

# Appendix

## Appendix A1.1 Study characteristics: Rouse & Krueger, 2004

| Characteristic                          | Description  |
|---|--|
| <b>Study citation</b>                   | Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a “scientifically based” reading program. <i>Economics of Education Review</i> , 23(4), 323–338.   |
| <b>Participants</b>                     | Groups were formed through a multistep process. Authors first identified an eligible population of students from four schools within one urban school district, focusing on third- to sixth-grade students who scored in the bottom 20% on the state’s standardized reading test administered in the 2001–02 school year. Consent letters were sent to these students’ parents. Principals in the schools were asked to identify students who could not sit through the daily 90- to 100-minute use of <i>Fast ForWord</i> ®, those who had transferred to another school, and those students who might otherwise be unavailable (family away on long trip, for example). The remaining students were randomly assigned to either the treatment or control group, within each grade and school. In all, 237 students in the <i>Fast ForWord</i> ® group and 217 students in the comparison group were included in the analysis sample. |
| <b>Setting</b>                          | The study took place in four schools in an urban district in the northeastern United States. Forty percent of the district’s students were African-American and more than 50% were Hispanic. Almost 70% of students in the district qualified for the free or reduced-price lunch program, and 56% of the district’s students spoke a language other than English at home. The authors describe test scores in these schools as well below average and note that schools in the district adopted a whole-school reform, <i>Success for All</i> .   |
| <b>Intervention</b>                     | <i>Fast ForWord</i> ® was primarily an add-on to regular reading instruction. In three schools, students in the treatment condition were pulled out of their regular classroom instruction for 90–100 minutes of computerized <i>Fast ForWord</i> ® instruction per day and, in one school, they used <i>Fast ForWord</i> ® for that same amount of time before or after school. Each school had to find a way to fit the use of <i>Fast ForWord</i> ® into its unique schedule. In no case were students taken out of <i>Success for All</i> . The study reported students’ outcomes after six to eight weeks of program implementation.  |
| <b>Comparison</b>                       | The control group continued to receive the standard curriculum being used in district schools. Because the <i>Fast ForWord</i> ® students used <i>Fast ForWord</i> ® either during subjects such as math, science, language arts, special subjects (such as art, music, or gym), or homeroom, or—in the case of one school—before or after school, the counterfactual condition for the control group students was mixed.  |
| <b>Primary outcomes and measurement</b> | For both the pretest and posttest, the authors administered the <i>Success for All</i> assessment, the Clinical Evaluation of Language Fundamentals–Third Edition (the receptive portion and the Listening to Paragraph supplemental test), and a state standardized reading test (the authors did not indicate which state). For a more detailed description of test outcome measures, see Appendix A2.4.   |
| <b>Staff/teacher training</b>           | <i>Fast ForWord</i> ® staff provided training for <i>Fast ForWord</i> ® instructors (those interacting with students) at the beginning of the study. Phone support was also provided for the duration of the study. Detailed information on the training of instructors was not provided.  |

## Appendix A1.2 Study characteristics: Scientific Learning Corporation, 2007a

| Characteristic                          | Description  |
|---|--|
| <b>Study citation</b>                   | Scientific Learning Corporation. (2007a). <i>Students in Western Australia improve language and literacy skills: Educator's briefing</i> . Oakland, CA: Author.  |
| <b>Participants</b>                     | Students between ages 5 and 14 identified by classroom teachers as having difficulties in language, literacy, auditory processing, attention, and/or behavior were randomly assigned to immediate or delayed treatment conditions, with 72 students in each group. The intervention group that received <i>Fast ForWord</i> ® either between February and April or May and July of 2006 was compared to the group of students who had not received <i>Fast ForWord</i> ® as of April 2006. In all, 68 students in the <i>Fast ForWord</i> ® group and 69 students in the comparison group were included in the analysis sample.  |
| <b>Setting</b>                          | The study took place at four primary schools in the Perth metropolitan area in Western Australia.  |
| <b>Intervention</b>                     | <i>Fast ForWord</i> ® participation was scheduled during class time for most students, generally in place of their language-arts lesson. A few students participated before school and during recess and/or lunch breaks. All <i>Fast ForWord</i> ® sessions were monitored by trained parent volunteers under the supervision of the school's <i>Fast ForWord</i> ® coordinator. Participants in the <i>Fast ForWord</i> ® group used (1) the 50-minute <i>Fast ForWord</i> ® <i>Language</i> protocol or the 48-minute <i>Fast ForWord</i> ® <i>Middle and High School</i> protocol and (2) the 50-minute <i>Fast ForWord</i> ® <i>Language to Reading</i> protocol. These protocols called for participants to use <i>Fast ForWord</i> ® each day, five days a week, for 8 to 12 weeks. The study reported students' outcomes after three months of program implementation. |
| <b>Comparison</b>                       | The counterfactual in this study is regular classroom instruction. The comparison group used <i>Fast ForWord</i> ® on a delayed schedule, either between May and July or July and September 2006.  |
| <b>Primary outcomes and measurement</b> | All tests were administered by speech pathology and occupational therapy students who were trained in the assessment process by qualified speech pathologists. Study students' skills were measured both before and after use of the intervention. Alphabetic skills were measured by the Queensland University Inventory of Literacy (QUIL), whereas students' skills in comprehension were measured by the Clinical Evaluation of Language Fundamentals (CELF)–Fourth Edition. For a more detailed description of these outcome measures, see Appendices A2.1 and A2.4.  |
| <b>Staff/teacher training</b>           | Sonic Hearing, a private clinical practice with expertise in the <i>Fast ForWord</i> ® programs, provided training for the parent monitors and support for the <i>Fast ForWord</i> ® coordinator at each school. All <i>Fast ForWord</i> ® sessions were monitored by these trained parent volunteers, under the supervision of the school's <i>Fast ForWord</i> ® coordinator. In addition, the lab supervisors at the schools were trained in current and established findings on the neuroscience of how phonemic awareness and the acoustic properties of speech affect development of language and reading skills, information on the efficacy of the products, effective implementation techniques, and monitoring student progress.   |

## Appendix A1.3 Study characteristics: Beattie, 2000

| Characteristic                          | Description   |
|---|---|
| <b>Study citation</b>                   | Beattie, K. K. (2000). The effects of intensive computer-based language intervention on language functioning and reading achievement in language-impaired adolescents (Doctoral dissertation, George Mason University, 2000). <i>Dissertation Abstracts International</i> , 61(08A), 194–3116.  |
| <b>Participants</b>                     | Eighty-one 11- to 16-year-old students who scored in the bottom quartile on standardized reading or language tests were randomly assigned by computer-generated procedures to one of four intervention groups or to a control group in a two-step process. <sup>1</sup> The researchers first assigned 18 students to the two intervention groups that received a phase of <i>SuccessMaker</i> and <i>Fast ForWord</i> ® and also concomitantly participated in a functional resonance imaging research project. Then, the remaining participants were randomly assigned across the five groups. To ensure an equal distribution among groups, fewer students were placed in the first two groups at the second step of randomization. For this review, the WWC reported results from 12 students in the <i>Fast ForWord</i> ® group who were compared to 12 students in the comparison group. <sup>2</sup> Although the overall attrition rate was higher than 20%, the post-attrition intervention and comparison groups were equivalent on the pretest achievement measures. |
| <b>Setting</b>                          | The study took place in two middle schools and one middle-high school located in the suburbs of a large metropolitan area in northern Virginia.   |
| <b>Intervention</b>                     | Students worked on <i>Fast ForWord</i> ® for 90–94 minutes a day, five days a week. The intervention ended after each student completed 64–80 hours on the program. The study reported students' outcomes after two months of program implementation.   |
| <b>Comparison</b>                       | The control group received the standard instruction provided as part of the regular school curriculum.  |
| <b>Primary outcomes and measurement</b> | For both pre- and posttests, the author administered the Gray Oral Reading Test, four subtests of the Woodcock-Johnson Psycho-Educational Battery (Letter-Word Identification, Word Attack, Passage Comprehension, and Auditory Processing), the Spelling subtest of the Wide Range Achievement Test, and the Receptive Language subtest of the Clinical Evaluation of Language Fundamentals. For a more detailed description of these outcome measures, see Appendices A2.1–A2.4.  |
| <b>Staff/teacher training</b>           | No information on training for the teachers and staff in this study was provided. To facilitate the use of <i>Fast ForWord</i> ®, computers were procured or updated to meet criteria for running <i>Fast ForWord</i> ® software.   |

1. The first intervention group received two phases of *Fast ForWord*®; the second intervention group received two phases of *SuccessMaker*; and the third and fourth intervention groups received a phase of *Fast ForWord*® and a phase of *SuccessMaker*.
2. The analysis samples for the *Fast ForWord*® and *SuccessMaker* groups were not shown to be equivalent at baseline. Two other groups which combined *Fast ForWord*® and *SuccessMaker* are not appropriate counterfactuals because the measures of effects cannot be attributed solely to the *Fast ForWord*® program.

