





## Research *(continued)*

students were randomly assigned to either the treatment or control group.<sup>7</sup> Participants were included in the study based on low English and Spanish reading achievement. The program was used as a supplement to the regular reading program. Individual

findings in the reading achievement domain were statistically significant and substantively important. The English language development domain had no statistically significant or substantively important findings.

## Effectiveness Findings

The WWC review of English language learner interventions addresses student outcomes in three domains: reading achievement, mathematics achievement, and English language development.<sup>8</sup>

*Reading achievement.* Vaughn, Cirino, et al. (2006) found no statistically significant effects for any reading achievement measures. However, five of the seven effect sizes, as well as the average effect size, were large enough to be considered substantively important using WWC criteria. Vaughn, Mathes, et al. (2006) found statistically significant differences favoring the *Enhanced Proactive Reading* students in two of the four reading measures (Word Attack and Passage Comprehension). Further, the average effect size for reading achievement in this study was statistically significant, as calculated by the WWC, and large

enough to be considered substantively important using WWC criteria.

*English language development.* Neither study found a statistically significant or substantively important effect on English language development.

## Rating of effectiveness

The WWC rates interventions 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, the size of the difference between participants in the intervention condition and the comparison condition, and the consistency in findings across studies (see the [WWC Intervention Rating Scheme](#)).

## The WWC found *Enhanced Proactive Reading* to have potentially positive effects on reading achievement and no discernible effects on English language development

### 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 as well as 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 entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results. The average improvement index for the reading achievement domain is +19 percentile points, with a range of +2 to +43 percentile

7. The authors report 10 cases of failed assignment due to scheduling conflicts. These students were replaced with alternates prior to onset of the intervention. There was also differential attrition between the two groups, but the authors were able to provide evidence of pretest equivalence for the post-attrition samples.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate the statistical significance. In the case of *Enhanced Proactive Reading*, a correction for multiple comparisons was needed.





## Appendix A1.2 Study characteristics: Vaughn, Mathes, et al., 2006 (randomized controlled trial with randomization problems<sup>1</sup>)

Characteristic	Description
<b>Study citation</b>	Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., Cardenas-Hagan, E., & Franics, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. <i>Elementary School Journal</i> , 107(2), 153–180. <sup>2</sup>
<b>Participants</b>	The study involved 41 first-grade Hispanic English language learners (50% female) from 14 classrooms. The students were randomly assigned <sup>2</sup> to either the intervention or comparison group. Participants were included in the study based on low English and Spanish reading achievement. All participants were prescreened to assess their English and Spanish reading and language ability. To be included in the study, students had to meet two inclusion criteria: scoring below the 25th percentile for first grade on the Letter Word Identification subtest of the Woodcock Language Proficiency Battery in both Spanish and English and reading between zero and one word from a list of five two- to four-letter words in English and Spanish.
<b>Setting</b>	Four Texas schools with a large population of English language learners served as sites for the study. On average, 98% of the students were Hispanic and more than 80% (ranging from 85% to 100%) qualified for the free or reduced-price lunch program. Schools were located in large urban areas or on an urban border.
<b>Intervention</b>	The intervention group received <i>Enhanced Proactive Reading</i> . The curriculum was implemented as a supplemental reading program for these students. They were taught in small groups of three to five students from October to May, receiving a total of 120, 50-minute lessons. Checklists and observations were conducted and found that the intervention was delivered with acceptable fidelity.
<b>Comparison</b>	The comparison group English language learners did not participate in the supplemental reading program but received the same core reading instruction as the intervention group. The study authors note that 14 of these students received one or more reading interventions, in addition to their core instruction, during the study. More specifically, comparison students received an average of 63.7 hours of supplemental instruction, but there was wide variability, with some receiving relatively few hours and others receiving well over 100 hours (compared with the 80 hours received by students in the intervention group).
<b>Primary outcomes and measurement</b>	Pre- and post-intervention assessments included measures of English language development and reading assessments in both English and Spanish. Only the English language measures are within the scope of this review, and the WWC did not consider Spanish outcomes when assessing the intervention. The reading measures included various subtests from the Woodcock-Johnson battery of assessments, the Dynamic Indicators of Basic Early Literacy Skills, and a measure of students' ability to identify letters in the English alphabet and to provide at least one corresponding sound for each letter. The Woodcock Language Proficiency Battery-Revised served as the English language development measure (composite scores and some related subtests). (See Appendices A2.1 and A2.2 for more detailed descriptions of outcome measures.)
<b>Teacher training</b>	Teachers received 12 hours of professional development training prior to implementation of the intervention and six additional hours six weeks after the start of the intervention. Teachers also participated in frequent staff development sessions and on-site coaching.

