Carnegie Learning Curricula and Cognitive Tutor®

Report Contents

Program Description[1]

Carnegie Learning Curricula and Cognitive Tutor®, published by Carnegie Learning, is a secondary math curricula that offers textbooks and interactive software to provide individualized, self-paced instruction based on student needs. The program includes pre-Algebra, Algebra I, Algebra II, and Geometry, as well as a three-course series that integrates numeric, algebraic, geometric, and statistical content. The developer indicates that the program is aligned with most state standards and the standards set by the National Council of Teachers of Mathematics. The program can be customized to meet other state-specific standards.

Research[2]

The What Works Clearinghouse (WWC) identified six studies of *Carnegie Learning Curricula and Cognitive Tutor®* that both fall within the scope of the High School Mathematics topic area and meet WWC evidence standards. Three studies meet WWC standards without reservations and three studies meet WWC standards with reservations, and together, they included 2,553 high school students from 39 schools.[3]

The WWC considers the extent of evidence for *Carnegie Learning Curricula and Cognitive Tutor®* on the math performance of high school students to be medium to large for the mathematics achievement domain, the only domain for the High School Mathematics review protocol.

Effectiveness

Carnegie Learning Curricula and Cognitive Tutor® was found to have mixed effects on mathematics achievement for high school students.

Table 1. Summary of findings[4]

		Improvement index *(percentile points)*				
Outcome domain	Rating of effectiveness	Average	Range	Number of studies	Number of students	Extent of evidence
Mathematics achievement	Mixed effects	–1	–8 to +36	6	2,553	Medium to large

Program Information

Background

Carnegie Learning Curricula and Cognitive Tutor® was developed and is distributed by Carnegie Learning Inc. Address: Carnegie Learning Inc., Frick Building, 20th Floor, 437 Grant St., Pittsburgh, PA, 15219. Email: info@carnegielearning.com. Web: http://www.carnegielearning.com. Phone: (888) 851-7094.

Program details

Carnegie Learning Curricula and Cognitive Tutor® can be implemented using a textbook, adaptive software, or as a part of a blended implementation that combines textbook and software activities. In a blended implementation, three periods per week are spent using the *Carnegie Learning Curricula and Cognitive Tutor*® text for classroom activities. The textbooks aim to foster a collaborative classroom environment in which students develop skills to work cooperatively to solve problems, participate in investigations, and propose and compare solutions. Two periods per week are spent in the computer lab using the *Carnegie Learning Curricula and Cognitive Tutor*® software. Students learn with the adaptive software at their own pace. The math problems are designed to emphasize connections between verbal, numeric, graphic, and algebraic representations.

Cost

The cost for *Carnegie Learning Curricula and Cognitive Tutor*® varies depending the number of students who will be using the program and the type of implementation (blended, textbook only, or software only). Cost information is available from the developer.

Research Summary

The WWC identified 27 studies that investigated the effects of *Carnegie Learning Curricula and Cognitive Tutor®* on math performance for high school students.

The WWC reviewed 11 of those studies against group design evidence standards. Three studies (Cabalo, Jaciw, & Vu, 2007; Campuzano, Dynarski, Agodini, & Rall, 2009; & Pane, McCaffrey, Slaughter, Steele, & Ikemoto, 2010) are randomized controlled trials that meet WWC evidence standards without reservations, and three studies (Shneyderman, 2001; Smith, 2001; & Wolfson, Koedinger, Ritter, & McGuire, 2008) are randomized controlled trials or quasi-experimental designs that meet WWC evidence standards with reservations. These six studies are summarized in this report. Five studies do not meet WWC evidence standards. The remaining 16 studies do not meet WWC eligibility screens for review in this topic area. Citations for all 27 studies are in the References section, which begins on p. 7.

Table 2. Scope of reviewed research[5]

Grade	9, 10, 11, 12
Delivery method	Whole class
Program type	Curriculum
Studies reviewed	27 studies
Group design studies that meet WWC standards	
• without reservations	3 studies
• with reservations	3 studies

Summary of studies meeting WWC evidence standards without reservations

Cabalo et al. (2007) randomly assigned 22 classrooms to receive either the *Carnegie Learning Curricula and Cognitive Tutor®* Algebra I program or the standard curriculum. The study took place in six schools in Hawaii and included nine teachers. The analysis sample consisted of 182 intervention students and 162 comparison students who had taken both the pretest (in fall 2005) and the posttest (in May 2006).

Campuzano et al. (2009) randomly assigned teachers in high-poverty schools to intervention and comparison groups as part of a national study of software products. During the second year of the study (presented in this report), the *Carnegie Learning Curricula and Cognitive Tutor®* Algebra I program was implemented in nine schools in four districts. Nine teachers were randomly assigned to use the intervention, and nine were assigned to the comparison condition and used traditional instructional methods. The fall and spring tests were administered to 145 intervention students and 131 comparison students in eighth and ninth grades.