## Appendix A1.4 Study characteristics: Borman & Benson, 2006

| Characteristic                          | Description   |
|---|---|
| <b>Study citation</b>                   | Borman, G. D., & Benson, J. (2006). <i>Can brain research and computers improve literacy? A randomized field trial of the Fast ForWord® Language computer-based training program</i> (WCER working paper no. 2006-5). Madison, WI: University of Wisconsin–Madison, School of Education.  |
| <b>Participants<sup>1</sup></b>         | Students were eligible for the study if they scored below national norms on the total reading outcome for the district-administered Comprehensive Test of Basic Skills–Fifth Edition (CTBS/5) during the spring of 2000. These students also tended to have below-average outcomes on language skills. A total of 274 of these academically at-risk seventh-grade students took pretests (CTBS/5) in the spring of 2001. Random assignment was conducted separately within each of seven schools. Of the initial intervention and comparison students, listwise deletion of students with missing pretest or posttest data was conducted. Additionally, 13 students (eight from the treatment group and five from the control group) were dropped from the sample because they were determined to be outliers based on a substantial drop from pre- to posttest. In all, 90 students in the <i>Fast ForWord®</i> group and 98 students in the control group were included in the analysis sample (therefore, overall attrition was 31%). Although differential attrition between the treatment and control groups was 8%, the treatment and control groups were shown to be similar to each other at baseline. The groups primarily consisted of African-American (66.3% of both the intervention and comparison groups) and economically disadvantaged students (73.3% of the intervention group and 84.7% of the comparison group received free lunch). |
| <b>Setting</b>                          | The study took place in seven middle schools in the Baltimore City Public School System.  |
| <b>Intervention</b>                     | In addition to their regular reading instruction, students randomly assigned to the intervention condition used the <i>Fast ForWord® Language</i> software program in school resource rooms. The resource rooms served as a targeted pullout program offered during the regular school day to supplement the regular classroom literacy instruction. Students received the program 100 minutes a day, five days a week, for at least 20 days under the supervision of an <i>Fast ForWord®</i> -trained teacher. The study reported students' outcomes after two months of program implementation.   |
| <b>Comparison</b>                       | In addition to their regular reading instruction, comparison group students received nonliteracy instruction or participated in special activities and classes not related to literacy, such as art and gym.  |
| <b>Primary outcomes and measurement</b> | The eligible outcomes are standardized (normal curve equivalent) CTBS/5 Terra Nova Language and Reading test scores. These tests were administered both before and after the intervention. For a more detailed description of test outcome measures, see Appendix A2.4.   |
| <b>Staff/teacher training</b>           | Before the start of the program, Scientific Learning provided training sessions for teachers operating the <i>Fast ForWord®</i> programs at the schools. No detailed information about these training sessions was provided by the authors.   |

1. In addition to the 188 students included in the analysis sample, the study also included 112 second-grade students who were excluded from the findings in this report because they did not fall in the grade range specified in the protocol.

## Appendix A1.5 Study characteristics: Overbay & Baenen, 2002

| Characteristic                          | Description   |
|---|---|
| <b>Study citation</b>                   | Overbay, A., & Baenen, N. (2002). <i>Fast ForWord® evaluation, 2002–03</i> (Evaluation and Research, report no. 03.24). Raleigh, NC: Wake County Public School System.  |
| <b>Participants</b>                     | During the 2002–03 school year, 616 third- to eighth-grade students received the <i>Fast ForWord®</i> program. Of these, 426 were matched with students from non- <i>Fast ForWord®</i> schools based on race, limited English proficiency status, a special programs code, free and reduced-price lunch status, and reading pretest scores. The remaining 190 were missing either pre- or posttest scores and, therefore, were not included in the matching process. The analysis sample for this review included students in grades 4–8: 355 students in the <i>Fast ForWord®</i> group and 355 in the comparison group. <sup>1</sup> <i>Fast ForWord®</i> was used in 10 elementary, middle, and high schools, and the comparison students were selected from schools that did not use <i>Fast ForWord®</i> . Additional findings reflecting students' outcomes by grade can be found in Appendix A4. |
| <b>Setting</b>                          | The study took place in one school district (10 treatment schools) in Raleigh, North Carolina. <sup>2</sup>   |
| <b>Intervention</b>                     | During the school year, the intervention group used <i>Fast ForWord® Language</i> , <i>Fast ForWord® Language to Reading</i> , and <i>Fast ForWord® Reading</i> . Most of the <i>Fast ForWord®</i> participants (91.4%) used <i>Fast ForWord® Language</i> ; the majority (60%) used more than one level of the program. The 8.6% who did not use <i>Fast ForWord® Language</i> had completed it in 2001–02.  |
| <b>Comparison</b>                       | The counterfactual in this study is regular classroom instruction. However, the study authors note that students in the comparison group may have been exposed to a variety of other programs or interventions that were not controlled for in this study.  |
| <b>Primary outcomes and measurement</b> | For both pre- and posttests, the authors used the End of Grade Reading Subtest. For a more detailed description of this outcome measure, see Appendix A2.3.   |
| <b>Staff/teacher training</b>           | No information about teacher or staff training was provided.  |

1. The study also presented data for students in grade 3, attending a total of six elementary schools, but these students do not fall within the age range of the WWC's Adolescent Literacy reviews, so they are not included in this report.
2. The number of control schools is not available.

## Appendix A1.6 Study characteristics: Scientific Learning Corporation, 2004a

| Characteristic                          | Description   |
|---|---|
| <b>Study citation</b>                   | Scientific Learning Corporation. (2004a). Improved Ohio reading proficiency test scores by students in the Springfield City School District who used <i>Fast ForWord</i> ® products. <i>MAPS for Learning: Educator Reports</i> , 8(8), 1–6.  |
| <b>Participants</b>                     | Fourth-grade students who did not pass the fall 2002 Ohio Proficiency Test from four Title I designated schools were eligible to participate in the study. Each elementary school established its own method of identifying treatment and comparison group students for the study. The comparison group was formed by selecting 50 students with test scores from both fall 2002 and spring 2003 who had no exposure to <i>Fast ForWord</i> ® products. <sup>1</sup> The intervention and comparison groups were shown to be equivalent on the Ohio Reading Proficiency Test pretest scores. In all, 41 students who used the <i>Fast ForWord</i> ® products and 50 students in the comparison group were included in the analysis sample.  |
| <b>Setting</b>                          | The study took place in four elementary schools in the Springfield City School District in Ohio.  |
| <b>Intervention</b>                     | The study used <i>Fast ForWord</i> ® <i>Language</i> , <i>Fast ForWord</i> ® <i>Language to Reading</i> , and <i>Fast ForWord</i> ® <i>to Reading 3</i> products. The <i>Fast ForWord</i> ® <i>Language</i> protocol called for students to use the product for 100 minutes a day, five days a week, for four to eight weeks. The <i>Fast ForWord</i> ® <i>Language to Reading</i> and <i>Fast ForWord</i> ® <i>to Reading 3</i> protocols called for use of the product for 90 minutes a day, five days a week, for four to eight weeks. Students included in the treatment group were required to have used <i>Fast ForWord</i> ® products for 20 or more days. Schools used different implementation models, with some schools having students use the products in the back of the classroom, and other schools sending students to computer labs that served between 7 and 24 students. The study reported students' outcomes after one semester of program implementation. |
| <b>Comparison</b>                       | The study did not describe the comparison condition. Presumably, the comparison group received the regular school curriculum.   |
| <b>Primary outcomes and measurement</b> | The Ohio Reading Proficiency Test (a statewide assessment) was administered in the year of the study, before and after the intervention. For a more detailed description of this outcome measure, see Appendix A2.3.  |
| <b>Staff/teacher training</b>           | At each participating school, educators were trained in current and established neuroscience findings on how phonemic awareness and the acoustic properties of speech affect development of language and reading skills, information on the efficacy of the products, methods for assessment of potential candidates for participation, the selection of appropriate measures for testing and evaluation, effective implementation techniques, approaches for using Progress Tracker reports to monitor student performance, and techniques for measuring the gains students have achieved after they have finished using <i>Fast ForWord</i> ® products.   |

