure.
<b>Comparison</b>	The control group received no reading instruction beyond the regular language arts time. Typical reading instruction in the school was a 90-minute reading block.
<b>Primary outcomes and measurement</b>	Students were tested before and after the intervention using the DIBELS subtests for Initial Sounds Fluency (kindergarten only), Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency (first grade only) and Oral Reading Fluency (first grade only). For a more detailed description of these outcome measures, see Appendices A2.1 and A2.2.
<b>Staff/teacher training</b>	No information on teacher training was provided. The <i>Lexia Early Reading</i> group worked in a computer lab, with minimal teacher instruction.

## Appendix A1.3 Study Characteristics: Macaruso & Walker, 2008

Characteristic	Description
<b>Study citation</b>	Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. <i>Reading Psychology, 29</i> (3), 266–287.
<b>Participants</b>	Six kindergarten classes from two elementary schools participated in the study. The six classes included morning and afternoon classes for each of three teachers. The authors randomly assigned the six classes to treatment ( <i>Lexia Early Reading</i> ) or comparison (extra time spent in language-related classroom activities), blocked by teacher. These six classes included a total of 94 students. After randomly assigning classrooms, the authors dropped from the analysis 11 students (9 intervention, 2 comparison) who were designated as English Language Learners or special education. At the end of the study, the authors excluded another 12 students from the treatment group who had not completed their minimum criterion of more than 45 sessions with <i>Lexia Early Reading</i> . The final analysis sample consisted of 26 students in the <i>Lexia Early Reading</i> group and 45 students in the comparison group. The authors demonstrated that there were no statistically significant pre-intervention differences between the two analysis groups on the baseline measures (DIBELS: Initial Sounds Fluency and DIBELS: Letter Naming Fluency).
<b>Setting</b>	The participating schools were two urban elementary schools near Boston, Massachusetts. Twenty-nine percent of families in the school system spoke a language other than English at home, and the median household income in the school district was \$37,000 (compared to a state median of \$50,000). More than half of the students in the district qualified for free or reduced-price lunch.
<b>Intervention</b>	Classes in the intervention condition began using <i>Lexia Early Reading</i> in November and continued for approximately six months. Students used the software in two to three weekly sessions of 15 to 20 minutes each. On average, students in the analysis sample completed 52 sessions with the software. <i>Lexia Early Reading</i> contains nine activities involving sound identification, rhyming, segmenting and blending of sounds, and application of letter-sound correspondences for subsets of consonants and vowels. Each activity consists of several units; students progress to the next activity only after mastering skills in the prior activity.
<b>Comparison</b>	Students in the comparison condition spent extra time engaged in language-related classroom activities.
<b>Primary outcomes and measurement</b>	At the end of the study period, the students were tested using the DIBELS subtests for Letter Naming Fluency and Phoneme Segmentation Fluency, as well as both subtest and overall composite scores for the Gates-MacGintie Reading Test, Level PR. Because the composite score for the Gates-MacGintie measure spans the alphabets and comprehension domains, the subtests results for alphabets and comprehension are presented as the main findings in Appendices A3.1 and A3.3 and the composite score results are presented as supplemental findings in Appendix A4.3. For a more detailed description of these outcome measures, see Appendices A2.1 and A2.3.
<b>Staff/teacher training</b>	Kindergarten teachers and computer lab staff participated in an orientation and training session for <i>Lexia Early Reading</i> software implementation.