Pane et al. (2010) randomly assigned students in eight high schools in the Baltimore County Public School district to receive either the *Carnegie Learning Curricula and Cognitive Tutor®* Geometry curriculum or the standard geometry curriculum. The study took place over three academic school years (2005–06 to 2007–08). During each year, participating schools held a morning and an afternoon geometry class, with an intervention and a comparison classroom offered in each period. Two teachers from each participating school were randomly assigned between the morning classrooms and then taught the opposite curriculum in the afternoon period. This allowed each teacher to deliver the intervention and comparison curricula across the two class periods. The analytic sample included 60 classrooms (30 intervention and 30 comparison) and 699 students (348 intervention and 351 comparison). At the end of the academic year, the Baltimore County Public School district geometry assessment was administered to the student sample.

Summary of studies meeting WWC evidence standards with reservations

Shneyderman (2001) conducted a quasi-experiment in six senior high schools in Miami–Dade County, Florida that implemented the *Carnegie Learning Curricula and Cognitive Tutor®* Algebra I program and had an operational computer lab during the 2000–01 school year. For each school, two teachers were randomly selected from all teachers using the intervention. One class for each teacher was randomly selected to form an intervention sample of 12 classrooms. The comparison sample was composed of 12 sampled nonintervention Algebra I classrooms in the same six schools. The analyses were conducted on 276 intervention and 382 comparison students in the ninth and tenth grades.

Smith (2001) conducted a randomized controlled trial that was compromised because the analysis did not include all students that were randomly assigned—the analysis excluded students who were randomly assigned but did not complete their three-semester Algebra I requirement. However, the study's analysis did demonstrate baseline equivalence of the analysis sample on a pretest and made necessary statistical adjustments in estimating program effects. Therefore, the study meets WWC evidence standards with reservations. The study involved all students in seven high schools in Virginia Beach City Public Schools who completed a three-semester Algebra I sequence during the 1999–2000 and 2000–01 school years. Students were assigned to either a sequence in which the math teacher was willing to implement the *Carnegie Learning Curricula and Cognitive Tutor®* program (229 students) or a sequence with the traditional curriculum (216 students).

Wolfson et al. (2008) conducted a quasi-experimental study during the 1993–94 school year with 26 Algebra I classrooms across three high schools in the Pittsburgh Public Schools District. Students in the intervention classrooms used an early version of *Carnegie Learning Curricula and Cognitive Tutor®*, which at the time was referred to as the Pittsburgh Urban Mathematics Project curriculum plus Practical Algebra Tutor program.[6] Comparison students received their usual Algebra I curriculum. Intervention and comparison classes were matched on the basis of student math grades from the previous school year. At the end of the spring semester, all students were administered the Iowa Test of Basic Skills assessment. In addition, students were administered two assessments randomly selected from a set of three assessments. These three assessments included a subset of the Math SAT and two researcher-developed tests: Problem Situation and Representation. This WWC report presents findings for only the Math SAT and Problem Situation outcomes because the intervention and comparison groups in the analysis samples used for the other two outcomes were not equivalent in baseline mathematics achievement.

Effectiveness Summary

The WWC review of *Carnegie Learning Curricula and Cognitive Tutor®* for the High School Mathematics topic includes student outcomes in one domain: mathematics achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Carnegie Learning Curricula and Cognitive Tutor®* on high school students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 23.

Summary of effectiveness for the mathematics achievement domain

Six studies reported findings in the mathematics achievement domain.

Cabalo et al. (2007) did not report a statistically significant difference between the *Carnegie Learning Curricula and Cognitive Tutor®* group and the comparison group when using the Northwest Evaluation Association (NWEA) Algebra End-of-Course Achievement Level Test/Measures of Academic Progress as an outcome measure. The effect size was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as an indeterminate effect.

Campuzano et al. (2009) did not report a statistically significant difference between the *Carnegie Learning Curricula and Cognitive Tutor®* group and the comparison group on the Educational Testing Service (ETS) Algebra I End-of-Course Assessment. The effect size was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as an indeterminate effect.

Pane et al. (2010) reported, and the WWC confirmed, a statistically significant negative difference between the *Carnegie Learning Curricula and Cognitive Tutor®* Geometry group and the comparison group on the Baltimore County Public School district geometry assessment. The WWC characterizes this study finding as a statistically significant negative effect.

Shneyderman (2001) did not report a statistically significant difference between the *Carnegie Learning Curricula and Cognitive Tutor®* group and the comparison group on the Florida Comprehensive Assessment Test (FCAT) Norm-Referenced Component. The effect size was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as an indeterminate effect.

Smith (2001) did not report a statistically significant difference between the *Carnegie Learning Curricula and Cognitive Tutor®* group and the comparison group on the Virginia Standards of Learning (SOL) Algebra Assessment. The effect size was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as an indeterminate effect.

Wolfson et al. (2008) reported, and the WWC confirmed, a statistically significant positive difference between the *Carnegie Learning Curricula and Cognitive Tutor®* group and the comparison group on the Problem Situation assessment. The study also reported a statistically significant positive difference between the *Carnegie Learning Curricula and Cognitive Tutor®* group and the comparison group on a subset of the Math SAT. However, when this result was adjusted for clustering, the WWC found that this difference was no longer statistically significant. The effect sizes of both differences were large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as a statistically significant positive effect.

