

What Works Clearinghouse



SuccessMaker®

Program Description¹

The *SuccessMaker*® program is a set of computer-based courses used to supplement regular classroom reading instruction in grades K–8. Using adaptive lessons tailored to a student’s reading level, *SuccessMaker*® aims to improve understanding in areas such as phonological awareness, phonics, fluency, vocabulary, comprehension, and concepts of print. “Foundations” courses aim to help students develop and maintain reading skills. “Exploaware”

courses aim to provide opportunities for exploration, open-ended instruction, and development of analytical skills. The computer analyzes students’ skills development and assigns specific segments of the program, introducing new skills as they become appropriate. As the student progresses through the program, performance is measured by the probability of the student answering the next exercise correctly, which determines the next steps of the lesson.²

Research

Three studies of *SuccessMaker*® meet What Works Clearinghouse (WWC) evidence standards³ with reservations. The three studies included 450 students, ranging in age from nine to 16 years, who attended elementary, middle, and middle-high schools in Alabama, Illinois, and Virginia.⁴

Based on these three studies, the WWC considers the extent of evidence for *SuccessMaker*® to be small for alphabetics, reading fluency, and general literacy achievement, and medium to large for comprehension.⁵

Effectiveness

SuccessMaker® was found to have no discernible effects on alphabetics and reading fluency, and potentially positive effects on comprehension and general literacy achievement.

	Alphabetics	Reading fluency	Comprehension	General literacy achievement
Rating of effectiveness	No discernible effects	No discernible effects	Potentially positive effects	Potentially positive effects
Improvement index⁶	Average: +1 percentile point Range: –8 to +5 percentile points	+9 percentile points na	Average: +11 percentile points Range: +1 to +15 percentile points	+11 percentile points na

na = not applicable

1. The descriptive information for this program was obtained from a publicly available source: the developer’s website (<http://www.pearsoned.com>, downloaded December 2008). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
2. The most current version of the program is called *SuccessMaker*®. Earlier versions were called *SuccessMaker*® Enterprise and *Computer Curriculum Corporation (CCC) SuccessMaker*®. We were unable to obtain documentation on the similarities and differences between these versions from the developer.
3. The studies included in this report were reviewed using WWC Evidence Standards, Version 1.0 (see the WWC Standards).
4. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
5. 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.”
6. These numbers show the average and range of student-level improvement indices for all findings across the studies.

Additional program information

Developer and contact

The research underlying *SuccessMaker*[®] was initiated by Patrick Suppes at Stanford University during the 1960s, continued by Mario Zanotti at the Computer Curriculum Corporation (Suppes and Zanotti, 1996), and extended and distributed by Pearson Digital Learning. Address: One Lake Street, Upper Saddle River, NJ 07458. Email: communications@pearsoned.com. Web: www.pearsoned.com. Telephone: (201) 236-7000.

Scope of use

According to the developers, *SuccessMaker*[®] has been used in more than 17,000 schools across the world. The program has been used with at-risk and accelerated learners, general and special education students, and English language learners.

Teaching

The software is a supplemental program that can be used in conjunction with existing language arts programs. “Foundations”

Research

Thirty-six studies reviewed by the WWC investigated the effects of *SuccessMaker*[®]. Three studies (Beattie, 2000; Campbell, 2000; Gallagher, 1996), one randomized controlled trial and two quasi-experimental designs, meet WWC evidence standards with reservations. Of the remaining studies, 33 studies do not meet WWC evidence standards or eligibility screens.

Beattie (2000) conducted a randomized controlled trial of middle and middle-high school students in suburban northern Virginia. Students with language deficits, ranging in age from 11 to 16 years, were randomly assigned by computer-generated procedures to one of five groups (Appendix A1.1 provides more details about these groups). The WWC based its effectiveness ratings on findings from comparisons of 14 students that received *SuccessMaker*[®] and 12 control group students that received regular reading instruction. Although these analytic samples were shown to be equivalent at baseline, differential attrition between groups led to the study’s rating of *meets*

courses contain basic skills-building exercises, while “Explore-ware” courses focus on application and literature-based reading aimed at building higher level analytical skills. Each student progresses through the computerized lessons at his or her own pace. The proportion of instruction across concept areas is adjusted for the individual so that weaker areas receive more emphasis. If a student continually struggles with a new concept, rather than staying on the difficult concept, *SuccessMaker*[®] sets the material aside to be reintroduced at a later point. This individualization allows each student to progress on his or her own time schedule. *SuccessMaker*[®] also periodically checks the student’s recollection of material previously mastered. Professional development for using *SuccessMaker*[®] is available and focuses on instructional strategies to incorporate *SuccessMaker*[®] into the curricula and customized on-site support for teachers.

