What Works Clearinghouse™ U.S. DEPARTMENT OF EDUCATION

WWC Intervention Report

A summary of findings from a systematic review of the evidence

ies INSTITUTE OF EDUCATION SCIENCES

Adolescent Literacy Updated November 2015

SuccessMaker®

Program Description[1]

SuccessMaker® is a set of computer-based courses designed to supplement regular K–8 reading instruction. The program aims to improve skills in areas such as phonological awareness, phonics, fluency, vocabulary, comprehension, concepts of print, grammar, and spelling. The software adapts instruction to match students' skill level and progress. "Foundations" courses contain basic skill-building exercises, while "Exploreware" courses focus on reading and writing activities aimed at building higher level analytical skills. The program analyzes students' progress and assigns specific segments of the lesson, introducing new skills as they become appropriate. As the student progresses, an algorithm calculates the probability of the student answering the next exercise correctly, which determines the next steps of the lesson.

Research[2]

The What Works Clearinghouse (WWC) identified one study of *SuccessMaker*® that falls within the scope of the Adolescent Literacy topic area and meets WWC group design standards. This study meets WWC group design standards without reservations. This study included 1,094 adolescent readers in grades 5 and 7 in nine schools located in seven states across the United States.

The WWC considers the extent of evidence for *SuccessMaker*® on the reading skills of adolescent readers to be small for two student outcome domains—comprehension and reading fluency. There were no studies that met standards in the two other domains, so this intervention report does not report on the effectiveness of *SuccessMaker*® for those domains. (See the Effectiveness Summary on p. 4 for more details of effectiveness by domain.)

Effectiveness

SuccessMaker® was found to have no discernible effects on comprehension and reading fluency for adolescent readers.

Table 1. Summary of findings[3]

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
Comprehension	No discernible effects	+3	+2 to +5	1	1,094	Small
Reading fluency	No discernible effects	–3	–4 to –1	1	1,087	Small

Report Contents

This intervention report presents findings from a systematic review of *SuccessMaker*® conducted using the WWC Procedures and Standards Handbook, version 3.0, and the Adolescent Literacy review protocol, version 3.0.

Program Information

Background

SuccessMaker® was developed by Patrick Suppes at Stanford University and Mario Zanotti at the Computer Curriculum Corporation in the late 1960s (Suppes & Zanotti, 1996).[4] The program is now owned and distributed by Pearson Education. Earlier versions of the program were called *SuccessMaker® Enterprise* and *Computer Curriculum Corporation (CCC) SuccessMaker*®. The most current version of the program is called *SuccessMaker® 8*. Address: One Lake Street, Upper Saddle River, NJ 07458. Email: communications@pearsoned.com. Web: www.pearsoned.com. Telephone: (201) 236-7000.

Program details

The *SuccessMaker*® software, also referred to as an integrated learning system by authors of studies included in this review, is a supplemental program used in conjunction with existing language arts curricula. The program includes an instructional management system, formative assessments, a reporting system with information on student progress, and individualized reading curriculum resources for elementary and middle school instruction. Program activities (practice, tutoring, and games) are based on selections from classic literature for children and adolescents. Initial program courses, "Foundations," contain basic skill-building exercises, while "Exploreware" courses focus on reading and writing activities aimed at building higher level analytical skills. The program offers approximately 43 hours of instruction per grade. Each student progresses through the computerized lessons at his or her own pace. The program individualizes instruction and provides real-time feedback and tutorials for students who encounter challenges during a lesson. If a student continually struggles with a new concept, *SuccessMaker*® sets the material aside to be reintroduced at a later point. *SuccessMaker*® also periodically checks the student's recollection of previously mastered material. 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

Cost information for *SuccessMaker*® is available from the distributor.

Research Summary

The WWC identified nine eligible studies that investigated the effects of *SuccessMaker*® on the reading skills of adolescent readers. An additional 57 studies were identified but do not meet WWC eligibility criteria for review in this topic area. Citations for all 66 studies are in the References section, which begins on p. 6.