## Appendix A2.1 Outcome measures for the alphabetic domain

Characteristic	Description
<b>Phonological Awareness</b>	
<b>Gates-MacGintie Reading Test, Level BR: Letter-Sound Correspondences subtest</b>	Students are required to match letters with their appropriate sounds. It is one of four subtests on the Gates-MacGintie Reading Test, Level BR (as cited in Macaruso, Hook, & McCabe, 2006).
<b>Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Initial Sounds Fluency subtest</b>	Students are presented with four pictures that are named by the examiner. The examiner then asks the student to identify the picture that begins with a sound presented orally by the examiner. The student is also asked to orally provide the initial sound in a word presented orally by the examiner. The score is calculated by totaling the amount of time that it takes the student to identify or produce the correct sounds and converting that time into the number of correct onsets in a minute (as cited in Gale, 2006).
<b>Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Phoneme Segmentation Fluency subtest</b>	In this task, the student is given a word and asked to provide the individual phonemes that make up the word. Words are continuously presented for one minute. The score is calculated by how many phonemes the student correctly segments in one minute (as cited in Gale, 2006).
<b>Gates-MacGintie Reading Test, Level PR: Oral Language Concepts subtest</b>	This subtest requires children to identify pictures with names that begin or end with the same sound or identify pictures that have rhyming names (as cited in Macaruso & Walker, 2008).
<b>Letter Knowledge</b>	
<b>Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Letter Naming Fluency subtest</b>	This task requires the student to orally identify upper- and lowercase letters presented in random order on a piece of paper. The student names as many letters (out of 120) as he or she can in one minute with the examiner providing the name if the student hesitates for three seconds. The score is calculated by the number of correctly named letters in one minute (as cited in Gale, 2006).
<b>Gates-MacGintie Reading Test, Level PR: Letters and Letter-Sound Correspondences subtest</b>	Children identify when two letters match and match letters with pictures that begin with sounds corresponding to the letters (as cited in Macaruso & Walker, 2008).
<b>Phonics</b>	
<b>Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Nonsense Words Fluency subtest</b>	In this measure, the student is presented with randomly ordered vowel-consonant and consonant-vowel-consonant nonsense words on a sheet of paper and asked to produce either the individual sounds or the whole nonsense word. The child has one minute to produce as many letter sounds or words as he or she can (as cited in Gale, 2006).
<b>Print Awareness</b>	
<b>Gates-MacGintie Reading Test, Level PR: Literacy Concepts subtest</b>	This subtest assesses students' basic knowledge of printed text, such as finding the first letter in a word (as cited in Macaruso & Walker, 2008).

## Appendix A2.2 Outcome measures for the fluency domain

Outcome measure	Description
<b>Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Oral Reading Fluency</b>	Oral Reading Fluency is a measure of accuracy and fluency with connected text. Students are presented with a passage calibrated at their grade level and asked to read aloud for one minute. Scoring is based on mispronunciations, omissions, substitutions, and hesitations (as cited in Gale, 2006).

## Appendix A2.3 Outcome measures for the comprehension domain

Outcome measure	Description
<i>Vocabulary Development</i>	
<b>Gates-MacGintie Reading Test, Level BR: Basic Story Words subtest</b>	Children are tested on their ability to recognize words that appear most commonly in written text and do not require decoding. It is one of four subtests on the Gates-MacGintie Reading Test, Level BR (as cited in Macaruso, Hook, & McCabe, 2006).
<i>Reading Comprehension</i>	
<b>Gates-MacGintie Reading Test, Level PR: Listening Comprehension subtest</b>	This subtest asks children to listen to a passage and select a picture that most closely reflects the meaning of the passage (as cited in Macaruso & Walker, 2008).

## Appendix A2.4 Outcome measures for the general reading achievement domain

Outcome measure	Description
<b>Gates-MacGintie Reading Test, Level BR: Form S</b>	This test contains four subtests: (1) letter-sound correspondences for initial consonants and consonant clusters, (2) letter-sound correspondences for final consonants and consonant clusters, (3) letter-sound correspondences for vowels, and (4) recognizing basic story words (as cited in Macaruso, Hook, & McCabe, 2006).

## Appendix A3.1 Summary of study findings included in the rating for the alphabets domain<sup>1</sup>

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Lexia Reading-comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Lexia Reading group	Comparison group				
<b>Gale, 2006<sup>7,8</sup></b>								
<i>Comparison #1: Lexia Early Reading vs. Control</i>								
<b>Construct: Phonological Awareness</b>								
DIBELS: Initial Sounds Fluency	Kindergarten	26	10.07 (5.01)	5.21 (3.00)	4.86	1.14	Statistically significant	+37
DIBELS: Phoneme Segmentation Fluency	Kindergarten	26	1.31 <sup>9</sup> (0.63)	0.00 <sup>10</sup> (0.00)	1.31	2.85	Statistically significant	+50
DIBELS: Phoneme Segmentation Fluency	Grade 1	24	37.66 (13.71)	31.02 (10.57)	6.64	0.52	ns	+20
<b>Construct: Letter Knowledge</b>								
DIBELS: Letter Naming Fluency	Kindergarten	26	16.92 <sup>9</sup> (12.91)	13.08 <sup>10</sup> (10.00)	3.84	0.32	ns	+13
DIBELS: Letter Naming Fluency	Grade 1	24	48.11 (14.33)	38.02 (8.97)	10.09	0.81	ns	+29
<b>Construct: Phonics</b>								
DIBELS: Nonsense Word Fluency	Grade 1	24	40.87 (15.12)	26.11 (11.44)	14.76	1.06	ns	+36
<b>Average for alphabets, Comparison #1 (Gale, 2006)<sup>11</sup></b>						<b>1.12</b>	<b>Statistically significant</b>	<b>+37</b>
<i>Comparison #2: Lexia Early Reading vs. Earobics<sup>®</sup></i>								
<b>Construct: Phonological Awareness</b>								
DIBELS: Initial Sounds Fluency	Kindergarten	26	10.07 (5.01)	13.72 (4.61)	-3.65	-0.73	ns	-27
DIBELS: Phoneme Segmentation Fluency	Kindergarten	26	1.31 <sup>9</sup> (0.63)	1.31 <sup>10</sup> (0.75)	0.00	0.00	ns	0
DIBELS: Phoneme Segmentation Fluency	Grade 1	25	37.66 (13.71)	47.75 (8.08)	-10.09	-0.88	ns	-31