Thus, for the mathematics achievement domain, one study showed a statistically significant negative effect, one study showed a statistically significant positive effect, and four studies showed indeterminate effects. This results in a rating of mixed effects, with a medium to large extent of evidence.

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

Rating of effectiveness	Criteria met
Mixed effects *Evidence of inconsistent effects.*	In the six studies that reported findings, the estimated impact of the intervention on outcomes in the *mathematics achievement domain* was inconsistent. One study showed a statistically significant negative effect, one study showed a statistically significant positive effect, and four studies showed indeterminate effects.
Extent of evidence	**Criteria met**
Medium to large	One study that included 2,553 students in 39 schools reported evidence of effectiveness in the *mathematics achievement domain.*

References

Studies that meet WWC evidence standards without reservations

Cabalo, J. V., Jaciw, A., & Vu, M.-T. (2007). *Comparative effectiveness of Carnegie Learning's Cognitive Tutor Algebra I curriculum: A report of a randomized experiment in the Maui School District.* Palo Alto, CA: Empirical Education, Inc.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Additional source:

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., … Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of Cognitive Tutor geometry. *Journal of Research on Educational Effectiveness, 3*(3), 254–281.

Studies that meet WWC evidence standards with reservations

Shneyderman, A. (2001). *Evaluation of the Cognitive Tutor Algebra I program.* Unpublished manuscript. Miami, FL: Miami–Dade County Public Schools, Office of Evaluation and Research.

Smith, J. E. (2001). *The effect of the Carnegie Algebra Tutor on student achievement and attitude in introductory high school algebra* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University, Blacksburg.

Wolfson, M., Koedinger, K., Ritter, S., & McGuire, C. (2008). *Cognitive Tutor Algebra I: Evaluation of results (1993–1994)*. Pittsburgh, PA: Carnegie Learning, Inc.

Additional source:

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*(1), 30–43.

Studies that do not meet WWC evidence standards

Arbuckle, W. J. (2005). *Conceptual understanding in a computer-assisted Algebra 1 classroom.* Norman: University of Oklahoma. The study does not meet WWC evidence standards because it does not provide adequate information to determine whether it uses an outcome that is valid or reliable.

Bibi, T. (2010). *Analysis of Cognitive Tutor Geometry Curriculum*. Ames: Iowa State University. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

Salden, R., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science, 38*(3), 289–307. The study does not meet WWC evidence standards because it is a randomized controlled trial in which the combination of overall and differential attrition rates exceeds WWC standards for this area, and the subsequent analytic intervention and comparison groups are not shown to be equivalent.

Sarkis, H. (2004). *Cognitive Tutor Algebra I program evaluation: Miami–Dade County Public Schools.* Lighthouse Point, FL: The Reliability Group. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Voloshin, D. (2010). An evaluation of a computer-assisted, remedial algebra curriculum on attitudes and performance of ninth grade English learners. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 70*(7-A), 2431. The study does not meet WWC evidence standards because the measures of effec-

tiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

Studies that are ineligible for review using the High School Mathematics Evidence Review Protocol

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*(2), 147. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*(2), 101–128. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research, 19*(2), 185–224. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.

Carnegie Learning, Inc. (2010). *Cognitive Tutor effectiveness.* Pittsburgh, PA: Author. Retrieved from http://www.carnegielearning.com/static/web_docs/2010_Cognitive_Tutor_Effectiveness.pdf. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Corbett, A. T. (2001). *Cognitive Tutor results report: 7th grade*. Pittsburgh, PA: Carnegie Learning, Inc. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Corbett, A. T. (2002). *Cognitive Tutor results report: 8th & 9th grade.* Pittsburgh, PA: Carnegie Learning, Inc. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Dickensheets, K. (2001). Not just computers: Learning by doing. *Multimedia Schools, 8*(1), 40. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Epper, R., & Delott Baker, E. (2009). *Technology solutions for developmental math: An overview of current and emerging practices.* Paper prepared for the Bill & Melinda Gates Foundation. Retrieved from http://www.gatesfoundation.org/learning/Documents/technology-solutions-for-developmental-math-jan-2009.pdf. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Plano, G. S., Ramey, M., & Achilles, C. M. (2005). *Implications for student learning using a technology-based algebra program in a ninth-grade algebra course.* Unpublished manuscript. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Additional source:

Plano, G. S. (2004). The effects of the Cognitive Tutor Algebra on student attitudes and achievement in a 9th-grade algebra course. *Dissertation Abstracts International* 65(04A), 47-291. (AAI3130130)

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin and Review, 14*(2), 249–255. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Ritter, S., Haverty, L., Koedinger, K., Hadley, W., & Corbett, A. (2008). Integrating intelligent software tutors with the math classroom. In G. Blume & K. Heid (Eds.), *Research on technology and the teaching and learning of mathematics: Vol. 2. Cases and perspectives.* Charlotte, NC: Information Age Publishing. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Ritter, S., Kulikowich, J., Lei, P., McGuire, C., & Morgan, P. (2007). What evidence matters? A randomized field trial of Cognitive Tutor Algebra I. In T. Hirashima, H. U. Hoppe, & S. Shwu-Ching Young (Eds.), *Supporting learning flow through integrative technologies* (pp. 13–20). Netherlands: IOS Press. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Additional source:

Morgan, P., & Ritter, S. (2002). *An experimental study of the effect of Cognitive Tutor Algebra I on student knowledge and attitude.* Retrieved from http://www.carnegielearning.com/web_docs/morgan_ritter_2002.pdf.