Table 2. Scope of reviewed research

Grades	5, 7
Delivery method	Individual
Program type	Supplement

The WWC reviewed nine eligible studies against group design standards. One study (Gatti, 2011) is a randomized controlled trial that meets WWC group design standards without reservations. This study is summarized in this report. Eight studies do not meet WWC group design standards.

Summary of study meeting WWC group design standards without reservations

Gatti (2011) conducted a cluster randomized controlled trial that examined the effects of *SuccessMaker*® on fifth- and seventh-grade students[5] attending nine schools in Arizona, California, Indiana, Kansas, Michigan, Missouri, and Texas. Within each school, English language arts classes[6] were randomly assigned either to receive the *SuccessMaker*® program as a supplement to current instruction or to receive the schools' regular English language arts program. The WWC based its effectiveness ratings on findings from the 641 fifth-grade students who participated in the study; 342 students in the *SuccessMaker*® group and 299 students in the regular English language arts program, and the 453 seventh-grade students who participated in the study; 254 students in the *SuccessMaker*® group and 199 students in the regular English language arts program. The study reported student outcomes after 1 year of program implementation.

Summary of studies meeting WWC group design standards with reservations

No studies of *SuccessMaker*® met WWC group design standards with reservations.

Effectiveness Summary

The WWC review of *SuccessMaker*® for the Adolescent Literacy topic area includes student outcomes in four domains: alphabetics, comprehension, general literacy achievement, and reading fluency. The one study of *SuccessMaker*® that meets WWC group design standards reported findings in two of the four domains: (a) comprehension and (b) reading fluency. The findings below present the author's estimates and WWC-calculated estimates of the size and statistical significance of the effects of *SuccessMaker*® on adolescent readers. Additional comparisons are presented as supplemental findings in the appendix. The supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 19.

Summary of effectiveness for the comprehension domain

One study that meets WWC group design standards without reservations reported findings in the comprehension domain.

Gatti (2011) found that *SuccessMaker*® had statistically significant positive effects on the Overall Score, Passage Comprehension, and Sentence Comprehension subtests of the Group Reading Assessment and Diagnostic Evaluation (GRADE) for fifth- and seventh-grade students when compared to the regular English language arts program alone. The WWC could not confirm the statistical significance of these findings after adjusting of the clustering of students within classrooms. The average effect size across the two grades on the GRADE overall score was not large enough to be considered substantively important, according to WWC criteria. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the comprehension domain, one study showed indeterminate effects. This results in a rating of no discernible effects, with a small extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the comprehension domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *comprehension* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 1,094 students in nine schools reported evidence of effectiveness in the *comprehension* domain.

Summary of effectiveness for the reading fluency domain

One study that meets WWC group design standards without reservations reported findings in the reading fluency domain.

Gatti (2011) found that *SuccessMaker*® had a statistically significant negative effect on the AIMSweb Reading Curriculum-Based Measurement (AIMSweb R-CBM) number of words read correctly (WRC) for fifth-grade students when compared to the regular English language arts program alone. The WWC could not confirm the statistical significance of this finding after adjusting for the clustering of students within classrooms. The author did not find statistically significant effects of *SuccessMaker*® on the AIMSweb R-CBM WRC for seventh-grade students when compared to the regular English language arts program alone. The WWC confirmed the lack of statistical significance of this finding. The average effect size across the two grades was not large enough to be considered substantively important, according to WWC criteria. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the reading fluency domain, one study showed indeterminate effects. This results in a rating of no discernible effects, with a small extent of evidence.

Table 4. Rating of effectiveness and extent of evidence for the reading fluency domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *reading fluency* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 1,087 students in five schools reported evidence of effectiveness in the *reading fluency* domain.

References

Study that meets WWC group design standards without reservations

Gatti, G. (2011). *Pearson SuccessMaker reading efficacy study 2010–11 final report*. Pittsburgh, PA: Gatti Evaluation, Inc.

Studies that meet WWC group design standards with reservations

None.

Studies that do not meet WWC group design standards

Beattie, K. K. (2000). *The effects of intensive computer-based language intervention on language functioning and reading achievement in language-impaired adolescents* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9983709) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

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.