(continued)

**Appendix A3.1 Summary of study findings included in the rating for the alphabets domain<sup>1</sup> (continued)**

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Lexia Reading–comparison)	WWC calculations		
			Lexia Reading group	Comparison group		Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
<b>Construct: Letter Knowledge</b>								
DIBELS: Letter Naming Fluency	Kindergarten	26	18.31 <sup>9</sup> (12.91)	21.08 <sup>10</sup> (11.74)	-2.77	-0.22	ns	-9
DIBELS: Letter Naming Fluency	Grade 1	25	48.11 (14.33)	50.26 (13.83)	-2.15	-0.15	ns	-6
<b>Construct: Phonics</b>								
DIBELS: Nonsense Word Fluency	Grade 1	25	40.87 (15.12)	47.72 (19.65)	-6.85	-0.38	ns	-15
<b>Average for alphabets, Comparison #2 (Gale, 2006)<sup>11</sup></b>						<b>-0.39</b>	<b>ns</b>	<b>-15</b>
<b>Average for alphabets, Entire study (Gale, 2006)<sup>11</sup></b>						<b>0.36</b>	<b>ns</b>	<b>+14</b>
<b>Macaruso &amp; Walker, 2008<sup>7,12</sup></b>								
<b>Construct: Phonological Awareness</b>								
DIBELS: Phoneme Segmentation Fluency	Kindergarten	6/71	28.00 (13.30)	30.90 (19.10)	-2.90	-0.17	ns	-7
Gates-MacGintie: Oral Language Concepts	Kindergarten	6/71	14.80 (4.00)	12.80 (3.50)	2.00	0.54	ns	+20
<b>Construct: Letter Knowledge</b>								
DIBELS: Letter Naming Fluency	Kindergarten	6/71	38.30 (16.90)	38.50 (17.00)	-0.20	-0.01	ns	0
Gates-MacGintie: Letters and Letter-Sound Correspondences	Kindergarten	6/71	24.70 (4.50)	23.70 (5.40)	1.00	0.19	ns	+8
<b>Construct: Print Awareness</b>								
Gates-MacGintie: Literacy Concepts	Kindergarten	6/71	16.80 (2.80)	15.70 (3.00)	1.10	0.37	ns	+14
<b>Average for alphabets (Macaruso &amp; Walker, 2008)<sup>11</sup></b>						<b>0.18<sup>9</sup></b>	<b>ns</b>	<b>+7<sup>9</sup></b>
<b>Domain average for alphabets across all studies<sup>11</sup></b>						<b>0.27</b>	<b>na</b>	<b>+11</b>

ns = not statistically significant

na = not applicable

(continued)

## Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain<sup>1</sup> (continued)

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the alphabetics domain. Subtest and subgroup findings from Macaruso and Walker (2008) are not included in these ratings but are reported in Appendix A4.1.
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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that 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.
7. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Gale (2006), corrections for multiple comparisons were needed, and in the case of Macaruso & Walker (2008), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original studies.
8. Unless otherwise noted, means from this study are posttest means, ANCOVA-adjusted for pretest differences, as reported in Gale (2006).
9. The *Lexia Reading* group mean equals the comparison group mean plus the mean difference. The study author did not provide adjusted means for this outcome, so the WWC calculated the mean difference in outcomes, taking into account the pretest difference between the study groups. For further details, please see Technical Details of WWC-Conducted Computations.
10. Unadjusted posttest mean as reported in Gale (2006).
11. 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 sizes.
12. In this study, the authors did an ANCOVA-adjustment for pretest scores when calculating statistical significance but presented raw means and standard deviations.