Salden, R., Aleven, V., Renkl, A., & Schwonke, R. (2009). Worked examples and tutored problem solving: Redundant or synergistic forms of support. *Topics in Cognitive Science, 1,* 203–213. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Stylianou, D. A., & Shapiro, L. (2002). Revitalizing algebra: The effect of the use of a cognitive tutor in a remedial course. *Journal of Educational Media, 27*(3), 147. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Viadero, D. (2009). Reading, math software found to have little effect on scores. *Education Week, 28*(25), 8. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Zimmerman, J. E. (2004). The impact of Cognitive Tutor software on student performance in college intermediate algebra. *Dissertation Abstracts International 64*(9), 3229 A. (UMI No. 3103843) The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Appendix A.1: Research details for Cabalo et al., 2007

Cabalo, J. V., Jaciw, A., & Vu, M.-T. (2007). *Comparative effectiveness of Carnegie Learning's Cognitive Tutor Algebra I curriculum: A report of a randomized experiment in the Maui School District.* Palo Alto, CA: Empirical Education, Inc.

Table A1. Summary of findings — **Meets WWC evidence standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	6 schools/344 students	–7	No

Setting The study took place in five schools within the Maui School District and in Maui Community College, all located in Maui County, Hawaii. According to the authors, Maui County is a mixed suburban and rural community located on one of the seven islands of Hawaii. Nine teachers and 22 classrooms participated in the study. Within the participating schools, students were 32% Filipino, 28% part-Hawaiian, 11% White, 8% Japanese, 5% Hawaiian, 3% Hispanic, and 14% other. The distribution of ethnicities at Maui Community College was similar. Approximately 27% of students participated in the National School Lunch Program, and approximately 6% were designated as limited English proficient.

Study sample After an informational session with a group of teachers in the Maui School District, nine teachers volunteered to participate in a study of the effectiveness of the *Carnegie Learning Curricula and Cognitive Tutor®* Algebra I program. When possible, classes were paired based on class size and achievement level, with a coin toss determining which one of the pair would be assigned to the intervention group. Classes that were unable to be paired (when a teacher had an odd number of classes) were assigned to the intervention or comparison group by coin toss. Pre-intervention mathematics achievement data were collected in fall 2005, and a posttest evaluation was administered in May 2006. Only students with both tests were included in the analysis. Of the initial sample of 541 students (281 intervention and 260 comparison), 344 (182 intervention and 162 comparison) had both pre- and posttest scores. At the beginning of the study, students in grades 9–12 comprised 73% of the sample, with 19% in grade 8 and 7% enrolled at Maui Community College.

Intervention group Classrooms selected for the intervention group implemented the *Carnegie Learning Curricula and Cognitive Tutor®* Algebra I program. Selected classrooms utilized the intervention for six months, from October/November through the end of the 2005–06 school year.

Comparison group For the comparison classrooms, teachers continued to follow the textbook program in use at the time of study implementation, one of several branded Algebra I textbooks.

Outcomes and measurement

Student mathematics achievement was measured by the NWEA Algebra End-of-Course Achievement Level Test/Measure of Academic Progress. A paper version of the assessment was administered to participating students enrolled in the Maui School District, and a computer-adapted version of the assessment was administered to participating students enrolled at Maui Community College. For a more detailed description of these outcome measures, see Appendix B. Results from both tests were combined by the authors and in the results presented in Appendix C.1. The disaggregated results by subscale are presented in Appendix D.1; these findings do not factor into the determination of the intervention rating.

Support for implementation

Teachers utilizing the *Carnegie Learning Curricula and Cognitive Tutor®* Algebra I program received three days of professional development, led by a consultant from the developer. Teachers were observed briefly in the classroom and given an opportunity to ask the trainer questions early in the implementation period. No ongoing technical assistance was provided.

Appendix A.2: Research details for Campuzano et al., 2009

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). ***Effectiveness of reading and mathematics software products: Findings from two student cohorts.*** **Washington, DC: U.S. Department of Education, Institute of Education Sciences.**

Table A2. Summary of findings — **Meets WWC evidence standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	9 schools/276 students	–6	No

Setting

This national study of software products included an examination of algebra products. Findings from the first year of the study are pooled between *Carnegie Learning Curricula and Cognitive Tutor*® and another intervention. Therefore, this report includes findings from the second year of the study only, which are disaggregated by intervention. *Carnegie Learning Curricula and Cognitive Tutor*® was implemented in nine schools in four districts. Districts were located in urban and urban fringe areas, averaging 230 schools and 133,000 students. Within each of the nine study schools, one teacher was randomly assigned to use the intervention, and one was assigned to the comparison condition, with at least a pair of intervention and comparison teachers in each school. Teachers averaged 16 years of experience, and 47% had a master's degree.