Cost

Not available online.

standards with reservations. The study reported student outcomes after two months of program implementation.

Campbell (2000) conducted a quasi-experiment that examined the effects of *SuccessMaker*[®] on students in upper elementary grades in Alabama. The schools that used *SuccessMaker*[®] and traditional instruction (*Accelerated Reader* in conjunction with a basal reader) were matched to schools that used only traditional instruction based on the intellectual ability, poverty level, and demographic characteristics of students in each school. The WWC based its effectiveness ratings on findings for grade 4 students: 143 students in four intervention schools and 186 students in four comparison schools. The study reported student outcomes after one year of program implementation.

Gallagher (1996) conducted a quasi-experiment that examined the effects of *SuccessMaker*[®] on at-risk students in grades 4–7 at an inner city elementary school in Chicago, IL. Students in each classroom were sorted by either reading achievement test score or student identification number (ID), and then

Research *(continued)*

alternately assigned to treatment and control groups.⁷ The WWC based its effectiveness ratings on findings from comparisons of the 48 students that received two reading components of *SuccessMaker*[®] (Readers Workshop and Reading Adventures) and the 47 control group students that received math components of *SuccessMaker*[®]. Both groups received their regular reading curriculum outside of the *SuccessMaker*[®] instruction. The study reported student outcomes after six weeks of program implementation.

Effectiveness Findings

The WWC review of interventions for *SuccessMaker*[®] addresses student outcomes in four domains: alphabetics, reading fluency, comprehension, and general literacy achievement. The studies included in this report cover all four domains. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *SuccessMaker*[®] on students.⁹

Alphabetics. Beattie (2000) did not find statistically significant effects of *SuccessMaker*[®] on alphabetics measures, including the Woodcock-Johnson subtests of Letter-Word Identification,

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.⁸

The WWC considers the extent of evidence for *SuccessMaker*[®] to be small for alphabetics, reading fluency, and general literacy achievement, and medium to large for comprehension.

Word Attack, and Auditory Processing, and the Wide Range Achievement Spelling subtest. The WWC-calculated average effect size across the four outcomes was not large enough to be considered substantively important according to WWC criteria (that is, an effect size at least 0.25).¹⁰

Reading fluency. Beattie (2000) did not find a statistically significant effect of *SuccessMaker*[®] on the Gray Oral Reading Test, and the effect was not large enough to be considered substantively important according to WWC criteria.

Comprehension. Beattie (2000) did not find statistically significant effects of *SuccessMaker*[®] on the Woodcock-Johnson

7. The authors *either* sorted the students by student identification numbers (ID) or Iowa Test of Basic Skills (ITBS) reading comprehension scores, and *then* assigned students to groups in an alternating fashion, but it is not clear which method was used from the text. If they sorted by student ID and then assigned students to groups, the assignment might be functionally random, but if they sorted by ITBS score, and always assigned students in an alternating fashion (starting with the treatment group, for example), the groups would be imbalanced, because they were always assigning the lower (or higher) scores to the treatment group. The WWC could not confirm that the assignment was truly random, as the authors had not responded to the WWC query at the time of publication of this review.
8. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *SuccessMaker*[®] is in Appendix A6.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, 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 levels may differ from those reported in the original study. In the case of Campbell (2000), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study. In the case of Gallagher (1996), no corrections for clustering or multiple comparisons were needed.
10. The WWC computes an average effect size (ES) as a simple average of the ESs across all individual findings within the study domain. For information on how the WWC characterizes study effects, consult the WWC Procedures and Standards Handbook, Appendix E.

Effectiveness *(continued)*

Passage Comprehension subtest, but the effect size was large enough to be considered substantively important according to WWC criteria (that is, an effect size at least 0.25). Campbell (2000) did not find statistically significant effects of *SuccessMaker*® on either measure of comprehension examined (the Stanford Achievement Reading Vocabulary and Reading Comprehension subtests). The WWC-calculated average effect size across the two outcomes was not large enough to be considered substantively important according to WWC criteria. Gallagher (1996) found a statistically significant effect of *SuccessMaker*® on the reading comprehension subtest of the Iowa Test of Basic Skills. The WWC found that the effect was not statistically significant but large enough to be considered substantively important according to WWC criteria.¹¹

The WWC found *SuccessMaker*® to have no discernible effects on alphabets and reading fluency, and potentially positive effects on comprehension and general literacy achievement

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see WWC Procedures and Standards Handbook, Appendix F). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is based entirely on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analyses. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.¹²

The average improvement index for alphabets is +1 percentile point (based on findings from one study), with a range of -8 to +5 percentile points across findings. The improvement index

General literacy achievement. Beattie (2000) did not find statistically significant effects of *SuccessMaker*® on the Clinical Evaluation of Language Fundamentals Receptive Language Score, but the effect size was large enough to be considered substantively important according to WWC criteria (that is, an effect size at least 0.25).