## Appendix A3.2 Summary of study findings included in the rating for the fluency domain<sup>1</sup>

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome <sup>2</sup> (standard deviation) <sup>3</sup>		Mean difference <sup>4</sup> (Lexia Reading–comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Lexia Reading group	Comparison group				
<b>Gale, 2006<sup>8</sup></b>								
<i>Comparison #1: Lexia Early Reading vs. Control</i>								
DIBELS: Oral Reading Fluency	Grade 1	24	21.31 (9.65)	13.81 (7.83)	7.50	0.82	ns	+30
<i>Comparison #2: Lexia Early Reading vs. Earobics<sup>®</sup></i>								
DIBELS: Oral Reading Fluency	Grade 1	25	21.31 (9.65)	27.35 (18.53)	–6.04	–0.39	ns	–15
<b>Domain average for fluency<sup>9</sup></b>						<b>0.22</b>	<b>ns</b>	<b>+9</b>

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the fluency domain.
2. Means are posttest means, ANCOVA-adjusted for pretest differences, as reported in Gale (2006).
3. 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.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor 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 and that 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, when 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Gale (2006), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

### Appendix A3.3 Summary of study findings included in the rating for the comprehension domain<sup>1</sup>

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Lexia Reading-comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Lexia Reading group	Comparison group				
<b>Macaruso &amp; Walker, 2008<sup>7,8</sup></b>								
<b>Construct: Reading Comprehension</b>								
Gates-MacGintie: Listening Comprehension	Kindergarten	6/71	13.60 (3.80)	12.60 (3.50)	1.00	0.27	ns	+11
<b>Domain average for comprehension<sup>9</sup></b>						<b>0.27</b>	<b>ns</b>	<b>+11</b>

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain. Subtest and subgroup findings from Macaruso and Walker (2008) are not included in these ratings, but are reported in Appendix A4.2.
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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that 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.
7. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Macaruso and Walker (2008), corrections for clustering and multiple comparisons were needed, so the significance level may differ from that reported in the original study.
8. In this study, the authors did an ANCOVA-adjustment for pretest scores when calculating statistical significance but presented raw means and standard deviations.
9. 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 A3.4 Summary of study findings included in the rating for the general reading achievement domain<sup>1</sup>

Outcome measure	Study sample	Sample size (clusters/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome <sup>2</sup> (standard deviation) <sup>3</sup>		Mean difference <sup>4</sup> (Lexia Reading– comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Lexia Reading group	Comparison group				
<b>Macaruso, Hook, &amp; McCabe, 2006<sup>8</sup></b>								
Gates-MacGintie Reading Test, Level BR: Form S	Grade 1	10/167	63.70 (14.10)	60.40 (14.10)	3.30	0.23	ns	+9
<b>Domain average for general reading achievement<sup>9</sup></b>						<b>0.23</b>	<b>ns</b>	<b>+9</b>

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the general reading achievement domain. Subtest and subgroup findings from Macaruso, Hook, and McCabe (2006) are not included in these ratings, but are reported in Appendix A4.3.
2. Means and standard deviations are ANCOVA-adjusted for pretest differences, as reported in communication with the author.
3. 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.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor 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 and that 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, when 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Macaruso, Hook, and McCabe (2006), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. 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 subscale and subgroup findings for the alphabetics domain<sup>1</sup>

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Lexia Reading</i> –comparison)	WWC calculations		
			<i>Lexia Reading</i> group	Comparison group		Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
<b>Macaruso, Hook, &amp; McCabe, 2006<sup>7,8,9</sup></b>								
Gates-MacGintie Reading Test: Letter-Sound Correspondences	Grade 1: Title I Students	10/30	39.80 (5.50)	34.80 (5.50)	5.00	0.88	Statistically significant	+31
<b>Macaruso &amp; Walker, 2008<sup>8,10</sup></b>								
<b>Construct: Phonological Awareness</b>								
DIBELS: Phoneme Segmentation Fluency	Kindergarten: Low Performers	6/24	29.00 (11.00)	28.00 (21.20)	1.00	0.06	ns	+2
Gates-MacGintie: Oral Language Concepts	Kindergarten: Low Performers	6/24	16.00 (2.20)	12.40 (3.60)	3.60	1.17	Statistically significant	+38
<b>Construct: Letter Knowledge</b>								
DIBELS: Letter Naming Fluency	Kindergarten: Low Performers	6/24	39.20 (12.40)	38.40 (12.70)	0.80	0.06	ns	+2
Gates-MacGintie: Letters and Letter-Sound Correspondences	Kindergarten: Low Performers	6/24	25.60 (2.60)	22.30 (5.40)	3.30	0.75	ns	+27
<b>Construct: Print Awareness</b>								
Gates-MacGintie: Literacy Concepts	Kindergarten: Low Performers	6/24	17.10 (2.50)	15.30 (2.90)	1.80	0.64	ns	+24