1. The study authors did not provide detailed information on how comparison group students were selected (stating that comparison group students were “pseudo-randomly” selected).

## Appendix A1.7 Study characteristics: Scientific Learning Corporation, 2004b

| Characteristic                          | Description   |
|---|---|
| <b>Study citation</b>                   | Scientific Learning Corporation. (2004b). Improved reading achievement by students in the school district of Philadelphia who used <i>Fast ForWord</i> ® products. <i>MAPS for Learning: Educator Reports</i> , 8(21), 1–6.   |
| <b>Participants</b>                     | Three groups of students in grades 2 to 8 (mainly fourth and fifth graders) participated in <i>Fast ForWord</i> ® supplemental instruction during the 2003–04 school year. Groups 1 and 2 comprised the treatment group for this study. Group 1 used <i>Fast ForWord</i> ® between September and November, and group 2 used <i>Fast ForWord</i> ® between December and February. Group 3 served as the comparison group (and used <i>Fast ForWord</i> ® between March and May). The participating schools determined which students were placed in the three groups. Students were assessed in September and March. In all, 256 students in the <i>Fast ForWord</i> ® treatment group and 37 students in the comparison group were included in the analysis sample. Additional findings reflecting students' outcomes by grade and intervention group (1 versus 2) can be found in Appendix A4. |
| <b>Setting</b>                          | The study took place in 16 schools in the Philadelphia School District in Pennsylvania.   |
| <b>Intervention</b>                     | Students participating in the <i>Fast ForWord</i> ® group used a variety of <i>Fast ForWord</i> ® products. All students used either the <i>Fast ForWord</i> ® <i>Language</i> or <i>Fast ForWord</i> ® <i>Middle and High School</i> product for an average of 25 days. In addition, about half of the students used <i>Fast ForWord</i> ® <i>Language to Reading</i> products (which are part of the <i>Fast ForWord</i> ® <i>Language</i> series), and one-tenth of the students used <i>Fast ForWord</i> ® <i>Reading 3</i> products (which are part of the <i>Fast ForWord</i> ® <i>Reading</i> series). <i>Fast ForWord</i> ® was used as a supplement to the regular reading curriculum. The study reported students' outcomes after three months of program implementation.   |
| <b>Comparison</b>                       | Before March 2004, comparison group students received their regular reading curriculum.   |
| <b>Primary outcomes and measurement</b> | The eligible outcome on this study is the Gates–MacGinitie Reading Test, which was administered both before and after the intervention. For a more detailed description of this outcome measure, see Appendix A2.3.   |
| <b>Staff/teacher training</b>           | Teachers were trained in current and established findings on the neuroscience of how phonemic awareness and acoustic properties of speech impact development of language and reading skills; information on the efficacy of the products; methods for assessment of potential product participants; the selection of appropriate standardized language measures for testing and evaluation; effective implementation techniques; instruction on the product, Progress Tracker, and the reports generated by the product that allow educators and coaches to monitor student performance; and techniques for measuring the progress and gains students achieve after they have finished using the product.   |

## Appendix A1.8 Study characteristics: Scientific Learning Corporation, 2007b

| Characteristic                          | Description   |
|---|---|
| <b>Study citation</b>                   | Scientific Learning Corporation. (2007b). Improved reading skills by students in the South Madison Community School Corporation who used <i>Fast ForWord</i> ® products. <i>MAPS for Learning: Educator Reports</i> , 11(34), 1–7.  |
| <b>Participants</b>                     | Two schools that used <i>Fast ForWord</i> ® during the spring of 2007 selected students in grades 2 to 5 for the study based on their scores on the Measures of Academic Progress (MAP) assessment. To form a comparison group, school personnel individually matched—by grade level and fall and winter scores from the MAP Reading subtest—80 students in the <i>Fast ForWord</i> ® group to 80 students not using <i>Fast ForWord</i> ®. The study sample included 78 treatment and 78 comparison students. The analysis sample for this review included students in grades 4 and 5: 35 students in the <i>Fast ForWord</i> ® group and 35 students in the comparison group. |
| <b>Setting</b>                          | This study took place in East Elementary and Maple Ridge Elementary in the South Madison Community School Corporation of Pendleton, Indiana.  |
| <b>Intervention</b>                     | The intervention groups used <i>Fast ForWord</i> ® <i>Language</i> and <i>Fast ForWord</i> ® <i>Language to Reading</i> products. The South Madison Community School Corporation chose to use the 50-minute <i>Fast ForWord</i> ® protocols, which called for students to use the product for 50 minutes a day, five days per week, for 6 to 10 weeks. The study reported students' outcomes after three months of program implementation.  |
| <b>Comparison</b>                       | The comparison group received the standard district reading curriculum.   |
| <b>Primary outcomes and measurement</b> | The outcomes on this study are students' reading and language scores on the MAP assessment, which was administered both before and after the intervention was used for the study. For a more detailed description of this outcome measure, see Appendices A2.3–A2.4.  |
| <b>Staff/teacher training</b>           | Educators were trained in current and established neuroscience findings on how phonemic awareness and the acoustic properties of speech impact development of language and reading skills, information on the efficacy of the products, methods for assessing potential candidates for participation, the selection of appropriate measures for testing and evaluation, effective implementation techniques, approaches for using Progress Tracker reports to monitor student performance, and techniques for measuring the gains students have achieved after they have finished using <i>Fast ForWord</i> ® products.   |

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

| Outcome measure  | Description  |
|--|--|
| <i>Phonemic awareness</i>  |  |
| <b>Woodcock-Johnson Psycho-Educational Battery–Revised, Tests of Cognitive Ability (WJ-R COG) (Auditory Processing Cluster for Phonemic Awareness)</b> | This composite is a standardized measure of a student’s ability to identify patterns among speech-based auditory stimuli. The score on this composite is derived from scores on three subtests: (1) the Sound Blending subtest measures the ability to synthesize sequences of sounds into whole words, (2) the Incomplete Words subtest measures the ability to identify a word with missing sounds, and (3) the Sound Patterns subtest measures the ability to indicate whether pairs of computer-generated sound sequences are the same or different (as cited in Beattie, 2000). |
| <i>Phonics</i>   |  |
| <b>Woodcock-Johnson Psycho-Educational Battery–Revised, Tests of Achievement (WJ-R ACH) (Letter-Word Identification)</b>                               | This standardized subtest, which assesses students’ ability to identify words and letters, requires students to read aloud isolated letters and real words that range in frequency and difficulty (as cited in Beattie, 2000).   |
| <b>Woodcock-Johnson Psycho-Educational Battery–Revised, Tests of Achievement (WJ-R ACH) (Word Attack)</b>  | This standardized subtest measures phonemic decoding skills by asking students to read “pseudo” words (e.g., plurp, fronkett). Students are aware that the words are not real (as cited in Beattie, 2000).   |
| <b>Wide Range Achievement Test–Third Edition (WRAT-3) (Spelling subtest)</b>   | This standardized subtest is a paper-and-pencil assessment that measures students’ ability to write their names, as well as letters and words from dictation. The dictated letters and words followed either phonetically regular or irregular patterns (as cited in Beattie, 2000).   |
| <b>Queensland University Inventory of Literacy (QUIL)</b>  | The QUIL is a standardized clinical assessment tool for measuring the phonological awareness skills of school-age children as they pertain to literacy. Three of the 10 subtests were administered to all students: Nonword Spelling, Phoneme Segmentation, and Phoneme Manipulation. In addition, students in years 4–7 were administered the Spoonerisms subtest, which assesses students’ metalinguistic phoneme awareness (as cited in Scientific Learning Corporation, 2007a).  |