Rating of effectiveness

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

for reading fluency is +9 percentile points for a single finding from one study. The average improvement index for comprehension is +11 percentile points across three studies, with a range of +1 to +15 percentile points across findings. The improvement index for general literacy achievement is +11 percentile points for a single finding from one study.

Summary

The WWC reviewed 36 studies on *SuccessMaker*®. Three of these studies meet WWC evidence standards with reservations. Of the remaining studies, 33 studies do not meet WWC evidence standards or eligibility screens. Based on the three studies, the WWC found no discernible effects in alphabets and reading fluency, and potentially positive effects in comprehension and general literacy achievement. The conclusions presented in this report may change as new research emerges.

11. The study is not consistent in reporting the numbers of students allocated to treatment and control groups. The WWC calculated the groups' sample sizes, means, and standard deviations from the raw data presented in the study appendices.

12. For information on how to interpret the improvement index, consult WWC Procedures and Standards Handbook, Appendix F.

References **Meet WWC evidence standards with reservations**

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). *Dissertation Abstracts International*, 61(08A), 194–3116.

Additional source:

Given, B. K., Wasserman, J. D., Chari, S. A., Beattie, K., & Eden, G. F. (2008). A randomized, controlled study of computer-based intervention in middle school struggling readers. *Brain & Language*, 106(2), 83–97.

Campbell, J. P. (2000). A comparison of computerized and traditional instruction in the area of elementary reading (Doctoral dissertation, University of Alabama, 2000). *Dissertation Abstracts International*, 61(03A), 77–952.

Gallagher, E. M. (1996). Utilization of an ILS to increase reading comprehension (integrated learning systems, CAI) (Doctoral dissertation, Northern Illinois University, 1996). *Dissertation Abstracts International*, 58(05A), 79–1591.

Studies that fall outside the Adolescent Literacy protocol or do not meet WWC evidence standards

Aeby, V. G., Powell, J. V., & Carpenter-Aeby, T. (2000). Effects of *SuccessMaker* computerized curriculum on the behavior of disruptive students. *Journal of Educational Technology Systems*, 28(4), 335–347. The study is ineligible for review because it does not use a comparison group.

Brush, T. A. (1998). *An evaluation of the effectiveness of the Computer Curriculum Corporation's (CCC) Foundations and Explores software on students in grades one through five*. Unpublished manuscript. The study is ineligible for review because it does not use a comparison group.

Domenech, D. (2002). *Project Excel interim evaluation report, year 2. Fairfax County Public Schools*. Salt Lake City, UT: Waterford Institute. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Donnelly, L. F. (2004). *Year two results: Evaluation of the implementation and effectiveness of SuccessMaker during 2002–2003*. Charleston, SC: Charleston County School District, Division of Student Learning Services, Division of Student Assessment, Program Evaluation and Data Management. The study is ineligible for review because it only includes outcomes that are overlapped with the intervention or measured in a way that is inconsistent with the protocol.

Education Commission of the States. (1999). *SuccessMaker*. Denver, CO: Author. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Hargrave, S. (2003). *Can SuccessMaker be used as a predictor of proficiency testing achievement?* Unpublished master's thesis, Franciscan University of Steubenville. The study is ineligible for review because it does not use a comparison group.

Hauswirth, K. (2006). Using comprehensive digital learning systems. *Learning & Leading with Technology*, 34(3), 32–33. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. Baker, & H. O'Neil (Eds.), *Technology assessment in education and training* (pp. 9–33). Hillsdale, NJ: Lawrence Erlbaum Associates. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say. Final report*. Arlington, VA: SRI International. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Levitt, J. L. (2000). *An interim evaluation of Operation Safety Net: A five-year project*. Miami-Dade, FL: Office of Evaluation and Research, Miami-Dade County. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