Study sample

Schools were eligible to be in the study if they were in high-poverty areas, had no prior software product use, and had enough teachers in each grade. Teachers in the participating schools were randomly assigned to intervention and comparison groups, and students were allocated to classrooms based on conventional school methods. The fall and spring tests were administered to 276 students, who were age 14 on average and 51% female. Eighteen percent of students were in eighth grade, and 82% were in ninth grade.[7]

Intervention group

The intervention group consisted of nine teachers from nine schools in four school districts. The intervention was delivered as a full curriculum that covered proportional reasoning, solving linear equations and inequalities, solving systems of linear equations, analyzing data, and using polynomial functions, powers, and exponents.

Comparison group

The comparison group consisted of nine other teachers from the same schools. The students in these classes received traditional algebra instruction using standard district materials.

Outcomes and measurement

The study team administered the ETS Algebra I End-of-Course Assessment. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

Teachers in the intervention group received four days of initial training in the summer of 2004 at a school or district location. They were given information on classroom management and curriculum, along with opportunities to practice using the product. Phone and email support was available.

Appendix A.3: Research details for Pane et al., 2010

Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of Cognitive Tutor geometry. *Journal of Research on Educational Effectiveness, 3*(3), 254–281.

Table A3. Summary of findings **Meets WWC evidence standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	8 schools/699 students	–8	Yes

Setting The study was conducted in eight high schools in the Baltimore County Public School district. Of these schools, two participated in each of three academic years (2005–06 to 2007–08), three participated for two years, and three participated for one year. The district included 25 schools with a student enrollment of about 32,000 students. Thirty-nine percent of students in the district were minorities, and 17% were eligible for free or reduced-price meals.

Study sample In each academic year, two geometry teachers at each participating school took part in the study. Each school scheduled a morning geometry period and an afternoon geometry period, with each period offering an intervention classroom using the *Carnegie Learning Curricula and Cognitive Tutor®* Geometry curriculum and a comparison classroom. The two teachers were randomly assigned between the morning classrooms, and then switched the alternate curriculum for the afternoon class. This allowed each teacher to deliver both the intervention and comparison curricula. Prior to the start of the school year, students who were on the rosters for these classes were randomly assigned between the intervention and comparison classrooms. After excluding the students who never attended the classes, the non-randomized students who entered after the start of school, and the randomized students with missing posttest scores, the analytic sample included 699 students (351 intervention and 348 comparison). Among the analytic student sample, 76% were minorities, and 36% were eligible for free or reduced-price meals.

Intervention group The intervention classrooms were taught using the *Carnegie Learning Curricula and Cognitive Tutor®* Geometry curriculum for one academic school year. The curriculum included teacher-directed classroom instruction (60% of classroom time) and computer-guided individual instruction (40% of classroom time). The curriculum included a pacing guide designed by the district to ensure that the required geometry content would be covered in the intervention classes.

Comparison group Students in the comparison classrooms received the regular geometry curriculum ("business–as-usual").

Outcomes and measurement The outcome for this study was the Baltimore County Public School district geometry assessment. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation Teachers received three days of training prior to use of the curriculum and one day of follow-up instruction on the software and classroom-based activities of the curriculum.

Appendix A.4: Research details for Shneyderman, 2001

Shneyderman, A. (2001). *Evaluation of the Cognitive Tutor Algebra I program.* Unpublished manuscript. Miami, FL: Miami–Dade County Public Schools, Office of Evaluation and Research.

Table A4. Summary of findings — **Meets WWC evidence standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	6 schools/658 students	+2	No

Setting

Within Miami–Dade County Public Schools, nine senior high schools used the *Carnegie Learning Curricula and Cognitive Tutor*® Algebra I program during the 2000–01 school year. Of those, the six schools that had a computer lab for use of the software program were selected for the study.

Study sample

For each of the six schools, two teachers were randomly sampled from all teachers participating in the program (excluding those working with classes of predominantly exceptional education students). One class for each teacher was randomly sampled, creating an intervention sample of 12 classrooms with 325 students. The comparison sample was composed of 12 classrooms with 452 students, randomly sampled from a pool of classrooms not implementing the program in the same six schools.

Initial proportions of student recipients of free and reduced-price lunch were identical (54%) for the two groups, and ethnic (30% Black, 56% Hispanic, and 13% White for intervention; 27% Black, 62% Hispanic, and 10% White for comparison) and gender (46% and 48% female for intervention and comparison, respectively) distributions were similar. Most of the students in both groups were in ninth and tenth grades: 79% and 18% for the intervention group, and 88% and 11% for the comparison group. The analyses were conducted on 276 intervention and 382 comparison students in ninth and tenth grades.

Intervention group

The intervention group received the *Carnegie Learning Curricula and Cognitive Tutor*® Algebra I program covering a full year Algebra I course. Whereas one school had a functional computer lab at the beginning of the school year, the other four schools did not have an operational computer lab until October, thereby possibly affecting the implementation of the software component within these schools.