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

| Outcome measure                                      | Description   |
|--|---|
| <b>Gray Oral Reading Test–Third Edition (GORT-3)</b> | In this standardized test, students are required to read orally a variety of graded passages to measure reading rate, word identification, and comprehension skills. The Passage subtest assesses a combination of rate and accuracy. The Comprehension subtest requires a student to respond to five multiple choice questions following each story. The Oral Reading Quotient reflects a total measure of a student’s oral reading performance and is calculated by combining the Passage and Comprehension scores (as cited in Beattie, 2000). |

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

| Outcome measure   | Description   |
|---|---|
| <b>Reading comprehension and vocabulary development</b>   |   |
| <b>Woodcock-Johnson Psycho-Educational Battery—Revised, Tests of Achievement (WJ-R ACH) (Passage Comprehension)</b> | In this standardized test, comprehension is measured by having students fill in missing words in a short paragraph (e.g., “Woof,” said the _____, biting the hand that fed it.) (as cited in Beattie, 2000).  |
| <b>Ohio Proficiency Test (OPT), Reading subtest</b>   | This statewide assessment is administered to students in 4th, 6th, and 9th grade. The Reading subtest includes multiple choice, short answer, and extended response questions across four subscales: constructing meaning from fiction, examining/extending meaning in fiction, constructing meaning from nonfiction, and examining/extending meaning in nonfiction. The subtest contains two or three fiction or poetry selections and two or three nonfiction selections, which may include pamphlets, instruction booklets, and newspaper and magazine articles. The selections total about 1,200 to 1,500 words. Students may be asked to summarize or retell a story, to interpret vocabulary, or to infer information. Students may also be asked to make predictions, to distinguish facts from opinions, or to fill in a chart or diagram with information from the selection. Word usage, grammar, spelling, and mechanics do not affect scoring, unless the student’s ideas are not clear to the evaluator (as cited in Scientific Learning Corporation, 2004a).  |
| <b>North Carolina End of Grade Test</b>   | The North Carolina End of Grade test measures students’ achievement of the goals and objectives specified in the 2004 North Carolina English Language Arts Standard Course of Study (Content Standards). Reading comprehension is assessed by having students read authentic selections and then answer questions directly related to the selections. Knowledge of vocabulary is assessed indirectly through application and understanding of terms within the context of selections and questions. The authentic selections in the reading tests are chosen to reflect reading for various purposes such as literary experience, gaining information, and performing a task (as cited in Overbay & Baenen, 2002).  |
| <b>Gates–MacGinitie Reading Test (GMRT)</b>   | The GMRT is used to assess a student’s decoding, vocabulary, and passage comprehension skills. <sup>1</sup> The test has two components that independently assess reading vocabulary and comprehension skills. The Vocabulary subtest measures each student’s reading vocabulary by asking the student to choose one word or phrase that means most nearly the same as a presented word. The subtest contains 45 questions. The Comprehension subtest measures each student’s ability to read and understand different types of prose. The subtest contains 11 passages of various lengths and subjects, and 48 questions (as cited in Scientific Learning Corporation, 2004b).   |
| <b>Measures of Academic Progress (MAP), Reading test</b>  | Developed by the Northwest Evaluation Association (NWEA), the MAP are state-aligned computerized adaptive tests that reflect the instructional level of each student and measure growth over time. The MAP is appropriate for students in grades 2 through 10. The untimed assessment typically features between 40 and 50 items. The assessment is usually tailored to the specific needs of individual organizations, but all NWEA MAP assessments draw from the same item bank. The Reading test draws items from the following areas: word meaning (such as use of context clues; use of synonyms, antonyms, and homonyms; use of component structure; or interpretation of multiple meanings), literal comprehension (such as recalling details, interpreting directions, sequencing details, classifying facts, or identifying main ideas), interpretive comprehension (such as drawing inferences, recognizing cause and effect, predicting events, or summarizing and synthesizing), and evaluative comprehension (such as distinguishing fact and opinion, recognizing elements of persuasion, evaluating validity and point of view, evaluating conclusions, or detecting bias and assumptions) (as cited in Scientific Learning Corporation, 2007b). |
| <b>Comprehensive Test of Basic Skills (CTBS/5) Terra Nova Reading Composite</b>                                     | This assessment combines selected-response items with constructed-response items that allow students to produce short and extended responses. The Reading composite score is the average of Reading Comprehension and Vocabulary subtest scores (as cited in TerraNova Prepublication Technical Manual, July 1996). The Reading Comprehension subtest items focus on five objectives: (1) oral comprehension of passages read aloud, (2) basic understanding of literal meanings of passages, (3) analyzing text, (4) evaluating and extending meaning, and (5) identifying reading strategies. The Vocabulary subtest focuses on three objectives: (1) understanding word meaning, (2) identifying multi-meaning words, and (3) inferring words in context (as cited in Borman & Benson, 2006).  |

1. At levels D (4th grade) and up, either subtest or the combination of both subtests falls into the comprehension domain. At levels A, B, and C (grades 1, 2, and 3), the vocabulary measure, which taps decoding skills rather than word meanings, would fall in the alphabetic domain. For the Scientific Learning Corporation (2004b) study, which included students from grades 2–8, the WWC classified the Gates–MacGinitie Reading Test as a comprehension measure, as the majority of study participants came from grades 4 and 5 (levels D and up).