1. The authors report 10 cases of failed assignment due to scheduling conflicts, the implication of which is the study meets WWC standards with reservations.

2. This article will be published in the November 2006 issue of *Elementary School Journal*.

## Appendix A2.1 Outcome measures<sup>1</sup> in the reading achievement domain

Outcome measure	Description
<b>Letter Sound Identification</b>	Letter Naming and Letter Sound Identification. Students were asked to identify all 26 letters in the English alphabet and to provide at least one sound for each letter. Raw scores were based on the number of correctly named letters and the number of letters where an appropriate sound was provided. The internal consistency of this measure for first-grade bilingual students ranged from 0.94 to 0.97 (as cited in Vaughn, Cirino et al., 2006, and Vaughn, Mathes et al., 2006).
<b>Word Attack</b>	Woodcock Language Proficiency Battery: Word Attack subtest. This is a standardized measure that assesses the student's phonemic awareness skills. Students read a list of nonsense words. Results are available in raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, and grade-level equivalencies (as cited in Vaughn, Mathes et al., 2006, and Vaughn, Cirino et al., 2006).
<b>Passage Comprehension</b>	Woodcock Language Proficiency Battery: Passage Comprehension subtest. This is a standardized measure. Students silently read a short passage and then fill in the missing word. Scores are available in raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, and grade-level equivalencies (as cited in Vaughn, Cirino et al., 2006, and Vaughn, Mathes et al., 2006).
<b>DIBELS1 &amp; DIBELS2</b>	Dynamic Indicators of Basic Early Literacy Skills: Oral reading fluency & accuracy measure (DIBELS). This measure requires students to read a passage geared at the student's grade level. Children are scored on the number of words read correctly minus the number of words read incorrectly within the time limit. In both studies, two versions of a DIBELS measure were used (passage1 and passage 2). Both passages were designed to be of comparable difficulty (as cited in Vaughn, Cirino et al., 2006, and Vaughn, Mathes et al., 2006 <sup>2</sup> ).
<b>Word Reading Efficiency</b>	Test of Word Reading Efficiency: Word Reading Efficiency subtest. This test requires students to read isolated words of increasing difficulty in 45 seconds. Students are scored on the number of correct words they read during the allotted time period (as cited in Vaughn, Cirino et al., 2006).
<b>Letter-Word Identification</b>	Woodcock Language Proficiency Battery: Letter-Word Identification subtest. <sup>3</sup> This is a standardized measure that assesses the student's phonemic awareness skills. Students first identify a list of letters and then read a list of words. Scores are available in raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, and grade-level equivalencies (as cited in Vaughn, Cirino et al., 2006).

1. Some outcome measures reported by the authors, in both studies, are not reported here because they do not apply to the domain.

2. DIBELS2 outcomes for Vaughn, Mathes et al. (2006) are not reported here. Although posttest scores are available, pretest scores are not. For reasons outlined in Appendix A3.1, the WWC reports gain score effect sizes (that is, pre- to post-change in the treatment group versus the comparison group) for this study.

3. This measure was used for screening purposes (not as an outcome measure) in the Vaughn, Mathes, et al. (2006) study

## Appendix A2.2 Outcome measures in the English language development domain

Outcome measure	Description
<b>English Language Composite</b>	Woodcock Language Proficiency Battery—Revised, Oral Language Composite. This is a standardized measure of oral language proficiency composed by a series of subtests in picture vocabulary, listening comprehension, and verbal analogies (as cited in Vaughn, Cirino et al., 2006, and Vaughn, Mathes et al., 2006).
<b>Picture Vocabulary</b>	Woodcock Language Proficiency Battery: Picture Vocabulary subtest. This is a standardized measure. This subtest is a measure of expressive language skills. Students are asked to name familiar and unfamiliar pictured objects. Scores are available in raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, and grade-level equivalencies (as cited in Vaughn, Cirino et al., 2006, and Vaughn, Mathes et al., 2006).
<b>Listening Comprehension</b>	Woodcock Language Proficiency Battery: Listening Comprehension subtest. This is a standardized measure. Students are asked to listen to a passage and supply the missing word at the end of a statement using oral cloze procedure. Scores are available in raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, and grade-level equivalencies (as cited in Vaughn, Cirino et al., 2006, and Vaughn, Mathes et al., 2006).
<b>Verbal Analogies</b>	Woodcock Language Proficiency Battery: Verbal Analogies subtest. This is a standardized measure. Students answer questions about logical relationships that increase in difficulty. Scores are available in raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, and grade-level equivalencies (as cited in Vaughn, Cirino et al., 2006, and Vaughn, Mathes et al., 2006).