Comparison group

Comparison group students received Algebra I instruction using a different curriculum not specified in the study.

Outcomes and measurement

Algebra performance was measured using the FCAT Norm-Referenced Component and the ETS Algebra I End-of-Course Assessment. However, based on data received by the WWC in response to a query, the intervention and comparison groups used in the ETS Algebra I End-of-Course Assessment analysis sample were not equivalent at baseline on the ETS assessment (0.14), and the analysis did not adjust for the pretest differences. Therefore, only findings related to the FCAT are included in this review. For a more detailed description of the FCAT outcome measure, see Appendix B.

Support for implementation

No information was provided about the training or support offered to implement the intervention.

Appendix A.5: Research details for Smith, 2001

Smith, J. E. (2001). *The effect of the Carnegie Algebra Tutor on student achievement and attitude in introductory high school algebra* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University, Blacksburg.

Table A5. Summary of findings **Meets WWC evidence standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	7 schools/445 students	–3	No

Setting The study involved 10 high schools in Virginia Beach City Public Schools, a large, urban, K–12 school district in Virginia. Of the 10 high schools, one opted not to participate in the program, and two did not keep students in the intervention program together for all three semesters of the study. Therefore, seven schools were used for the analysis. The student population was 33.5% minority, including 25% Black.

Study sample The target population included all students who completed a three-semester algebra sequence during the 1999–2000 school year and the fall semester of the 2000–01 school year. This sequence is part of the district's core curriculum and covers the standard Algebra I material at a slower pace than the traditional math sequence—it uses the same textbook and follows the same curriculum as the year-long Algebra I course but is taught over three semesters. Students are recommended for this sequence by previous math teachers because they have struggled with lower-level math courses. Thus, the sample population consisted of 445 students (229 intervention and 216 comparison) who followed this course progression in one of the seven schools included in the study. Students were randomly assigned to available classes through a computer-scheduling program. However, random assignment was compromised because the analysis did not include all students that were randomly assigned. The analysis only included students who completed their assigned algebra curriculum.

Intervention group Each high school identified a math teacher who was willing to implement the intervention rather than the traditional curriculum. Each teacher had their students spend 40% of class time on the computer and 60% of class time receiving instruction outside the computer lab.

Comparison group Comparison classes used traditional instruction based on the city curriculum and textbook, without use of computers or the tutoring software.

Outcomes and measurement At the conclusion of the three-semester Algebra sequence, students took the Virginia SOL Assessment for Algebra I. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation Each teacher participated in a three-day training program on how to implement the intervention. Two-thirds of the intervention group teachers were replaced in the second year, and the new teachers did not receive training.

Appendix A.6: Research details for Wolfson et al., 2008

Wolfson, M., Koedinger, K., Ritter, S., & McGuire, C. (2008). *Cognitive Tutor Algebra I: Evaluation of results (1993–1994)*. Pittsburgh, PA: Carnegie Learning, Inc.

Table A6. Summary of findings | **Meets WWC evidence standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	3 schools/131 students	+28	Yes

Setting

The study took place in three high schools in the urban Pittsburgh Public Schools District. Within these schools, 50% of the student body was African American, 50% came from one-parent families, and 15% attended college.

Study sample

Students in the intervention and comparison Algebra I classes were matched based on their math grades from the prior year. The intervention group included 26 classes from the three schools, and the comparison group included five classes from two of the three schools. The Math SAT sample consisted of 102 intervention students and 29 comparison students.

Among the full study sample, 34% were African American, 56% were female, and 60% were eligible for free or reduced-price lunch. Almost two-thirds of the sample were in ninth grade, 24% were in tenth grade, 8% were in eleventh grade, and 2% were in twelfth grade.

Intervention group

The intervention group used an early version of *Carnegie Learning Curricula and Cognitive Tutor®* software, then referred to as the Pittsburgh Urban Mathematics Project curriculum plus Practical Algebra Tutor program.[5] The curriculum emphasized the use of functional models, such as tables, graphs, and symbols, to solve “real-world” problems. Students in the intervention group used the tutoring software in about 25 of the 180 class periods.

Comparison group

Students in the comparison group received the traditional Algebra I curriculum.

Outcomes and measurement

Four assessments were administered over two days at the end of the spring semester. The Iowa Test of Basic Skills was given to all students on the first day. According to information provided to the WWC by the study authors, on the second day, students were administered two assessments randomly selected from a set of three assessments: the Math SAT subset, the Problem Situation assessment, and the Representations assessment. However, based on data received from the authors by the WWC in response to a query, the intervention and comparison groups were not equivalent at baseline on the Iowa assessment and the Representations test. Therefore, only findings related to the sample of students who took the Math SAT and the sample of students who took the Problem Situation assessment are included in this review. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

No information was provided about the training or support offered to implement the intervention.