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

| Outcome measure  | Description   |
|--|---|
| <i>General reading achievement</i>   |   |
| <b>Comprehensive Test of Basic Skills (CTBS/5) Terra Nova Language Composite</b>               | This assessment combines selected-response items with constructed-response items that allow students to produce short and extended responses. The Language Composite score is the average of scores on the Language and Language Mechanics subtests (as cited in TerraNova Prepublication Technical Manual, July 1996). The Language subtest covers four objectives: (1) introduction to print, (2) understanding sentence structure, (3) writing strategies, and (4) editing skills. The Language Mechanics subtest focuses on three objectives: (1) appropriate construction of sentences, phrases, and clauses; (2) appropriate writing conventions; and (3) editing skills (as cited in Borman & Benson, 2006).   |
| <b>Success for All assessment</b>  | The <i>Success for All</i> assessments (which are administered every 6–8 weeks) are a set of reading assessments closely aligned with the <i>Success for All</i> curriculum. The total score of the assessments reflects (1) students' scores on a paper-and-pencil assessment and (2) a more subjective assessment (by the evaluator) of the student's class work during the time period. For example, the subjective assessment might evaluate how well children understand the learning objective, how their writing has progressed, and how well they comprehend what is read to them. Therefore, the total score not only reflects students' reading and writing achievement, but it can also reflect educational behaviors and habits (e.g., note taking, direction following, attention and focus). The version of the paper-and-pencil assessment administered to students depends on students' ability level and language proficiency. The assessments are designed to closely match the individual state's assessment in both content and format (as cited in Rouse & Krueger, 2004). |
| <b>State Standardized Reading Test</b>   | This is the state's criterion-referenced standardized test (the study authors did not specify which state). The exam is designed to be aligned with the curriculum standards of the state as well as to parallel critical aspects of the National Assessment of Educational Progress (NAEP). The state administers tests in reading, math, and writing annually (as cited in Rouse & Krueger, 2004).  |
| <b>Clinical Evaluation of Language Fundamentals—Third Edition (CELF-3), Receptive Language</b> | This standardized assessment measures a student's ability to interpret and execute commands of increasing complexity and to understand relationships between words and categories. It addresses sentence structure, concepts and directions, and word classes (as cited in Beattie, 2000). The Receptive Language portion of the assessment includes five components: (1) sentence structure, in which students point to one of four pictures in response to an orally presented stimulus; (2) concepts and directions, in which students identify pictures of geometric shapes in response to orally presented direction; (3) semantic relations, in which students listen to four facts and then select two of four visually presented options; (4) word classes, in which students select two out of three or four orally presented words that go together; and (5) recalling sentences, in which students imitate an orally presented sentence (as cited in Rouse & Krueger, 2004).   |
| <b>CELF-4—Australian Standard Edition, Receptive Language</b>                                  | CELF-4 is a standardized test widely used to measure a student's overall oral language ability. The Receptive Language index is a cumulative measure of students' performance on subtests designed to best probe receptive aspects of language including comprehension and listening. The subtests cover topics such as Concepts & Following Directions, Word Classes, and Sentence Structure (as cited in Scientific Learning Corporation, 2007a).   |
| <b>CELF-4—Australian Standard Edition, Expressive Language</b>                                 | CELF-4 is a standardized test widely used to measure a student's overall oral language ability. The Expressive Language index is a cumulative measure of students' performance on subtests that probe expressive aspects of language including oral language expression. The subtests cover topics such as Word Structure, Recalling Sentences, and Formulated Sentences (as cited in Scientific Learning Corporation, 2007a).  |
| <b>Measures of Academic Progress (MAP), Language Test</b>                                      | Developed by the Northwest Evaluation Association, the MAP are state-aligned computerized adaptive tests that reflect the instructional level of each student and measure growth over time. The MAP is appropriate for students in grades 2 through 10. The untimed assessment typically features between 40 and 50 items. The assessment is usually tailored to the specific needs of individual organizations, but all NWEA MAP assessments draw from the same item bank. The Language Test draws items from the following areas: writing process, composition structure, grammar/usage, punctuation, and capitalization (as cited in Scientific Learning Corporation, 2007b).  |

## Appendix A3.1 Summary of study findings for all domains<sup>1</sup>

|   | Domain                        |                                     |                                     |                               |
|---|-------------------------------|-------------------------------------|-------------------------------------|-------------------------------|
|   | Alphabetics                   | Reading fluency                     | Comprehension                       | General literacy achievement  |
| <b>Meets WWC evidence standards</b>                   |                               |                                     |                                     |                               |
| Rouse & Krueger (2004)                                | nr                            | nr                                  | nr                                  | ind                           |
| Scientific Learning Corporation (2007a)               | ind                           | nr                                  | nr                                  | ind                           |
| <b>Meets WWC evidence standards with reservations</b> |                               |                                     |                                     |                               |
| Beattie (2000)  | ind                           | (+)                                 | (+)                                 | ind                           |
| Borman & Benson (2006)                                | nr                            | nr                                  | ind                                 | ind                           |
| Overbay & Baenen (2002)                               | nr                            | nr                                  | ind                                 | nr                            |
| Scientific Learning Corporation (2004a)               | nr                            | nr                                  | (+)                                 | nr                            |
| Scientific Learning Corporation (2004b)               | nr                            | nr                                  | +                                   | nr                            |
| Scientific Learning Corporation (2007b)               | nr                            | nr                                  | ind                                 | ind                           |
| <b>Rating of effectiveness</b>                        | <b>No discernible effects</b> | <b>Potentially positive effects</b> | <b>Potentially positive effects</b> | <b>No discernible effects</b> |

nr = no reported outcomes under this domain

+ = study finding was positive and statistically significant

(+) = study finding was positive and substantively important, but not statistically significant

ind = study finding was indeterminate; that is, neither substantively important nor statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices in each domain. More detailed information on findings for all measures within the domains and the constructs that factor into the domains can be found in Appendices A3.2–A3.5.

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

| Outcome measure   | Study sample <sup>4</sup> | Sample size (students) | Authors' findings from the study                            |                  | WWC calculations  |                          |   |                                |
|---|---------------------------|------------------------|---|------------------|---|--------------------------|---|--------------------------------|
|   |                           |                        | Mean outcome <sup>2</sup> (standard deviation) <sup>3</sup> |                  | Mean difference <sup>5</sup> ( <i>Fast ForWord</i> <sup>®</sup> – comparison) | Effect size <sup>6</sup> | Statistical significance <sup>7</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>8</sup> |
|   |                           |                        | <i>Fast ForWord</i> <sup>®</sup> group                      | Comparison group |   |                          |   |                                |
| <b>Scientific Learning Corporation, 2007a<sup>8,9</sup></b>                         |                           |                        |   |                  |   |                          |   |                                |
| Queensland University Inventory of Literacy (QUIL)                                  | Ages 5–14                 | 137                    | 8.49<br>(2.31)  | 7.93<br>(2.58)   | 0.56  | 0.23                     | ns  | +9                             |
| <b>Average for alphabetics (Scientific Learning Corporation, 2007a)<sup>9</sup></b> |                           |                        |   |                  |   | <b>0.23</b>              | <b>ns</b>   | <b>+9</b>                      |
| <b>Beattie, 2000<sup>10</sup></b>   |                           |                        |   |                  |   |                          |   |                                |
| WJ-R ACH (Letter-Word Identification)   | Ages 12–17                | 24                     | 90.99<br>(21.29)  | 92.08<br>(13.15) | –1.09   | –0.06                    | ns  | –2                             |
| WJ-R ACH (Word Attack)  | Ages 12–17                | 24                     | 86.41<br>(14.34)  | 85.91<br>(12.87) | 0.50  | 0.04                     | ns  | +1                             |
| WJ-R COG (Auditory Processing Cluster for Phonemic Awareness)                       | Ages 12–17                | 24                     | 82.58<br>(14.14)  | 85.66<br>(15.61) | –3.08   | –0.20                    | ns  | –8                             |
| Wide Range Achievement Test–Third Edition (WRAT-3) (Spelling subtest)               | Ages 12–17                | 24                     | 82.58<br>(15.10)  | 85.66<br>(13.13) | –3.08   | –0.21                    | ns  | –8                             |
| <b>Average for alphabetics (Beattie, 2000)<sup>11</sup></b>                         |                           |                        |   |                  |   | <b>–0.11</b>             | <b>ns</b>   | <b>–4</b>                      |
| <b>Domain average for alphabetics across all studies<sup>9</sup></b>                |                           |                        |   |                  |   | <b>0.06</b>              | <b>na</b>   | <b>+2</b>                      |

ns = not statistically significant

na = not applicable

WJ-R = Woodcock-Johnson Psycho-Educational Battery–Revised

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the alphabetics domain.
2. The intervention group values are the comparison group means plus the difference in means gains between the intervention and comparison groups.
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. The Adolescent Literacy topic area reviews studies of interventions administered to students in grades 4–12 (or 9–18 years of age). For studies that include samples of students that span both the Adolescent Literacy (grades 4–12) and Beginning Reading (grades K–3) topic areas and cannot be disaggregated by grade level, the Adolescent Literacy topic area reviews any studies that include 5th-grade students or higher. For example, this appendix includes a combined sample of students aged 5–14 years (Scientific Learning Corporation, 2007a).
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 WWC Procedures and Standards Handbook, Appendix B.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.