References *(continued)*

- McWhirt, R., Mentavlos, M., Rose-Baele, J. S., & Donnelly, L. (2003). *Evaluation of the implementation and effectiveness of SuccessMaker*. Charleston, SC: Charleston County School District. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.
- Metis Associates. (1999). *Community school district six [New York City schools] integrated technology reading support project: Third year evaluation report*. New York, NY: Author. The study is ineligible for review because it does not use a comparison group.
- Miller, B. S. W. (1999). *Opinions of teachers regarding the effects of educational technology in the elementary classroom*. Greeneville, TN: Tusculum College. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Miller, L., DeJean, J., & Miller, R. (2000). The literacy curriculum and use of an integrated learning system. *Journal of Research in Reading*, 23(2), 123–135. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Mills, S. C., & Ragan, T. R. (1998). *An implementation model for integrated learning systems*. National Convention of the Association for Educational Communications and Technology (AECT), St. Louis, MO. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Oakley, G. (2003). Improving oral reading fluency (and comprehension) through the creation of talking books. *Reading Online*, 6(7). The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Parr, J. M. (1997). Computer assisted learning with an integrated learning system: Another front for raising literacy and numeracy amongst secondary students? *New Zealand Journal of Educational Studies*, 32(1), 37–51. The study is ineligible for review because it does not use a comparison group.
- Pearson Education Technologies. (2002). *SuccessMaker evidence of effectiveness: Selected evaluation summaries*. Scottsdale, AZ: Author. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.
- Perez, K. J. (1998). Predictors of achievement in math and reading by elementary ESOL and non-ESOL students using a computer-based integrated learning system (Doctoral dissertation, Barry University School of Education, 1998). *Dissertation Abstracts International*, 59(07A), 106–2354. The study is ineligible for review because it does not use a comparison group.
- Powell, J. V., Aeby, J., Victor, G., & Carpenter-Aeby, T. (2003). A comparison of student outcomes with and without teacher facilitated computer-based instruction. *Computers & Education*, 40(2), 183–191. The study is ineligible for review because it does not use a comparison group.
- Presland, A., & Wishart, J. (2004). Secondary school pupils' motivations to use an integrated learning system. *British Journal of Educational Technology*, 35(5), 663–668. The study is ineligible for review because it does not use a comparison group.
- Rhine, C. (2005). *Beyond fluency to meaning: Using SuccessMaker to build comprehension skills in English language learners*. Unpublished master's thesis, California State University–San Marcos. The study is ineligible for review because it does not use a comparison group.
- Simon, C., & Tingey, B. (2001). *Seminole county public schools relationship study for 2000–2001*. Upper Saddle River, NJ: Pearson Education. The study is ineligible for review because it does not use a comparison group.
- Slavin, R., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly*, 43(3), 290–322. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.
- Stanley, W. A. (2001). *Computer validity: An evaluation of CCC SuccessMaker*. Unpublished master's thesis, North Dakota State University. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

References *(continued)*

- Tingey, B., & Simon, C. (2001). *Relationship study for SuccessMaker levels and SAT-9 in Hueneme elementary district, school year 2000–2001, with growth analysis 2000–2001*. Upper Saddle River, NJ: Pearson Education. The study is ineligible for review because it does not use a comparison group.
- Tingey, B., & Thrall, A. (2000). *Duval county public schools evaluation report for 1999–2000*. Upper Saddle River, NJ: Pearson Education. The study is ineligible for review because it does not use a comparison group.
- Tingey, B., & Thrall, A. (2002). High stakes management. *Multimedia Schools*, 9(2), 1–7. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.
- Underwood, J. D. M. (2000). A comparison of two types of computer support for reading development. *Journal of Research in Reading*, 23(2), 136–148. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.
- Waxman, H. C., Lin, M., & Michko, G. M. (2003). *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes*. Naperville, IL: Learning Point Associates. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.
- Wheeler, J. M. (2007). *The impact of SuccessMaker on the FCAT reading scores of level-1 and level-2 sixth graders*. Unpublished doctoral dissertation, University of West Florida. The study is ineligible for review because it does not use a comparison group.
- Winters, J. L. (2000). Perceptions of middle school students concerning their language and reading abilities under different instructional interventions. (Doctoral dissertation, George Mason University, 2000). *Dissertation Abstracts International*, 61(02A), 163–569. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Wood, D., Underwood, J., & Avis, P. (1999). Integrated learning systems in the classroom. *Computers & Education*, 33(2-3), 91–108. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

For more information about specific studies and WWC calculations, please see the WWC SuccessMaker® Technical Appendices.