Bonnville, J. K. (2013). *The impact of two response-to-intervention tier 2 literacy programs on middle level achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3573287) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Campbell, J. P. (2000). *A comparison of computerized and traditional instruction in the area of elementary reading* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9966683) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Falke, T. R. (2013). *The effects of implementing a computer-based reading support program on the reading achievement of sixth graders* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3509372) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Gallagher, E. M. (1996). *Utilization of an ILS to increase reading comprehension* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9733534) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

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 group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Parmer, R. M. (2011). *SuccessMakerRTM software—the effects on reading fluency and reading comprehension: A true experimental design* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3492241) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

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 group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Additional sources:

Underwood, J., Cavendish, S., Dowling, S., Fogelman, K., & Lawson, T. (1994). *Integrated learning systems in UK schools, 1994*. Leicester, UK: Leicester University.

Underwood, J., Cavendish S., Dowling S., & Lawson, T. (1996). *Integrated learning systems: A study of sustainable learning gains. Final report*. Leicester, UK: Leicester University.

Studies that are ineligible for review using the Adolescent Literacy Evidence Review Protocol

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 an eligible design.

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

Chapin, S. (2011). *Teacher and student evaluations on the use of digital courseware and computerized assessment for adolescent reading comprehension intervention* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3434539) The study is ineligible for review because it does not use an eligible design.

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 aligned with 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 is out of scope of the protocol.

Education Commission of the States. (1999). *SuccessMaker.* Denver, CO: Author. The study is ineligible for review because it does not use an eligible design.

Gee, A. P. (2009). *An investigation of the impact of SuccessMaker on reading and math achievement at an elementary school* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3326592) The study is ineligible for review because it does not use a sample aligned with the protocol.

Hauswirth, K. (2006). Using comprehensive digital learning systems. *Learning & Leading with Technology, 34*(3), 32–33. The study is ineligible for review because it is out of scope of the protocol.

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 does not use an eligible design.

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 does not use an eligible design.

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 is ineligible for review because it does not use an eligible design.

Metis Associates. (1999). *Community school district six [New York City schools] integrated technology reading support project: Third year evaluation report*. New York: Author. The study is ineligible for review because it does not use an eligible design.

Additional sources:

Metis Associates. (1996). *Community school district 6 [New York City schools] integrated technology reading support project: First year evaluation report*. New York: Metis Associates.

Metis Associates. (1998). *Community school district 6 [New York City schools] integrated technology reading support project: Second year evaluation report*. New York: Metis Associates.

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 is out of scope of the protocol.

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 is out of scope of the protocol.

Mills, S. C., & Ragan, T. R. (1998, February). *An implementation model for integrated learning systems*. Paper presented at the National Convention of the Association for Educational Communications and Technology (AECT), St. Louis, MO. The study is ineligible for review because it is out of scope of the protocol.

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 is out of scope of the protocol.

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 an eligible design.

Pearson Education, Inc. (2010a). *Edinburg Consolidated Independent School District, Edinburg, Texas*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2010b). *Lake Weir High School, Ocala, Florida*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2010c). *Our Lady of Unity Catholic School, Kansas City, Kansas*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2010d). *Stout Middle School, Dearborn, Michigan*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2010e). *Tuttle Elementary School, Sarasota, Florida*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011a). *Atlanta Elementary School, Atlanta, Texas*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011b). C. E. *Rose Elementary School, Tucson, Arizona*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011c). *Cora Thomas Elementary School, Richmond, Texas*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011d). *Crestdale Elementary School, Richmond, Indiana*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011e). *Eliot Elementary School, Gilroy, California*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011f). *Great Falls Public Schools, Great Falls, Montana*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011g). *Hunt Elementary School, Cuero, Texas*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011h). *James Stephens International Academy, Fort Myers, Florida*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011i). *John F. Kennedy Magnet School, Port Chester, New York*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011j). *Killeen Independent School District, Killeen, Texas*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011k). *Lincoln Elementary School, Lawton, Oklahoma*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011l). *Rains Independent School District, Emory, Texas*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011m). *Reuther and West Middle Schools, Rochester Hills, Michigan*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011n). *Roseville Community Schools, Roseville, Michigan*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2011o). *South Colonie School District, Albany, New York*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2012a). *Cox Elementary School, Xenia, Ohio*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2012b). *John Page Middle School, Madison Heights, Michigan*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2012c). *Mt. Morris Consolidated Schools, Mt. Morris, Michigan*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2012d). *Radley Elementary School, East Helena, Montana*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2012e). *Village Charter School, Trenton, New Jersey*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education, Inc. (2012f). *West Side Elementary School, Sidney, Montana*. Upper Saddle River, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Pearson Education Technologies. (2002). *SuccessMaker evidence of effectiveness: Selected evaluation summaries*. Scottsdale, AZ: Author. The study is ineligible for review because it does not use an eligible design.