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

Outcome measure <sup>3</sup>	Study sample	Sample size (students/schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (Enhanced Proactive Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Enhanced Proactive Reading group	Comparison group				
<b>Vaughn, Cirino, et al., 2006 (randomized controlled trial)<sup>8</sup></b>								
Letter Sound Identification	Grade 1	90	23.02 (4.1)	21.45 (4.5)	1.57	0.26	ns	+10
Word Attack	Grade 1	78	99.92 (14.3)	94.40 (11.8)	5.52	0.42	ns	+16
Letter Word Identification	Grade 1	91	89.88 (19.1)	87.42 (18.5)	2.46	0.13	ns	+5
Passage Comprehension	Grade 1	90	86.93 (13.0)	86.13 (13.3)	0.80	0.06	ns	+2
DIBELS1	Grade 1	88	17.07 (17.9)	12.28 (11.6)	4.79	0.32	ns	+13
DIBELS2	Grade 1	88	14.66 (14.8)	11.26 (10.5)	3.40	0.27	ns	+10
Word Reading Efficiency	Grade 1	89	16.93 (10.3)	12.83 (9.5)	4.10	0.41	ns	+16
<b>Average<sup>9</sup> for reading achievement (Vaughn, Cirino, et al., 2006)</b>						.27	ns	+10
<b>Vaughn, Mathes, et al., 2006 (randomized controlled trial with randomization problems)<sup>8</sup></b>								
Letter Sound Identification	Grade 1	40	9.53 (2.2)	6.64 (6.9)	2.89	0.57	ns	+22
Word Attack	Grade 1	38	43.71 (12.3)	24.25 (12.4)	19.46	1.53	Statistically significant	+43
Passage Comprehension	Grade 1	39	22.82 (13.8)	4.88 (12.4)	17.94	1.32	Statistically significant	+41
DIBELS1	Grade 1	36	20.49 (14.5)	17.75 (23.8)	2.74	0.14	ns	+6

(continued)

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

Outcome measure <sup>3</sup>	Study sample	Sample size (students/schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (Enhanced Proactive Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Enhanced Proactive Reading group	Comparison group				
Average <sup>9</sup> for reading achievement (Vaughn, Mathes, et al., 2006)					0.89	Statistically significant	+28	
Domain average <sup>9</sup> for reading achievement across all studies					0.49	na	+19	

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
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. Some outcome measures reported by the authors, in both studies, are not reported here because they do not apply to the domain.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Note that this appendix shows two different types of mean outcomes. For the Vaughn, Cirino et al. (2006) study the WWC presents posttest differences between the treatment group and comparison group. For the Vaughn, Mathes et al. (2006) study, the WWC presents change score mean differences (that is, the pre- to post-difference in the treatment group minus the pre- to post-difference of the comparison group). Change scores were used in Vaughn, Mathes et al. (2006) because there were 10 cases of failed random assignment and severe differential attrition between the intervention and comparison groups.
5. For an explanation of the effect size calculation, please see the [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 favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of the two Vaughn et al. studies, a correction for multiple comparisons was needed for findings in the reading achievement domain, so the significance levels may differ from those reported in the original study.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

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

Outcome measure <sup>3</sup>	Study sample	Sample size (students/schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (Enhanced Proactive Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Enhanced Proactive Reading group	Comparison group				
Vaughn, Cirino, et al., 2006 (randomized controlled trial) <sup>8</sup>								
English Language Composite	Grade 1	88	55.66 (16.2)	58.66 (18.8)	-3.00	-0.17	ns	-7
Average <sup>9</sup> for English language development (Vaughn, Cirino, et al., 2006)						-0.17	ns	-7
Vaughn, Mathes, et al., 2006 (randomized controlled trial with randomization problems)								
English Language Composite	Grade 1	39	2.23 (16.9)	-0.19 (18.1)	2.42	0.13	ns	+5
Average <sup>9</sup> for English language development (Vaughn, Mathes, et al., 2006)						0.13	ns	+5
Domain average <sup>9</sup> for English language development across all studies						-0.02	na	-1