ns = not statistically significant

1. This appendix presents subscale and subgroup findings for measures that fall in alphabetics. Total group (for Macaruso & Walker, 2008) and total scale (for Macaruso, Hook, & McCabe, 2006) scores were used for rating purposes and are presented in Appendices A3.1 and A3.4, respectively.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that 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.
7. Means and standard deviations are ANCOVA-adjusted for pretest differences, as reported in communication with the author.

(continued)

## Appendix A4.1 Summary of subscale and subgroup findings for the alphabetics domain<sup>1</sup> *(continued)*

8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the cases of Macaruso, Hook, and McCabe (2006) and Macaruso and Walker (2008), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. The study did not include specific information on the number of clusters (classrooms) across which the sub-sample of Title I students were distributed. The WWC assumed 10 classrooms, the number of classes in the full sample. With any fewer than 10 classrooms, however, this comparison would no longer be statistically significant.
10. In this study, the authors did an ANCOVA-adjustment for pretest scores when calculating statistical significance but presented raw means and standard deviations.

## Appendix A4.2 Summary of subscale and subgroup findings for the comprehension domain<sup>1</sup>

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Lexia Reading-comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Lexia Reading group	Comparison group				
<b>Macaruso, Hook, &amp; McCabe, 2006<sup>7,8,9</sup></b>								
<b>Construct: Vocabulary Development</b>								
Gates-MacGintie: Basic Story Words Subtest	Grade 1: Title I Students	10/30	23.30 (3.50)	21.50 (3.50)	1.80	0.50	ns	+19
<b>Macaruso &amp; Walker, 2008<sup>8,10</sup></b>								
<b>Construct: Reading Comprehension</b>								
Gates-MacGintie: Listening Comprehension	Kindergarten: Low Performers	6/24	13.40 (4.10)	11.50 (3.60)	1.90	0.48	ns	+18

ns = not statistically significant

1. This appendix presents subscale and subgroup findings for measures that fall in comprehension. Total group (for Macaruso & Walker, 2008) and total scale (for Macaruso, Hook, & McCabe, 2006) scores were used for rating purposes and are presented in Appendices A3.1 and A3.4, respectively.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that 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.
7. Means and standard deviations are ANCOVA-adjusted for pretest differences, as reported in communication with the author.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the cases of Macaruso, Hook, and McCabe (2006) and Macaruso and Walker (2008), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. The study did not include specific information on the number of clusters (classrooms) across which the sub-sample of Title I students were distributed. The WWC assumed 10 classrooms, the number of classes in the full sample. With any fewer than 10 classrooms, however, this comparison would no longer be statistically significant.
10. In this study, the authors did an ANCOVA-adjustment for pretest scores when calculating statistical significance but presented raw means and standard deviations.

## Appendix A4.3 Summary of subgroup and composite score findings for the general reading achievement domain<sup>1</sup>

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Lexia Reading-comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Lexia Reading group	Comparison group				
<b>Macaruso, Hook, &amp; McCabe, 2006<sup>7,8,9</sup></b>								
Gates-MacGintie Reading Test, Level BR: Form S	Grade 1: Title I Students	10/30	62.10 (13.70)	49.70 (13.70)	12.40	0.88	Statistically significant	+31
<b>Macaruso &amp; Walker, 2008<sup>8,10</sup></b>								
Gates-MacGintie Reading Test, Level PR	Kindergarten: Full Sample	6/71	54.20 nr	46.40 nr	7.80	0.47	ns	+18
Gates-MacGintie Reading Test, Level PR	Kindergarten: Low Performers	6/24	55.80 nr	41.60 nr	14.20	1.51	Statistically significant	+43