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). Available from ProQuest Dissertations and Theses database. (UMI No. 9838849) The study is ineligible for review because it does not use an eligible design.

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 an eligible design.

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 an eligible design.

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 an eligible design.

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 out of the scope of the protocol.

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79*(4), 1391–1466. The study is ineligible for review because it does not use an eligible design.

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 an eligible design.

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 an eligible design.

Tingey, B., & Thrall, A. (2002). High stakes management. *Multimedia Schools, 9*(2), 1–7. The study is ineligible for review because it does not use an eligible design.

Washington, T. M. O. (2012). *A study of elementary students' use of a technology program and their performance on a standardized test* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3548127) The study is ineligible for review because it does not use an eligible design.

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 does not use an eligible design.

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, Pensacola. The study is ineligible for review because it does not use an eligible design.

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 out of the scope of the protocol.

Appendix A: Research details for Gatti (2011)

Gatti, G. (2011). *Pearson SuccessMaker reading efficacy study: 2010-11 final report*. Pittsburgh, PA: Gatti Evaluation Inc.

Table A. Summary of findings **Meets WWC group design standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Comprehension	1,094 students	+3	No
General literacy achievement	1,087 students	–3	No

Setting

The study was conducted in eight urban and suburban school districts located in seven states: Arizona, California, Indiana, Kansas, Michigan, Missouri, and Texas.

Study sample

Nine schools participated in the study. The schools had to meet the following conditions: they had to (1) have no prior exposure to *SuccessMaker®*; (2) have at least two teachers per study grade level; (3) be geographically diverse; (4) agree that teachers would uphold random assignment; and (5) agree that intervention group classrooms would have their students use *SuccessMaker®* at least 1 hour per week.

English language arts classes (or sections) within the selected schools and grade levels were randomly assigned to either the intervention or the business-as-usual comparison group.[7] The fifth-grade sample included 16 classrooms implementing *SuccessMaker®* and 14 implementing the school's regular English language arts program. The seventh-grade sample included 11 classrooms implementing *SuccessMaker®* and nine implementing the school's regular English language arts program.

Of the 641 fifth-grade students that participated in the study, 342 received *SuccessMaker®* and 299 received the school's regular English language arts program. Of the 453 seventh-grade students that participated in the study, 254 received *SuccessMaker®* and 199 received the school's regular English language arts program.

About 48% of the total sample were male, 39% were minority (about 23.6% Hispanic and 15.8% African American), and 100% received free or reduced-price lunch.

Intervention group

SuccessMaker® is an adaptive, computer-based learning program which includes an instructional management system, formative assessments, a progress reporting system, and individualized reading curriculum resources for elementary and middle school instruction. For this study, the program was typically implemented with the entire class in a computer laboratory during the regular reading instruction time. Intervention group students were expected to use the software for at least 1 hour each week. Over the course of the school year, most teachers went to the lab two or three times a week, with a median time of 22 hours for fifth-grade classes and 18 hours for seventh-grade classes.

Comparison group

Comparison classes received the business-as-usual English language arts instruction, which generally did not involve computer-based instruction. The majority of students in fifth grade (62%) received instruction from four widely-used published reading programs, while the remainder received instruction from non-published teacher-developed curricula. In contrast, the majority of students in seventh grade (63%) received instruction from non-published, largely teacher-created curricula, while the remainder received instruction from three different widely-used published literacy programs.