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendix A4.
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. Some outcome measures reported by the authors, in both studies, are not reported here because they do not apply to the domain.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Note that this appendix shows two different types of mean outcomes. For the Vaughn, Cirino et al. (2006) study the WWC presents posttest differences between the treatment group and comparison group. For the Vaughn, Mathes et al. (2006) study, the WWC presents change score mean differences (that is, the pre- to post-difference in the treatment group minus the pre- to post-difference of the comparison group). Change scores were used in Vaughn, Mathes et al. (2006) because there were 10 cases of failed random assignment and severe differential attrition between the intervention and comparison groups.
5. For an explanation of the effect size calculation, please see the [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 favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of both Vaughn et al. studies, no corrections for clustering or multiple comparisons were needed for findings in the English language development domain.
9. 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 Summary of subscale findings for the English language development domain<sup>1</sup>

Outcome measure <sup>3</sup>	Study sample	Sample size (students/schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (Enhanced Proactive Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Enhanced Proactive Reading group	Comparison group				
<b>Vaughn, Cirino, et al., 2006 (randomized controlled trial)<sup>8</sup></b>								
WJ: Picture Vocabulary	Grade 1	88	56.43 (18.8)	59.68 (19.7)	-3.25	-0.17	ns	+07
WJ: Listening Comprehension	Grade 1	91	51.37 (17.2)	55.40 (18.9)	-4.30	-0.22	ns	+09
WJ: Verbal Analogies	Grade 1	91	77.42 (12.0)	78.79 (13.5)	-1.37	-0.11	ns	+04
<b>Vaughn, Mathes, et al., 2006 (randomized controlled trial)</b>								
WJ: Picture Vocabulary	Grade 1	39	8.3 (20.6)	4.38 (19.4)	3.92	0.19	ns	+08
WJ: Listening Comprehension	Grade 1	39	-7.53 (18.3)	-2.05 (17.1)	-5/48	-0.30	ns	-0.12
WJ: Verbal Analogies	Grade 1	37	4.91 (11.2)	1.73 (8.0)	3.18	0.31	ns	+12

ns = not statistically significant

1. This appendix presents subscale findings for measures that fall in the English language development domain. Total scale scores were used for rating purposes and are presented in Appendix A3.2.
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. Some outcome measures reported by the authors, in both studies, are not reported here because they do not apply to the domain.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Again, the WWC reported pre- to post-change scores for the Vaughn, Mathes et al. (2006) study.
5. For an explanation of the effect size calculation, please see the [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 favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of both Vaughn et al. studies, no correction for clustering was needed.

## Appendix A5.1 *Enhanced Proactive Reading* rating for the reading achievement domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of reading achievement, the WWC rated *Enhanced Proactive Reading* as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study that reported a statistically significant positive effect. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Enhanced Proactive 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 statistically significant positive effects, and two studies showed substantively important positive effects.

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

**Met.** Neither study showed a statistically significant or substantively important negative effect. The number of studies showing indeterminate effects (zero) was not greater than the number showing substantively important positive effects (two).

### 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.** *Enhanced Proactive Reading* had only one study that met WWC evidence standards for a strong design and showed statistically significant positive effects.

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

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

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

## Appendix A5.2 *Enhanced Proactive Reading* rating for the English language development domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of English language development, the WWC rated *Enhanced Proactive Reading* as having no discernible effects. It did not meet the criteria for positive effects because it had only one study, which had no statistically significant positive effects. Further, it did not meet the criteria for other ratings (potentially positive, mixed, potentially negative, or negative effects), because the study did not show statistically significant or substantively important effects, either positive or negative.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

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

**Met.** Neither study reported 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.** Although one of the two studies met WWC evidence standards for a strong design, no statistically significant positive effects were found.

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

**Met.** Neither study reported statistically significant or substantively important negative effects.

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

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

**Not met.** Neither study reported a statistically significant or substantively important positive effect.

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

**Not met.** Both studies showed indeterminate effects.

**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. 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.** Neither study reported a statistically significant or substantively important effect in this domain.

OR

(continued)

## Appendix A5.2 *Enhanced Proactive Reading* rating for the English language development domain *(continued)*

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

**Not met.** Neither study reported a statistically significant or substantively important effect, 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 WWC analysis found no statistically significant or substantively important negative effects in this domain.

- 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 WWC analysis found no statistically significant or substantively important positive effects in this domain.

**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 is based on a strong design.

**Not met.** Neither of the studies showed statistically significant or substantively important negative effects.

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

**Met.** The WWC analysis found no statistically significant or substantively important positive effects in this domain.

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.