ns = not statistically significant

nr = not reported

1. This appendix presents subgroup and composite score findings for measures that fall in general reading achievement. In the case of Macaruso, Hook, and McCabe (2006), total group scores were used for rating purposes and are presented in Appendix A3.4. In the case of Macaruso and Walker (2008), subtest scores were used for rating purposes and are presented in Appendices A3.1 and A3.3.
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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that 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.
7. Means and standard deviations are ANCOVA-adjusted for pretest differences, as reported in communication with the author.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the cases of Macaruso, Hook, and McCabe (2006) and Macaruso and Walker (2008), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. The study did not include specific information on the number of clusters (classrooms) across which the sub-sample of Title I students were distributed. The WWC assumed 10 classrooms, the number of classes in the full sample. With any fewer than 10 classrooms, however, this comparison would no longer be statistically significant.
10. In this study, the authors did an ANCOVA-adjustment for pretest scores when calculating statistical significance but presented raw means and standard deviations.

## Appendix A5.1 Lexia Reading rating for the alphabetics domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of alphabetics, the WWC rated *Lexia Reading* as potentially positive. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered, as *Lexia Reading* 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 a statistically significant positive effect.

#### 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, and only one 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.** Only one study showed a statistically significant positive effect.

#### AND

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

**Met.** Neither study showed 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. For a complete description, see the WWC Intervention Rating Scheme.

## Appendix A5.2 Lexia Reading rating for the fluency domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup> For the outcome domain of fluency, the WWC rated *Lexia Reading* as having no discernible effects.

### 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.** Only one study examined outcomes in this domain, and the effect was neither statistically significant nor substantively important.

### 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 outcomes in this domain.

### AND

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

**Met.** No study showed a statistically significant or substantively important negative effect.

**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.** Only one study examined outcomes in this domain, and that study did not show a statistically significant or substantively important positive effect.

### 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.** No study showed a statistically significant or substantively important effect, either positive or negative.

**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 study showed a statistically significant or substantively important effect, 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.** No study showed a statistically significant or substantively important effect.

(continued)

## Appendix A5.2 Lexia Reading rating for the fluency domain (continued)

**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.** No study showed a statistically significant or substantively important negative effect.

**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.** No study showed a statistically significant or substantively important positive effect.

**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.** No study showed a statistically significant negative effect.

**AND**

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

**Met.** No study showed a statistically significant or substantively important positive effect.

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. For a complete description, see the WWC Intervention Rating Scheme.

### Appendix A5.3 Lexia Reading rating for the comprehension domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of comprehension, the WWC rated *Lexia Reading* as potentially positive. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered, as *Lexia Reading* 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.** Only one study examined outcomes in this domain, and that study showed a substantively important positive effect.

#### 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.** Only one study examined outcomes in this domain, and that study showed a substantively important positive effect.

#### 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 outcomes in this domain.

#### AND

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

**Met.** Only one study examined outcomes in this domain, and that study showed a substantively important positive effect.

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. For a complete description, see the WWC Intervention Rating Scheme.

## Appendix A5.4 Lexia Reading rating for the general reading achievement 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.<sup>1</sup> For the outcome domain of general reading achievement, the WWC rated *Lexia Reading* as having no discernible effects.

### 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.** Only one study examined outcomes in this domain, and the effect was neither statistically significant nor substantively important.

### 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 outcomes in this domain.

### AND

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

**Met.** No study showed a statistically significant or substantively important negative effect.

**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.** Only one study examined outcomes in this domain, and that study did not show a statistically significant or substantively important positive effect.

### 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.** Only one study examined outcomes in this domain, and that study showed an indeterminate effect.

**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 a statistically significant or substantively important effect, 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.** No studies showed a statistically significant or substantially important effect.

(continued)

## Appendix A5.4 Lexia Reading rating for the general reading achievement domain (continued)

**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.** No study showed a statistically significant or substantively important negative effect.

**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.** No study showed a statistically significant or substantively important positive effect.

**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.** No study showed a statistically significant negative effect.

**AND**

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

**Met.** No study showed a statistically significant or substantively important positive effect.

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. For a complete description, see the WWC Intervention Rating Scheme.

## Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence <sup>1</sup>
		Schools	Students	
Alphabetics	2	3	147	Small
Reading fluency	1	1	37	Small
Comprehension	1	2	71	Small
General reading achievement	1	5	167	Small

1. A rating of “medium 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.”