Outcomes and measurement

Assessments were administered at the onset of the intervention and in the last month of the school year. Outcomes in the comprehension domain included Group Reading Assessment and Diagnostic Evaluation (GRADE) Overall Score and three subtest scores of the GRADE; Passage Comprehension, Sentence Comprehension, and Vocabulary. One measure in the reading fluency domain was administered, the AIMSweb Reading Curriculum-Based Measurement. Supplemental findings for the three GRADE subtest outcomes and for subgroups of students on the GRADE overall score and the AIMSweb outcome are presented in Appendix D.[8] The supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of these outcome measures, see Appendix B.

The study also included a researcher-designed student reading attitude survey. However, this outcome is not eligible for review based on the Adolescent Literacy review protocol (version 3.0).

Support for implementation

Teachers received 1 day of initial training, some before school started and some in the second month of the school year. The teachers also received a 3-hour follow-up training and additional assistance from Pearson when needed.

Appendix B: Outcome measures for each domain

Comprehension	
Group Reading Assessment and Diagnostic Evaluation (GRADE) Overall Score	This test is an untimed standardized group-administered assessment of reading comprehension published by Pearson Assessment. The overall score is based on scores on the Passage Comprehension, Sentence Comprehension, and Vocabulary subtests. Two parallel forms were administered in the study for each grade level: Form A at baseline and Form B at the end of school year. The test has reliability coefficients of .93 and above for the overall score (as cited in Gatti, 2011).
Reading comprehension	
GRADE Passage Comprehension Subtest	This 19-item assessment measures students' understanding of extended text through explicit and implicit multiple-choice questions requiring questioning, predicting, summarizing, and clarifying information from several paragraphs. Reliability coefficients ranged from .77 to .85 for the three GRADE subtests (as cited in Gatti, 2011). This outcome is included as a supplemental finding.
GRADE Sentence Comprehension Subtest	This 30-item assessment measures students' ability to understand a given sentence as a complete thought, by asking testers to identify the missing word in a sentence, the grammatical complexity of which varies by reading level. Reliability coefficients ranged from .77 to .85 for the three GRADE subtests (as cited in Gatti, 2011). This outcome is included as a supplemental finding.
Vocabulary development	
GRADE Vocabulary Subtest	This assessment measures students' knowledge and understanding of words. Students are asked to select the correct meaning of targeted words presented in short sentences or phrases. The subtest is comprised of 35 questions for grade 5 (level 5) and 40 questions for grade 7 (level M). Reliability coefficients ranged from .77 to .85 for the three GRADE subtests (as cited in Gatti, 2011). This outcome is included as a supplemental finding.
Reading fluency	
AIMSweb Reading Curriculum-Based Measurement	This assessment measures the number of words that a student reads accurately from a specified passage. The student reads three passages, each for 1 minute, and the middle score is recorded. Bivariate correlation coefficients between the three passages for all grades and testing sessions ranged from .89 to .93. The test is published by Pearson Assessment (as cited in Gatti, 2011).