Appendix

Appendix A1.1 Study characteristics: Beattie, 2000 (randomized controlled trial with severe attrition)

Characteristic	Description
Study citation	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.
Participants	Eighty-one students with language deficits, ranging in age from 11 to 16 years, were randomly assigned by computer-generated procedures to one of four intervention groups ¹ or to a control group in a two-step process. 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 14 students in the <i>SuccessMaker</i> [®] group who were compared to 12 students in the comparison group. ² Although the differential attrition rate was higher than 7%, the post-attrition intervention and comparison groups were equivalent on the pretest achievement measures.
Setting	Two middle schools and one middle-high school located in the suburbs of a large metropolitan area in northern Virginia.
Intervention	Students worked on <i>SuccessMaker</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 student outcomes after two months of program implementation.
Comparison	The control group received the standard instruction provided in the regular school curriculum.
Primary outcomes and measurement	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.
Staff/teacher training	No information on training for the teachers and staff in this study was provided. To facilitate the use of <i>SuccessMaker</i> [®] , computers were procured or updated to meet criteria for running <i>SuccessMaker</i> [®] software.

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

Appendix A1.2 Study characteristics: Campbell, 2000 (quasi-experimental design)

Characteristic	Description
Study citation	Campbell, J. P. (2000). A comparison of computerized and traditional instruction in the area of elementary reading (Doctoral dissertation, University of Alabama, 2000). <i>Dissertation Abstracts International</i> , 61(03A), 77–952.
Participants	Based on the School Ability Index score, five elementary schools that used both <i>SuccessMaker</i> ® and traditional instruction were matched to five elementary schools that used only traditional instruction. ¹ Poverty level and gender were similar across intervention and comparison schools. Although the overall and differential student attrition rates were high (58% and 37%, respectively), the post-attrition intervention and comparison samples of fourth-graders were equivalent on both subtests of the Stanford Achievement Test at baseline. ² After one year, 143 students in four <i>SuccessMaker</i> ® schools and 186 students in four comparison schools remained in the sample.
Setting	The analysis sample included eight elementary schools in Etowah County, Alabama.
Intervention	Students in the intervention group received 10 to 20 minutes of <i>SuccessMaker</i> ® instruction daily. They were also given traditional instruction that included the <i>Accelerated Reader</i> program in conjunction with a basal reader. The study was conducted during the first year of <i>SuccessMaker</i> ® program implementation.
Comparison	Comparison classrooms implemented the standard district curriculum, which used the <i>Accelerated Reader</i> program in conjunction with a basal reader.
Primary outcomes and measurement	For both pre- and posttests, the author used two subtests of the Stanford Achievement Test administered by schools. The Vocabulary and Reading Comprehension Otis Lennon School Ability test was also used in the study, but was not included in this report because it was outside the scope of the Adolescent Literacy review. For a more detailed description of the outcome measures included in this report, see Appendix A2.3.
Staff/teacher training	In order to maintain consistency in the administration of the outcome measure (SAT-9), all test administrators and proctors were trained in the areas of test security and proper administration techniques.

1. For the overall grade 5 analysis sample, the intervention and comparison groups were not shown to be equivalent at baseline and are, therefore, excluded from review. As a result, two schools were dropped from the analysis.
2. WWC aggregated reading achievement data across schools to conduct the analyses.

Appendix A1.3 Study characteristics: Gallagher, 1996 (quasi-experimental design)

Characteristic	Description
Study citation	Gallagher, E. M. (1996). Utilization of an ILS to increase reading comprehension (integrated learning systems, CAI) (Doctoral dissertation, Northern Illinois University, 1996). <i>Dissertation Abstracts International</i> , 58(05A), 79–159.
Participants	Students in grades 4–7 were pretested using the Iowa Test of Basic Skills (ITBS), sorted by either the ITBS score or student identification number (ID), and then alternately assigned to treatment or control groups within classrooms. ¹ All of the students were African-American and were eligible for the federal free lunch program. Students who scored below 3.0 on the reading comprehension subtest or who were part of the school's special education program were eliminated from the study's sample prior to the assignment. Although the overall attrition rate at posttest was 38%, the post-attrition intervention and comparison groups were equivalent on the reading achievement pretest measure (ITBS). In all, 48 students in the <i>SuccessMaker</i> ® group and 47 students in the comparison group were included in the analysis sample. Additional findings reflecting student outcomes by grade can be found in Appendix A4.
Setting	The study took place in an inner city elementary school in Chicago, Illinois.
Intervention	The intervention group spent a minimum of 40 minutes a day on the two reading components of the <i>SuccessMaker</i> ® program. The Readers Workshop component is an individualized basic skill building program. In the first 100 minutes a student participates in the program, the computer analyzes their skills development and assigns specific segments of the program appropriate to further develop the students' skills, introducing new skills as they become appropriate. The Reading Adventures component places each student at a reading level and provides stories and comprehension questions at that level. The student progresses through a semi-linear program where the only choice is among stories at the assigned level. Outside of the <i>SuccessMaker</i> ® instruction, the intervention group also received the regular reading curriculum. The study reported students' outcomes after six weeks of program implementation.
Comparison	The comparison group spent a minimum of 40 minutes a day on the math components of the <i>SuccessMaker</i> ® program (Math Concepts and Skills and Problem Solving). Comparison students also received the regular reading curriculum.
Primary outcomes and measurement	For both pre- and posttests, the author used the reading comprehension subtest of the Iowa Test of Basic Skills. For a more detailed description of this outcome measure, see Appendix A2.3.
Staff/teacher training	No information on training for the teachers and staff in this study was provided.