Appendix B: Outcome measures for each domain

Mathematics achievement	
NWEA Achievement Level Test/ Measures of Academic Progress	The Achievement Level Test is a paper-based end-of-course algebra exam. The Measures of Academic Progress test is a computerized version of the Achievement Level Test. The two tests are scored on a Rash Unit (RIT) scale, an equal-interval scale that yields a constant change in growth for a one-unit change, regardless of the numerical scale value. RIT scores range from about 150 to 300 and indicate a student's current achievement level along a curriculum scale for a particular subject. These results are combined by the study authors (as cited in Cabalo et al., 2007).
ETS Algebra I End-of-Course Assessment	This 50-question multiple-choice test is based on the Algebra I standards of the National Council of Teachers of Mathematics (as cited in Campuzano et al., 2009).
Baltimore County Public School District Geometry Assessment	This district-wide geometry assessment includes 30 multiple-choice items and 11 extended response items. Total scores can range between 0 and 50 (as cited in Pane et al., 2010).
FCAT Norm-Referenced Component	This 48-question multiple-choice test has questions ranging from problem solving to pre-calculus (as cited in Shneyderman, 2001).
Virginia SOL Algebra Assessment	This high-stakes assessment, which students need to pass to graduate from high school, consists of 50 questions that contribute to the student's score: 12 on expressions and operations, 12 on relations and functions, 18 on equations and inequalities, and eight on statistics (as cited in Smith, 2001).
Math SAT subset	The study used a subset of items drawn from the Math SAT and determined by the study authors to be appropriate for ninth graders (as cited in Koedinger et al., 1997).
Problem Situation test	This author-created test assesses students' abilities to investigate problem situations, presented verbally, that have algebraic content (as cited in Koedinger et al., 1997). The authors reported to the WWC that the internal consistency of the test is 0.91 and therefore meets WWC requirements.

Appendix C: Findings included in the rating for the mathematics achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation) Intervention group	Comparison group	WWC calculations Mean difference	Effect size	Improvement index	*p*-value
Cabalo et al., 2007[a]								
NWEA Algebra End-of-Course Achievement Level Test/ Measures of Academic Progress	Grades 8–13	6 schools/ 344 students	243.37 (7.67)	244.71 (7.47)	–1.34	–0.18	–7	0.23
Domain average for mathematics achievement (Cabalo et al., 2007)						**–0.18**	**–7**	**Not statistically significant**
Campuzano et al., 2009[b]								
ETS Algebra I End-of-Course Assessment	Grades 8–9	9 schools/ 276 students	29.78 (11.04)	31.88 (14.52)	–2.10	–0.16	–6	0.30
Domain average for mathematics achievement (Campuzano et al., 2009)						**–0.16**	**–6**	**Not statistically significant**
Pane et al., 2010[c]								
Baltimore County Public School District Geometry Assessment	High school grades	8 schools/ 699 students	–0.07 (0.99)	0.12 (1.00)	–0.19	–0.19	–8	0.03
Domain average for mathematics achievement (Pane et al., 2010)						**–0.19**	**–8**	**Statistically significant**
Shneyderman, 2001[d]								
FCAT Norm-Referenced Component	Grades 9–10	6 schools/ 658 students	683.66 (29.78)	682.47 (27.53)	1.19	0.04	+2	0.86
Domain average for mathematics achievement (Shneyderman, 2001)						**0.04**	**+2**	**Not statistically significant**
Smith, 2001[e]								
Virginia SOL Algebra Assessment	Grades 9+	7 schools/ 445 students	397.90 (32.90)	400.00 (29.10)	–2.10	–0.07	–3	0.46
Domain average for mathematics achievement (Smith, 2001)						**–0.07**	**–3**	**Not statistically significant**
Wolfson et al., 2008[f]								
Math SAT subset	Grades 9–12	3 schools/ 131 students	0.31 (0.16)	0.24 (0.10)	0.07	0.45	+17	0.01
Problem Situation Test subset	Grades 9–12	3 schools/ 85 students	8.83 (5.54)	3.29 (3.15)	5.54	1.07	+36	0.00
Domain average for mathematics achievement (Wolfson et al., 2008)						**0.76**	**+28**	**Statistically significant**
Domain average for mathematics achievement across all studies						**–0.02**	**–1**	**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 student outcomes, representing the average change expected for all students 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 student's percentile rank that can be expected if the student 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 each study's domain average was determined by the WWC. NWEA = Northwest Evaluation Association. FCAT = Florida Comprehensive Assessment Test. SOL = Standards of Learning.

[a] For Cabalo et al. (2007), no corrections for clustering or multiple comparisons were needed. The *p*-value presented here was reported in the original study. The outcome measure includes scores from the *NWEA Algebra End-of-Course Achievement Level Test* and the *Measures of Academic Progress*. The intervention group mean is the sum of the adjusted comparison group mean and the hierarchical linear modeling (HLM) coefficient for the difference between the two groups in the study. The standard deviations presented are adjusted standard deviations. This study is characterized as having indeterminate effects because no effects are statistically significant or substantively important. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, page 96.

[b] For Campuzano et al. (2009), no corrections for clustering or multiple comparisons and no difference-in-differences adjustment were needed. The *p*-value presented here was reported in the original study. The intervention group mean is the sum of the HLM coefficient and the reported effect size. The standard deviations presented were provided to the WWC by the authors. This study is characterized as having indeterminate effects because no effects are statistically significant or substantively important. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, page 96.