Appendix C.1: Findings included in the rating for the comprehension domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Gatti, 2011[a]								
Group Reading Assessment and Diagnostic Evaluation (GRADE) Overall Score	Grade 5	30 classrooms/ 641 students	60.59 (12.93)	60.03 (12.75)	0.56	0.04	+2	< .00
GRADE Overall Score	Grade 7	20 classrooms/ 453 students	54.56 (14.00)	52.70 (16.07)	1.86	0.12	+5	< .00
Domain average for comprehension (Gatti, 2011)						**0.08**	**+3**	**Not statistically significant**
Domain average for comprehension across all studies						**0.08**	**+3**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Gatti (2011), data reported in the table were obtained through author queries. The *p*-values presented here were reported in the original study. A correction for clustering was needed for all outcomes and resulted in WWC-computed *p*-values of .81 and .57 for grades 5 and 7, respectively; therefore, the WWC does not find any of the results to be statistically significant. The WWC calculated the program group means using a difference-in-differences approach (see the WWC Procedures and Standards Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. This study is characterized as having no discernible effects because the estimated impacts of the intervention on outcomes in the comprehension domain for fifth- and seventh-grade students were neither statistically significant nor large enough to be substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix C.2: Findings included in the rating for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Gatti, 2011[a]								
AIMSweb Reading Curriculum-Based Measurement	Grade 5	30 classrooms/ 639 students	152.20 (38.01)	156.01 (37.63)	–3.81	–0.10	–4	< .00
AIMSweb Reading Curriculum-Based Measurement	Grade 7	20 classrooms/ 448 students	164.47 (34.24)	165.37 (38.03)	–0.90	–0.02	–1	> .05
Domain average for reading fluency (Gatti, 2011)						**–0.06**	**–3**	**Not statistically significant**
Domain average for reading fluency across all studies						**–0.06**	**–3**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Gatti (2011), data reported in the table were obtained through author queries. The *p*-values presented here were reported in the original study. A correction for clustering was needed for all outcomes and resulted in WWC-computed *p*-values of .91 and .58, for grades 5 and 7, respectively; therefore, the WWC does not find any of the results to be statistically significant. The WWC calculated the program group mean using a difference-in-differences approach (see the WWC Procedures and Standards Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. This study is characterized as having no discernible effects because the estimated impacts of the intervention on outcomes in the reading fluency domain for fifth- and seventh-grade students were neither statistically significant nor large enough to be substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D.1: Description of supplemental findings for the comprehension domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Gatti, 2011[a]								
Group Reading Assessment and Diagnostic Evaluation (GRADE) Passage Comprehension Subtest	Grade 5	30 classrooms/ 641 students	23.40 (4.89)	23.04 (4.97)	0.36	0.07	+3	< .00
GRADE Sentence Comprehension Subtest	Grade 5	30 classrooms/ 641 students	15.81 (3.48)	15.48 (3.35)	0.33	0.10	+4	< .00
GRADE Vocabulary Subtest	Grade 5	30 classrooms/ 641 students	21.39 (6.01)	21.51 (6.09)	–0.12	–0.02	–1	.34
GRADE Passage Comprehension Subtest	Grade 7	30 classrooms/ 453 students	20.17 (5.38)	19.78 (6.02)	0.39	0.07	+3	< .02
GRADE Sentence Comprehension Subtest	Grade 7	20 classrooms/ 453 students	14.01 (3.85)	13.29 (4.32)	0.72	0.18	+7	< .00
GRADE Vocabulary Subtest	Grade 7	20 classrooms/ 453 students	20.37 (6.16)	19.63 (7.09)	0.74	0.11	+4	< .00
Grade 5 subgroups								
Low achieving								
GRADE Overall Score	Grade 5	30 classrooms/ 130 students	44.36 (10.11)	45.34 (11.54)	–0.98	–0.09	–4	> .05
African American								
GRADE Overall Score	Grade 5	30 classrooms/ 44 students	51.15 (13.94)	50.65 (13.51)	0.50	0.04	+1	> .05
Hispanic								
GRADE Overall Score	Grade 5	30 classrooms/ 207 students	57.43 (10.82)	55.48 (11.23)	1.75	0.16	+6	< .05
Male								
GRADE Overall Score	Grade 5	30 classrooms/ 298 students	61.28 (12.55)	60.31 (12.64)	0.97	0.08	+3	> .05
Female								
GRADE Overall Score	Grade 5	30 classrooms/ 343 students	60.01 (13.30)	59.80 (12.87)	0.21	0.02	+1	> .05
Reduced-price lunch								
GRADE Overall Score	Grade 5	30 classrooms/ 286 students	57.26 (11.32)	55.32 (12.53)	1.94	0.16	+6	< .05
Grade 7 subgroups								
Low achieving								
GRADE Overall Score	Grade 7	20 classrooms/ 144 students	39.89 (9.96)	37.33 (9.95)	2.56	0.26	+10	< .05
African American								
GRADE Overall Score	Grade 7	20 classrooms/ 129 students	46.61 (13.64)	45.13 (14.45)	1.48	0.10	+4	< .05