1. The authors *either* sorted the students by student identification numbers (ID) or Iowa Test of Basic Skills (ITBS) reading comprehension scores, and *then* assigned students to groups in an alternating fashion, but it is not clear which method was used from the text. If they sorted by student ID and then assigned students to groups, the assignment might be functionally random, but if they sorted by ITBS score, and always assigned students in an alternating fashion (starting with the treatment group, for example), the groups would be imbalanced, because they were always assigning the lower (or higher) scores to the treatment group. The WWC could not confirm that the assignment was truly random, as the authors had not responded to the WWC query at the time of publication of this review.

Appendix A2.1 Outcome measures for the alphabetic domain

Outcome measure	Description
<i>Phonemic awareness</i>	
Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R), Tests of Cognitive Abilities: Auditory Processing subtest	This composite is a standardized measure of a student's ability to appreciate patterns among speech-based auditory stimuli. The score 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 ability to indicate whether pairs of computer-generated sound sequences are the same or different (as cited in Beattie, 2000).
<i>Phonics</i>	
WJ-R, Tests of Achievement: Word Attack subtest	This standardized subtest measures phonemic decoding skills by asking students to read pseudowords (e.g., plurp, fronkett). Students are aware that the words are not real (as cited in Beattie, 2000 and http://www.concordspedpac.org/WJ-III-subtests.htm#Achievement).
WJ-R, Tests of Achievement: Letter-Word Identification subtest	This standardized subtest requires the student to read aloud isolated letters and real words that range in frequency and difficulty (as cited in Beattie, 2000).
Wide Range Achievement Test-Third Edition (WRAT-3): Spelling subtest	This standardized subtest is a paper-and-pencil task that tests students' ability to write their names, as well as letters and words from dictation. Dictated letters and words followed either phonetically regular or irregular patterns (as cited in Beattie, 2000).

Appendix A2.2 Outcome measures for the reading fluency domain

Outcome measure	Description
Gray Oral Reading Test-Third edition (GORT-3)	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 is reflective of a total measure of one'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
Vocabulary development	
The Stanford Achievement Test (SAT–9): Reading Vocabulary subtest	This standardized subtest is composed of multiple-choice and open-ended assessment questions that measure word reading and achievement. The open-ended reading section includes a narrative reading selection followed by nine questions. There are three types of reading selections: (1) recreational (material read for enjoyment or literary merit, including folk tales, historical fiction, contemporary fiction, humor, and poetry), (2) textual (expository material with content from the natural, physical, and social sciences, as well as other nonfiction general information materials) and (3) functional (material encountered in everyday life both inside and outside of school, including directions, forms, labels, schedules, and advertisements) (as cited in Campbell, 2000 and http://brighted.funeducation.com/Prepare/StateTests/?state=SAT-9).
Reading comprehension	
SAT–9: Reading Comprehension subtest	This standardized subtest is based on questions that range from interpreting simple sentences to understanding more complex paragraphs. The questions on complex paragraphs ask the student to recognize directly stated details or relationships, as well as implicit information and relationships that demand integration of what is provided in the text (as cited in Campbell, 2000).
WJ–R, Tests of Achievement: Passage Comprehension subtest	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 and http://www.concordspedpac.org/WJ-III-subtests.htm#Achievement).
The Iowa Test of Basic Skills: Reading Comprehension subtest	This standardized test consists of reading passages of varying length and difficulty and assesses three types of understanding: (1) factual questions tap students’ literal understanding of what is stated in the text; (2) inferential/interpretive questions require students to “read between the lines” to demonstrate their understanding of what is implied; and (3) analysis and generalization questions require students to “step back from” the text to generalize about a passage’s main points or ideas or to analyze aspects of the author’s viewpoint or use of language (as cited in http://www.riverpub.com/products/itbs/details.html).