(continued)

## Appendix A3.2 Summary of study findings included in the rating for the alphabetics domain<sup>1</sup> *(continued)*

8. 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 for the intervention group.
9. Results for the early elementary school students (in 3rd grade or below) in this study are traditionally considered under the Beginning Reading topic area reviews; however, because there was no separate analysis for students in 3rd grade or below (grades covered by the Beginning Reading topic area) and 4th grade and above (areas covered by the Adolescent Literacy topic area), we report on the total sample of students here.
10. 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 the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Beattie (2000), a correction for multiple comparisons was needed, so the significance level may differ from that reported in the original study. In the case of Scientific Learning Corporation (2007a), no corrections for clustering or multiple comparisons were needed.
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.

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

| Outcome measure  | Study sample | Sample size (students) | Author's findings from the study                            |                  | WWC calculations  |                          |   |                                |
|--|--------------|------------------------|---|------------------|---|--------------------------|---|--------------------------------|
|  |              |                        | Mean outcome <sup>2</sup> (standard deviation) <sup>3</sup> |                  | Mean difference <sup>4</sup> ( <i>Fast ForWord</i> <sup>®</sup> – comparison) | Effect size <sup>5</sup> | Statistical significance <sup>6</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>7</sup> |
|  |              |                        | <i>Fast ForWord</i> <sup>®</sup> group                      | Comparison group |   |                          |   |                                |
| <b>Beattie, 2000<sup>8</sup></b>                               |              |                        |   |                  |   |                          |   |                                |
| Gray Oral Reading Test–Third Edition (GORT-3)                  | Ages 12–17   | 24                     | 87.39<br>(16.47)  | 79.50<br>(17.74) | 7.89  | 0.44                     | ns  | +17                            |
| <b>Average for reading fluency (Beattie, 2000)<sup>9</sup></b> |              |                        |   |                  |   | <b>0.44</b>              | <b>ns</b>   | <b>+17</b>                     |

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the reading fluency domain.
2. The intervention group values are the comparison group means plus the difference in means gains between the intervention and comparison groups.
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 WWC Procedures and Standards Handbook, Appendix B.
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 for 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 the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Beattie (2000), no corrections for clustering and multiple comparisons were needed.
9. This row provides the study average, which in this instance is also the domain average. The domain improvement index is calculated from the average effect size.

**Appendix A3.4 Summary of study findings included in the rating for the comprehension domain<sup>1</sup>**

| Outcome measure   | Study sample <sup>4</sup> | Sample size (students) | Authors' findings from the study                            |                   | WWC calculations  |                          |   |                                |
|---|---------------------------|------------------------|---|-------------------|---|--------------------------|---|--------------------------------|
|   |                           |                        | Mean outcome <sup>2</sup> (standard deviation) <sup>3</sup> |                   | Mean difference <sup>5</sup> ( <i>Fast ForWord</i> <sup>®</sup> – comparison) | Effect size <sup>6</sup> | Statistical significance <sup>7</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>8</sup> |
|   |                           |                        | <i>Fast ForWord</i> <sup>®</sup> group                      | Comparison group  |   |                          |   |                                |
| <b>Beattie, 2000<sup>9</sup></b>  |                           |                        |   |                   |   |                          |   |                                |
| WJ-R ACH Passage Comprehension  | Ages 12–17                | 24                     | 97.17<br>(13.78)  | 93.25<br>(11.30)  | 3.92  | 0.30                     | ns  | +12                            |
| <b>Average for comprehension (Beattie, 2000)<sup>10</sup></b>                         |                           |                        |   |                   |   | <b>0.30</b>              | <b>ns</b>   | <b>+12</b>                     |
| <b>Borman &amp; Benson, 2006<sup>8</sup></b>  |                           |                        |   |                   |   |                          |   |                                |
| CTBS/5 Terra Nova Reading NCE Scores  | Grade 7                   | 188                    | 36.99<br>(14.11)  | 34.03<br>(14.92)  | 2.96  | 0.20                     | ns  | +8                             |
| <b>Average for comprehension (Borman &amp; Benson, 2006)<sup>9</sup></b>              |                           |                        |   |                   |   | <b>0.20</b>              | <b>ns</b>   | <b>+8</b>                      |
| <b>Overbay &amp; Baenen, 2002<sup>8,11</sup></b>                                      |                           |                        |   |                   |   |                          |   |                                |
| North Carolina End of Grade   | Grades 4–8                | 710                    | 154.37  | 155.37            | –1.00   | –0.14                    | ns  | –6                             |
| <b>Average for comprehension (Overbay &amp; Baenen, 2002)<sup>9</sup></b>             |                           |                        |   |                   |   | <b>–0.14</b>             | <b>ns</b>   | <b>–6</b>                      |
| <b>Scientific Learning Corporation, 2004a<sup>8</sup></b>                             |                           |                        |   |                   |   |                          |   |                                |
| Ohio Proficiency Test, Reading Test   | Grade 4                   | 91                     | 210.60<br>(16.65)   | 205.10<br>(16.97) | 5.50  | 0.32                     | ns  | +13                            |
| <b>Average for comprehension (Scientific Learning Corporation, 2004a)<sup>9</sup></b> |                           |                        |   |                   |   | <b>0.32</b>              | <b>ns</b>   | <b>+13</b>                     |
| <b>Scientific Learning Corporation, 2004b<sup>8,12</sup></b>                          |                           |                        |   |                   |   |                          |   |                                |
| Gates–MacGinitie Reading Test   | Grades 2–8                | 293                    | 30.39<br>(14.37)  | 25.00<br>(10.60)  | 5.39  | 0.39                     | Statistically significant                                   | +15                            |
| <b>Average for comprehension (Scientific Learning Corporation, 2004b)<sup>9</sup></b> |                           |                        |   |                   |   | <b>0.39</b>              | <b>Statistically significant</b>                            | <b>+15</b>                     |
| <b>Scientific Learning Corporation, 2007b<sup>8,13</sup></b>                          |                           |                        |   |                   |   |                          |   |                                |
| Measures of Academic Progress, Reading Test   | Grades 4–5                | 70                     | 34.90<br>(32.00)  | 30.50<br>(30.10)  | 4.40  | 0.14                     | ns  | +6                             |
| <b>Average for comprehension (Scientific Learning Corporation, 2007b)<sup>9</sup></b> |                           |                        |   |                   |   | <b>0.14</b>              | <b>ns</b>   | <b>+6</b>                      |
| <b>Domain average for comprehension across all studies<sup>9</sup></b>                |                           |                        |   |                   |   | <b>0.20</b>              | <b>na</b>   | <b>+8</b>                      |

(continued)

## Appendix A3.4 Summary of study findings included in the rating for the comprehension domain<sup>1</sup> (continued)

ns = not statistically significant

na = not applicable

WJ-R ACH = Woodcock-Johnson Psycho-Educational Battery–Revised, Tests of Achievement

CTBS/5 = Comprehensive Test of Basic Skills

NCE = Normal Curve Equivalent

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendix A4.
2. The intervention and control group values for Scientific Learning Corporation (2007b) are the ANCOVA adjusted mean values calculated using pretest scores as the covariates. For all other studies in this domain, the intervention group values are the comparison group means plus the difference in means gains between the intervention and comparison groups.
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. The Adolescent Literacy topic area reviews studies of interventions administered to students in grades 4–12 (or 9–18 years of age). For studies that include samples of students that span both the Adolescent Literacy (grades 4–12) and Beginning Reading (grades K–3) topic areas and cannot be disaggregated by grade level, the Adolescent Literacy topic area reviews any studies that include 5th-grade students or higher. For example, this appendix includes a combined sample of students from grades 2–8 (Scientific Learning Corporation, 2004b).
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 WWC Procedures and Standards Handbook, Appendix B.
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 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 for the intervention group.
9. 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 the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. For the *Fast ForWord*<sup>®</sup> studies summarized here, no corrections for clustering or multiple comparisons were needed.
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.
11. This study reported the mean values for the outcome measure, not the standard deviations. The effect size for each grade was calculated through the F-statistics from the one way ANOVA reported in the study. The average effect size reported here is based on effect sizes that have been weighted by the sample size for each grade.
12. The means and standard deviations were aggregated across two intervention groups.
13. This study separately reported results for students in grades 3 and below and for students in grades 4 and above, along with aggregated results across all of the grade levels. Results for the second- and third-grade students in this study will be considered under the Beginning Reading topic area reviews.