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Male								
GRADE Overall Score	Grade 7	20 classrooms/ 222 students	60.22 (14.70)	57.70 (17.43)	2.52	0.16	+6	< .05
Female								
GRADE Overall Score	Grade 7	20 classrooms/ 231 students	54.96 (13.22)	53.66 (14.68)	1.30	0.09	+4	< .05
Reduced-price lunch								
GRADE Overall Score	Grade 7	20 classrooms/ 239 students	47.50 (13.76)	46.08 (14.73)	1.42	0.10	+4	< .05

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards without reservations, but do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

[a] For Gatti (2011), data reported in the table were obtained through author queries. The *p*-values presented here were reported in the original study. A correction for clustering was needed for all outcomes and resulted in WWC-computed *p*-values ranging from .31 to 92; therefore, the WWC does not find any of the results to be statistically significant. The WWC calculated the program group mean using a difference-in-differences approach (see the WWC Procedures and Standards Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

Appendix D.2: Description of supplemental findings for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Gatti, 2011[a]								
Grade 5 subgroups								
Low achieving								
AIMSweb Reading Curriculum-Based Measurement	Grade 5	30 classrooms/ 129 students	116.63 (29.70)	118.85 (31.70)	–2.22	–0.07	–3	> .05
African American								
AIMSweb Reading Curriculum-Based Measurement	Grade 5	30 classrooms/ 44 students	143.74 (34.71)	133.52 (43.23)	10.22	0.25	+10	< .05
Hispanic								
AIMSweb Reading Curriculum-Based Measurement	Grade 5	30 classrooms/ 205 students	139.32 (32.59)	145.04 (31.32)	–5.72	–0.18	–7	< .05
Male								
AIMSweb Reading Curriculum-Based Measurement	Grade 5	30 classrooms/ 297 students	153.23 (37.50)	156.95 (38.01)	–3.72	–0.10	–4	< .05

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Female								
AIMSweb Reading Curriculum-Based Measurement	Grade 5	30 classrooms/ 342 students	151.71 (38.33)	155.23 (37.42)	–3.52	–0.09	–4	< .05
Reduced-price lunch								
AIMSweb Reading Curriculum-Based Measurement	Grade 5	30 classrooms/ 284 students	138.40 (33.52)	141.69 (33.98)	–3.29	–0.10	–4	< .05
Grade 7 subgroups								
Low achieving								
AIMSweb Reading Curriculum-Based Measurement	Grade 7	20 classrooms/ 141 students	134.57 (28.76)	139.23 (27.80)	–4.66	–0.16	–7	> .05
African American								
AIMSweb Reading Curriculum-Based Measurement	Grade 7	20 classrooms/ 128 students	150.93 (33.92)	153.28 (32.11)	–2.35	–0.07	–3	< .05
Hispanic								
AIMSweb Reading Curriculum-Based Measurement	Grade 7	20 classrooms/ 51 students	146.74 (34.99)	149.71 (23.31)	–2.97	–0.10	–4	< .05
Male								
AIMSweb Reading Curriculum-Based Measurement	Grade 7	20 classrooms/ 219 students	157.82 (32.99)	159.42 (36.54)	–1.60	–0.05	–2	> .05
Female								
AIMSweb Reading Curriculum-Based Measurement Evaluation	Grade 7	20 classrooms/ 229 students	170.84 (35.53)	170.97 (38.73)	–0.13	0	0	> .05
Reduced-price lunch								
AIMSweb Reading Curriculum-Based Measurement	Grade 7	20 classrooms/ 23 students	151.06 (32.13)	152.51 (30.55)	–1.45	–0.05	–2	> .05

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards without reservations but do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

[a] For Gatti (2011), data reported in the table were obtained through author queries. The *p*-values presented here were reported in the original study. A correction for clustering was needed for all outcomes and resulted in WWC-computed *p*-values ranging from .46 to .99; therefore, the WWC does not find any of the results to be statistically significant. The WWC calculated the program group mean using a difference-in-differences approach (see the WWC Procedures and Standards Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

Endnotes

[1] The descriptive information for this program was obtained from a publicly available source: the program's website (http://www.pearsonschool.com, downloaded November 2014). The WWC requests distributors review the program description sections for accuracy from their perspective. The program description was provided to the distributor in November 2014; however, the WWC received no response. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.