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

Outcome measure	Description
Clinical Evaluation of Language Fundamentals–Third Edition (CELF–3): Receptive Language Score	This standardized assessment measures a student’s ability to interpret and execute commands of increasing complexity and understand relationships between words and categories. It addresses sentence structure, concepts and directions, and word classes (as cited in Beattie, 2000).

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome ² (standard deviation) ³		Mean difference ⁴ (<i>SuccessMaker</i> [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			<i>SuccessMaker</i> [®] group	Comparison group				
Beattie, 2000 (randomized controlled trial with attrition)⁸								
WJ–R Letter-Word Identification subtest	11–16 yrs old	26	89.69 (9.48)	92.08 (13.15)	–2.39	–0.20	ns	–8
WJ–R Word Attack subtest	11–16 yrs old	26	86.99 (17.65)	85.91 (12.87)	1.08	0.07	ns	+3
WJ–R Auditory Processing subtest	11–16 yrs old	26	87.44 (13.38)	85.66 (15.61)	1.78	0.12	ns	+5
WRAT–3 Spelling subtest	11–16 yrs old	26	87.02 (12.66)	85.66 (13.13)	1.36	0.10	ns	+4
Average for alphabetics (Beattie, 2000)⁹						0.02	ns	+1

ns = not statistically significant

WJ–R = Woodcock-Johnson Revised

WRAT-3 = Wide Range Achievement Test–Third Edition

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 mean 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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate 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 levels may differ from those reported in the original study.
9. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the reading fluency domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome ² (standard deviation) ³		Mean difference ⁴ (<i>SuccessMaker</i> [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			<i>SuccessMaker</i> [®] group	Comparison group				
Beattie, 2000 (randomized controlled trial with attrition)⁸								
Gray Oral Reading test (GORT–3)	11–16 yrs old	26	83.18 (12.72)	79.50 (17.76)	3.68	0.23	ns	+9
Average for reading fluency (Beattie, 2000)⁹						0.23	ns	+9

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 mean 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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. 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 A3.3 Summary of study findings included in the rating for the comprehension domain¹

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study ²			WWC calculations		
			Mean outcome ³ (standard deviation) ⁴		Mean difference ⁵ (<i>SuccessMaker</i> [®] – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>SuccessMaker</i> [®] group	Comparison group				
Beattie, 2000 (randomized controlled trial with attrition)⁹								
WJ–R Passage Comprehension subtest	11–16 yrs old	26	97.03 (8.08)	93.25 (11.30)	3.78	0.38	ns	+15
Average for comprehension (Beattie, 2000)¹⁰						0.38	ns	+15
Campbell, 2000 (quasi-experimental design)⁹								
SAT–9 Reading Vocabulary subtest	Grade 4	8/329	60.54 (23.36)	60.01 (24.12)	0.53	0.02	ns	+1
SAT–9 Reading Comprehension subtest	Grade 4	8/329	60.29 (23.14)	58.08 (24.76)	2.21	0.09	ns	+4
Average for comprehension (Campbell, 2000)¹⁰						0.06	ns	+2
Gallagher, 1996 (quasi-experimental design)⁹								
ITBS Reading Comprehension subtest	Grades 4–7	95	30.25 (10.78)	26.72 (8.32)	3.53	0.36	ns	+14
Average for comprehension (Gallagher, 1996)¹⁰						0.36	ns	+14
Domain average for comprehension across all studies¹⁰						0.27	na	+11

ns = not statistically significant

na = not applicable

WJ–R = Woodcock–Johnson–Revised

SAT–9 = Stanford Achievement Test

ITBS = Iowa Test of Basic Skills

- This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain.
- For Gallagher (1996), the WWC calculated groups' sample sizes, means, and standard deviations from the raw data presented in the study appendices.
- The intervention group values are the comparison group means plus the difference in mean gains between the intervention and comparison groups.
- 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. For Campbell (2000), the WWC aggregated means and standard deviations across four schools.
- Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
- For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
- Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
- 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.
- The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. 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 Campbell (2000), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study. In the cases of Beattie (2000) and Gallagher (1996), no corrections for clustering or multiple comparisons were needed.
- The WWC-computed domain average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