**Appendix A3.5 Summary of study findings included in the rating for the general literacy achievement domain<sup>1</sup>**

| Outcome measure  | Study sample <sup>4</sup> | Sample size (students)                 | Authors' findings from the study                            |                  | WWC calculations  |                          |   |                                |
|--|---------------------------|--|---|------------------|---|--------------------------|---|--------------------------------|
|  |                           |  | Mean outcome <sup>2</sup> (standard deviation) <sup>3</sup> |                  | Mean difference <sup>5</sup> ( <i>Fast ForWord</i> <sup>®</sup> – comparison) | Effect size <sup>6</sup> | Statistical significance <sup>7</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>8</sup> |
|  |                           | <i>Fast ForWord</i> <sup>®</sup> group | Comparison group  |                  |   |                          |   |                                |
| <b>Rouse &amp; Krueger, 2004<sup>8</sup></b>   |                           |  |   |                  |   |                          |   |                                |
| Success for All Assessment   | Grades 3–6                | 373                                    | 4.06<br>(1.40)  | 4.03<br>(1.33)   | 0.03  | 0.02                     | ns  | +1                             |
| CELF-3, Receptive Language   | Grades 3–6                | 86                                     | 31.70<br>(18.43)  | 31.01<br>(16.59) | 0.69  | 0.04                     | ns  | +2                             |
| State Standardized Reading Test  | Grades 3–6                | 454                                    | 44.18<br>(24.79)  | 43.03<br>(24.01) | 1.15  | 0.05                     | ns  | +2                             |
| <b>Average for general literacy achievement (Rouse &amp; Krueger, 2004)<sup>9</sup></b>              |                           |  |   |                  |   | <b>0.04</b>              | <b>ns</b>   | <b>+1</b>                      |
| <b>Scientific Learning Corporation, 2007a<sup>8,9</sup></b>  |                           |  |   |                  |   |                          |   |                                |
| CELF-4, Receptive Language   | Ages 5–14                 | 137                                    | 91.00<br>(12.40)  | 88.40<br>(14.12) | 2.60  | 0.19                     | ns  | +8                             |
| CELF-4, Expressive Language  | Ages 5–14                 | 137                                    | 88.00<br>(12.40)  | 85.00<br>(13.29) | 3.00  | 0.23                     | ns  | +9                             |
| <b>Average for general literacy achievement (Scientific Learning Corporation, 2007a)<sup>9</sup></b> |                           |  |   |                  |   | <b>0.21</b>              | <b>ns</b>   | <b>+8</b>                      |
| <b>Beattie, 2000<sup>8</sup></b>   |                           |  |   |                  |   |                          |   |                                |
| CELF-3, Receptive Language   | Ages 12–17                | 24                                     | 86.08<br>(21.11)  | 86.83<br>(22.74) | –0.75   | –0.03                    | ns  | –1                             |
| <b>Average for general literacy achievement (Beattie, 2000)<sup>9</sup></b>                          |                           |  |   |                  |   | <b>–0.03</b>             | <b>ns</b>   | <b>–1</b>                      |
| <b>Borman &amp; Benson, 2006<sup>10</sup></b>  |                           |  |   |                  |   |                          |   |                                |
| CTBS/5 Terra Nova Language NCE Scores  | Grade 7                   | 188                                    | 40.52<br>(11.22)  | 40.14<br>(11.59) | 0.38  | 0.03                     | ns  | +1                             |
| <b>Average for general literacy achievement (Borman &amp; Benson, 2006)<sup>11</sup></b>             |                           |  |   |                  |   | <b>0.03</b>              | <b>ns</b>   | <b>+1</b>                      |
| <b>Scientific Learning Corporation, 2007b<sup>8,12</sup></b>   |                           |  |   |                  |   |                          |   |                                |
| Measures of Academic Progress, Language Test   | Grades 4–5                | 70                                     | 31.10<br>(27.90)  | 26.80<br>(26.60) | 4.30  | 0.16                     | ns  | +6                             |
| <b>Average for general literacy achievement (Scientific Learning Corporation, 2007b)<sup>9</sup></b> |                           |  |   |                  |   | <b>0.16</b>              | <b>ns</b>   | <b>+6</b>                      |
| <b>Domain average for general literacy achievement across all studies<sup>9</sup></b>                |                           |  |   |                  |   | <b>0.08</b>              | <b>na</b>   | <b>+3</b>                      |

(continued)

## Appendix A3.5 Summary of study findings included in the rating for the general literacy achievement domain<sup>1</sup> (continued)

ns = not statistically significant

na = not applicable

CELF-3 = Clinical Evaluation of Language Fundamentals–Third Edition

CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition

CTBS/5 = Comprehensive Test of Basic Skills

NCE = Normal Curve Equivalent

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the general literacy achievement domain.
2. The intervention group values are the comparison group means plus the difference in means gains between the intervention and comparison groups tested immediately after the intervention for Beattie (2000) and Scientific Learning Corporation (2007a). The intervention group values are the comparison group means plus the regression-adjusted impacts for Borman and Benson (2006) and Rouse and Krueger (2004). The intervention and control group values for Scientific Learning Corporation (2007b) are the ANCOVA adjusted mean values calculated using pretest scores as the covariates.
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. The Adolescent Literacy topic area reviews studies of interventions administered to students in grades 4–12 (or 9–18 years of age). For studies that include samples of students that span both the Adolescent Literacy (grades 4–12) and Beginning Reading (grades K–3) topic areas and cannot be disaggregated by grade level, the Adolescent Literacy topic area reviews any studies that include 5th-grade students or higher. For example, this appendix includes a combined sample of students from grades 3–6 (Rouse & Krueger, 2004) and students aged 5–14 years (Scientific Learning Corporation, 2007a).
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 WWC Procedures and Standards Handbook, Appendix B.
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 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 for the intervention group.
9. Results for the early elementary school students (in grades 3 and below) in this study are traditionally considered under the Beginning Reading topic area reviews; however, because there was no separate analysis for students in 3rd grade or below (grades covered by the Beginning Reading topic area) and 4th grade and above (areas covered by the Adolescent Literacy topic area), we report on the total sample of students here.
10. 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 the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. For all *Fast ForWord*® studies summarized here, except Beattie (2000) and Borman & Benson (2006), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
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. This study separately reported results for students in grades 3 and below and for students in grades 4 and above, along with aggregated results across all of the grade levels. Results for the 2nd- and 3rd-grade students in this study will be considered under the Beginning Reading topic area reviews.