[2] The literature search reflects documents publicly available by October 2014, and represents an updated systematic review and assessment of the available studies. The previous intervention report was released in June 2009 and included studies reviewed under the WWC evidence standards, (version 1.0). This report has been updated to include reviews of 34 studies that have been released since 2009. Of the additional studies, 30 were not within the scope of the review protocol for the Adolescent Literacy topic area, and three were within the scope of the review protocol but did not meet WWC group design standards. One study met WWC group design standards without reservations, and findings from this study are highlighted in this report. In addition, three studies (Beattie, 2000; Campbell, 2000; and Gallagher, 1996), which met WWC design standards with reservations in the previous report, do not meet WWC design standards in this report. These revised dispositions are due to changes in the design standards and the Adolescent Literacy review protocol. In particular, for Campbell (2000), a statistical adjustment for baseline differences between 0.05 and 0.25 is required in order to meet the baseline equivalence requirement. For Gallagher (1996), baseline differences either required a statistical adjustment for outcome analyses in grades 4, 5, and the full sample, or exceeded 0.25 standard deviation for outcome analyses in grades 6 and 7. For Beattie (2000), the randomized controlled trial analysis included outcomes that had a combination of overall and differential attrition rates that exceeded the WWC standards (version 3.0), and the subsequent analytic intervention and comparison groups were not shown to be equivalent. A complete list and disposition of all studies reviewed are provided in the references. The studies in this report were reviewed using the Standards from the WWC Procedures and Standards Handbook (version 3.0), along with those described in the Adolescent Literacy review protocol (version 3.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 19. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies.

[4] Suppes, P., & Zanotti, M. (1996). Mastery learning of elementary mathematics: Theory and data. In P. Suppes & M. Zanotti (Eds.), *Foundations of probability with applications* (pp. 149–188). New York: Cambridge University Press.

[5] The study also included third-grade students, which are out of scope of the Adolescent Literacy review protocol (version 3.0).

[6] Intact classrooms of students (clusters), as opposed to individual students, were randomly assigned to intervention conditions.

[7] In some schools, fifth- or seventh-grade teachers taught multiple sections of English language arts, and each section was randomly assigned, as there are references in the study to some teachers teaching both *SuccessMaker®* and comparison sections.

[8] Supplemental findings are presented for two domains in Appendix D. For the comprehension domain, findings are presented for the three GRADE subtests (Passage Reading Comprehension, Sentence Comprehension, and Vocabulary). For both domains, comprehension and reading fluency, findings are also presented for the following subgroups: low-achieving students, African-American students, Hispanic students, male students, female students, and students who qualified for reduced-price lunch. Note that analyses for seventh-grade Hispanic students met WWC group design standards with reservations for the AIMSweb Words Read Correctly outcome because there was high student attrition, but the intervention and comparison groups were equivalent at baseline. Analyses in the comprehension domain for seventh-grade Hispanic students did not meet WWC group design standards because, due to high student attrition, equivalence of the analytic intervention and comparison group is necessary and not demonstrated.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2015, November). *Adolescent Literacy intervention report: SuccessMaker®*. Retrieved from http://whatworks.ed.gov

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC group design standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC group design standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows 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, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important positive effects..
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 19.

Improvement index Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from –50 to +50.

Intervention An educational program, product, practice, or policy aimed at improving student outcomes.

Intervention report A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.

Rating of effectiveness The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 19.

Single-case design A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Glossary of Terms

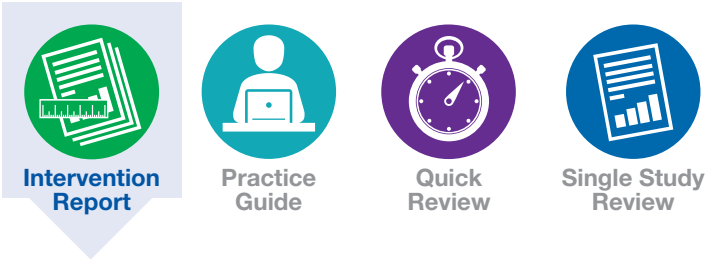
Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Systematic review A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.



An **intervention report** summarizes the findings of high-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.

This intervention report was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.