Appendix A3.4 Summary of study findings included in the rating for the general literacy achievement domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome ² (standard deviation) ³		Mean difference ⁴ (<i>SuccessMaker</i> [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			<i>SuccessMaker</i> [®] group	Comparison group				
Beattie, 2000 (randomized controlled trial with attrition)⁸								
Receptive Language subtest (CELF–3)	11–16 yrs old	26	92.81 (18.35)	86.63 (22.74)	5.98	0.28	ns	+11
Average for general literacy achievement (Beattie, 2000)⁹						0.28	ns	+11

ns = not statistically significant

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

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 mean 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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. 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 A4 Summary of subgroup findings for the comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study ²					
			Mean outcome ³ (standard deviation) ⁴		WWC calculations			
			SuccessMaker [®] group	Comparison group	Mean difference ⁵ (SuccessMaker [®] – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
Gallagher, 1996 (quasi-experimental design)⁹								
ITBS Reading comprehension subtest	Grade 4	32	31.69 (8.01)	29.19 (5.53)	2.50	0.35	ns	+14
ITBS Reading comprehension subtest	Grade 5	32	27.31 (10.74)	25.00 (9.64)	2.31	0.22	ns	+9
ITBS Reading comprehension subtest	Grade 6	20	34.60 (12.43)	24.80 (6.12)	9.80	0.96	Statistically significant	+33
ITBS Reading comprehension subtest	Grade 7	11	27.47 (14.22)	28.20 (14.13)	–0.73	–0.05	ns	–2

ns = not statistically significant

ITBS = Iowa Test of Basic Skills

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 A3.3.
2. For Gallagher (1996), the WWC calculated groups' sample sizes, means, and standard deviations from the raw data presented in the study appendices.
3. The intervention group values are the comparison group means plus the difference in mean gains between the intervention and comparison groups.
4. 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.
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. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate 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 Gallagher (1996), no correction for clustering was needed.

Appendix A5.1 *SuccessMaker*[®] rating for the alphabets 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.¹

For the outcome domain of alphabets, the WWC rated *SuccessMaker*[®] as having no discernible effects. It did not meet the criteria for positive, potentially positive, mixed, potentially negative, 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 statistically significant or substantively important effects, either positive or negative.

Other ratings considered

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

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

Not met. No studies showed statistically significant positive effects.

AND

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

Met. No studies showed statistically significant or substantively important negative effects.

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

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

Not met. 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, and one study showed indeterminate effects.

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

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

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

OR

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

Not met. No studies showed a statistically significant or substantively important effect, and one study showed indeterminate effects.

(continued)

Appendix A5.1 SuccessMaker® rating for the alphabetics domain (continued)

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

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

Not met. No studies showed a statistically significant or substantively important negative effect.

AND

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

Met. No studies showed statistically significant or substantively important positive effects.

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

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

Not met. No studies showed statistically significant or substantively important negative effects.

AND

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

Met. No studies showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5.2 SuccessMaker® 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.¹

For the outcome domain of reading fluency, the WWC rated *SuccessMaker*® as having no discernible effects. It did not meet the criteria for positive, potentially positive, mixed, potentially negative, 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 statistically significant or substantively important effects, either positive or negative.

Other ratings considered

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

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

Not met. No studies showed statistically significant positive effects.

AND

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

Met. No studies showed statistically significant or substantively important negative effects.

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

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

Not met. 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, and one study showed indeterminate effects.

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

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

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

OR

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

Not met. No studies showed a statistically significant or substantively important effect, and one study showed indeterminate effects.

(continued)

Appendix A5.2 SuccessMaker® rating for the reading fluency domain (continued)

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

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

Not met. No studies showed a statistically significant or substantively important negative effect.

AND

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

Met. No studies showed statistically significant or substantively important positive effects.

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

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

Not met. No studies showed statistically significant or substantively important negative effects.

AND

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

Met. No studies showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5.3 *SuccessMaker*[®] 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.¹ For the outcome domain of comprehension, the WWC rated *SuccessMaker*[®] as having potentially positive effects.

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. 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, and one study showed indeterminate effects.

Other ratings considered

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

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

Not met. No studies showed statistically significant positive effects.

AND

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

Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level 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 SuccessMaker® 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.¹ For the outcome domain of general literacy achievement, the WWC rated SuccessMaker® as having potentially positive effects.

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, and one study showed indeterminate effects.

Other ratings considered

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

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

Not met. No studies showed statistically significant positive effects.

AND

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

Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level 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 ¹
		Schools	Students	
Alphabetics	1	3	26	Small
Reading fluency	1	3	26	Small
Comprehension	3	12	450	Medium to large
General literacy achievement	1	3	26	Small

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.