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

| Outcome measure   | Study sample <sup>4</sup>          | Sample size (students) | Authors' findings from the study                            |                  | WWC calculations  |                          |   |                                |
|---|------------------------------------|------------------------|---|------------------|---|--------------------------|---|--------------------------------|
|   |                                    |                        | Mean outcome <sup>2</sup> (standard deviation) <sup>3</sup> |                  | Mean difference <sup>5</sup> ( <i>Fast ForWord</i> <sup>®</sup> – comparison) | Effect size <sup>6</sup> | Statistical significance <sup>7</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>8</sup> |
|   |                                    |                        | <i>Fast ForWord</i> <sup>®</sup> group                      | Comparison group |   |                          |   |                                |
| <b>Overbay &amp; Baenen, 2002<sup>9</sup></b>                 |                                    |                        |   |                  |   |                          |   |                                |
| North Carolina<br>End of Grade Test                           | Grade 4                            | 114                    | 148.39  | 150.90           | –2.51   | –0.35                    | ns  | –14                            |
| North Carolina<br>End of Grade Test                           | Grade 5                            | 148                    | 156.07  | 155.76           | 0.31  | 0.06                     | ns  | +2                             |
| North Carolina<br>End of Grade Test                           | Grade 6                            | 78                     | 149.80  | 151.59           | –1.79   | –0.23                    | ns  | –9                             |
| North Carolina<br>End of Grade Test                           | Grade 7                            | 224                    | 155.70  | 156.86           | –1.16   | –0.18                    | ns  | –7                             |
| North Carolina<br>End of Grade Test                           | Grade 8                            | 146                    | 157.70  | 158.18           | –0.48   | –0.08                    | ns  | –3                             |
| <b>Scientific Learning Corporation, 2004b<sup>10,11</sup></b> |                                    |                        |   |                  |   |                          |   |                                |
| Gates–MacGinitie<br>Reading Test                              | Group 1 vs. control;<br>grades 2–8 | 162                    | 30.70<br>(13.90)  | 25.00<br>(10.60) | 5.70  | 0.43                     | Statistically significant                                   | +17                            |
| Gates–MacGinitie<br>Reading Test                              | Group 2 vs. control;<br>grades 2–8 | 168                    | 30.10<br>(14.80)  | 25.00<br>(10.60) | 5.10  | 0.36                     | ns  | +14                            |
| Gates–MacGinitie<br>Reading Test                              | Group 1 vs. control;<br>grade 4    | 56                     | 26.90<br>(12.80)  | 23.20<br>(10.20) | 3.70  | 0.31                     | ns  | +12                            |
| Gates–MacGinitie<br>Reading Test                              | Group 2 vs. control;<br>grade 4    | 67                     | 27.00<br>(15.20)  | 23.20<br>(10.20) | 3.80  | 0.28                     | ns  | +11                            |
| Gates–MacGinitie<br>Reading Test                              | Group 1 vs. control;<br>grade 5    | 103                    | 35.40<br>(14.10)  | 30.60<br>(9.10)  | 4.80  | 0.34                     | ns  | +14                            |
| Gates–MacGinitie<br>Reading Test                              | Group 2 vs. control;<br>grade 5    | 83                     | 34.90<br>(13.40)  | 30.60<br>(9.10)  | 4.30  | 0.33                     | ns  | +13                            |

(continued)

## Appendix A4 Summary of subgroup findings for the comprehension domain<sup>1</sup> (continued)

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the comprehension domain. Total group scores were used for rating purposes and are presented in Appendix A3.4.
2. The intervention group values are the comparison group means plus the difference in means gains between the intervention and comparison groups.
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. The Adolescent Literacy topic area reviews studies of interventions administered to students in grades 4–12 (or 9–18 years of age). For studies that include samples of students that span both the Adolescent Literacy (grades 4–12) and Beginning Reading (grades K–3) topic areas and cannot be disaggregated by grade level, the Adolescent Literacy topic area reviews any studies that include 5th-grade students or higher. For example, this appendix includes a combined sample of students from grades 2–8 (Scientific Learning Corporation, 2004b).
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 WWC Procedures and Standards Handbook, Appendix B.
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 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.
9. This study reported the mean values only for the outcome measure, not the standard deviations. The effect size was calculated through the F-statistics from the one way ANOVA reported in the study.
10. Treatment group 1 received the intervention from September to November, and treatment group 2 received the intervention from December to February.
11. 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 the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. For the *Fast ForWord*<sup>®</sup> studies summarized here, no corrections for clustering were needed.

## Appendix A5.1 *Fast ForWord*<sup>®</sup> rating for the alphabetic 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 alphabetic, the WWC rated *Fast ForWord*<sup>®</sup> as having no discernible effects for adolescent learners. 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: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** No studies showed a statistically significant or substantively important effect, either positive or negative. Two studies 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.** No studies showed a statistically significant positive effect.

**AND**

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

**Met.** No studies 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.** No studies showed 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.

**Not met.** No studies showed a statistically significant or substantively important negative effect. Two studies showed indeterminate effects, and no studies showed a statistically significant or substantively important positive 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 positive effect, and no studies showed a statistically significant or substantively important negative effect.

**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 substantively important effect.

(continued)

## Appendix A5.1 *Fast ForWord*<sup>®</sup> rating for the alphabetic domain (continued)

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

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies 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: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No studies showed a statistically significant or substantively important effect, either positive or negative.

**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 studies showed a statistically significant negative effect.

**AND**

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

**Met.** No studies 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 Procedures and Standards Handbook, Appendix E.

## Appendix A5.2 *Fast ForWord*<sup>®</sup> rating for the reading 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 reading fluency, the WWC rated *Fast ForWord*<sup>®</sup> as having potentially positive effects for adolescent learners. It did not meet the criteria for positive effects because no studies showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, or negative effects) were not considered, as *Fast ForWord*<sup>®</sup> 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 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 studies showed a statistically significant or substantively important negative effect, no studies showed indeterminate effects, and one study showed substantively important positive effects.

### Other ratings considered

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

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

**Not met.** No studies showed statistically significant positive effects.

### AND

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

**Met.** No studies showed a statistically significant or substantively important negative 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 Procedures and Standards Handbook, Appendix E.

### Appendix A5.3 *Fast ForWord*<sup>®</sup> 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 *Fast ForWord*<sup>®</sup> as having potentially positive effects for adolescent learners. It did not meet the criteria for positive effects because only one study showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, or negative effects) were not considered, as *Fast ForWord*<sup>®</sup> 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 two studies showed substantively important positive effects.

#### AND

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

**Met.** No studies showed a statistically significant or substantively important negative effect. Three studies showed indeterminate effects, one study showed a statistically significant positive effect, and two studies showed substantively important positive effects.

#### Other ratings considered

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

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

**Not met.** Only one study showed a statistically significant positive effect.

#### AND

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

**Met.** No studies showed a statistically significant or substantively important negative 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 Procedures and Standards Handbook, Appendix E.

## Appendix A5.4 *Fast ForWord*<sup>®</sup> rating for the general literacy achievement 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 general literacy achievement, the WWC rated *Fast ForWord*<sup>®</sup> as having no discernible effects for adolescent learners. 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: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** No studies showed a statistically significant or substantively important effect, either positive or negative. Five studies 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.** No studies showed a statistically significant positive effect.

### AND

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

**Met.** No studies 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.** No studies showed 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.

**Not met.** No studies showed a statistically significant or substantively important negative effect. Five studies showed indeterminate effects, and no studies showed a statistically significant or substantively important positive 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 negative effect.

### 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 substantively important effect.

(continued)

## Appendix A5.4 *Fast ForWord*<sup>®</sup> rating for the general literacy achievement domain *(continued)*

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

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies 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: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No studies showed a statistically significant or substantively important effect, either positive or negative.

**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 studies showed a statistically significant negative effect.

**AND**

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

**Met.** No studies 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 Procedures and Standards Handbook, Appendix E.

## Appendix A6 Extent of evidence by domain

| Outcome domain               | Number of studies | Sample size      |                  | Extent of evidence <sup>1</sup> |
|------------------------------|-------------------|------------------|------------------|---------------------------------|
|                              |                   | Schools          | Students         |                                 |
| Alphabetics                  | 2                 | 7                | 161              | Small                           |
| Reading fluency              | 1                 | 3                | 24               | Small                           |
| Comprehension                | 6                 | >42 <sup>2</sup> | 1,376            | Medium to large                 |
| General literacy achievement | 5                 | 20               | 873 <sup>3</sup> | Medium to large                 |

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.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.
2. The number of control schools in Overbay and Baenen (2002) is unknown.
3. For Rouse and Krueger (2004), we counted the number of students as 454, which is from the state assessment. The actual number of students might be higher, as we do not know to what extent the number of students from the three outcome measures overlapped.