[c] For Pane et al. (2010), no corrections for clustering or multiple comparisons and no difference-in-differences adjustment were needed. The *p*-value presented here was reported in the original study. The intervention group mean is the sum of the adjusted comparison group mean and the HLM coefficient for the difference between the two groups in the study. The standardized group means and standard deviations were provided to the WWC by the authors. This study is characterized as having statistically significant negative effects because the effect for at the only measure within the domain is negative and statistically significant.

[d] For Shneyderman (2001), the means and standard deviations for both the intervention and comparison groups were computed using data on ninth- and tenth-grade samples obtained from the study author. The *p*-value presented here was calculated by the WWC and corrected for clustering. The study does not report findings pooled across ninth- and tenth-grade students; however, neither of the grade-level findings were reported as statistically significant. This study is characterized as having indeterminate effects because no effects are statistically significant or substantively important. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, page 96.

[e] For Smith (2001), no corrections for clustering or multiple comparisons and no difference-in-differences adjustment were needed. The *p*-value presented here was reported in the original study. This study is characterized as having indeterminate effects because no effects are statistically significant or substantively important. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, page 96.

[f] For Wolfson et al. (2008), a correction for clustering and multiple comparisons was needed and results in a significance level that differs from that in the original study for the *Math SAT* outcome, but not for the *Problem Situation* outcome. After adjusting for clustering, the WWC-calculated *p*-value for the *Math SAT* outcome was 0.11. Therefore, the WWC does not find this result statistically significant. The *p*-values presented here were reported in the original study. The standard deviations were provided by the study authors. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, page 96.

Appendix D: Summary of supplemental findings for the mathematics achievement domain

			Mean (standard deviation)		WWC calculations			
Outcome measure	Study sample	Sample size	Intervention group	Comparison group	Mean difference	Effect size	Improvement index	*p*-value
Cabalo et al., 2007[a]								
NWEA Algebra End-of-Course Achievement Level Test/Measures of Academic Progress—Quadratic Questions	Grades 8–13	6 schools/ 333 students	238.96 (11.24)	242.40 (9.98)	–3.44	–0.32	–13	0.02
NWEA Algebra End-of-Course Achievement Level Test/Measures of Academic Progress—Algebraic Operations	Grades 8–13	6 schools/ 345 students	241.03 (9.99)	243.50 (10.18)	–2.47	–0.24	–10	0.16
NWEA Algebra End-of-Course Achievement Level Test/Measures of Academic Progress—Linear Equations	Grades 8–13	6 schools/ 335 students	244.81 (9.57)	245.24 (7.94)	–0.43	–0.04	–2	0.80
NWEA Algebra End-of-Course Achievement Level Test/Measures of Academic Progress—Problem Solving	Grades 8–13	6 schools/ 338 students	246.67 (11.90)	246.38 (10.69)	0.29	0.03	1	0.86
Shneyderman, 2001[b]								
FCAT Norm-Referenced Component	Grade 10	6 schools/ 92 students	688.00 (24.69)	693.60 (24.23)	–5.60	–0.23	–9	> 0.05

Table Notes: The supplemental findings presented in this table are additional findings from the studies in this report that 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 student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. NWEA = Northwest Evaluation Association. FCAT = Florida Comprehensive Assessment Test.

[a] For Cabalo et al. (2007), a correction for multiple comparisons was needed and result in significance levels that differ from those in the original study. Due to the multiple comparison adjustment, the *p*-value of 0.02 for the *Quadratic Questions* measure was higher than the critical *p*-value of 0.01 for statistical significance; therefore, the WWC does not find the result to be statistically significant. The *p*-values presented here were reported in the original study. The intervention group mean is the sum of the adjusted comparison group mean and the HLM coefficient for the difference between the two groups in the study. The standard deviations presented are adjusted standard deviations.

[b] For Shneyderman (2001), a correction for clustering was needed but did not affect significance levels. The standard deviations were obtained by the WWC from the study author. The *p*-value presented here was reported in the original study. Subgroup findings related to grade 9 students are not presented here because the intervention and comparison groups in that analysis sample are not equivalent in baseline mathematics achievement (although the intervention and comparison groups in the analysis sample that pools students in grades 9 and 10, as presented in Appendix C, are equivalent in baseline mathematics achievement).

Endnotes

[1] The descriptive information for this program was obtained from a publicly available source: the developer's website (http://www.carnegielearning.com/, downloaded April 2010). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in January 2012, and the WWC incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2011.

[2] The previous report was released in August 2010. This report has been updated to include reviews of four studies that have been released since 2010. Of the additional studies, one was not within the scope of the protocol, two were within the scope of the topic area review but did not meet evidence standards, and one met WWC standards without reservations. The report includes reviews of all previous studies that met WWC evidence standards with or without reservations and did not result in revised dispositions of these studies. The report also includes a review of Wolfson et al. (2008) based on additional information provided to the WWC by the study authors pertaining to the study design and baseline equivalence of the research groups. The disposition of this study was revised from "does not meet WWC standards" to "meets WWC standards with reservations." A complete list and disposition of all studies reviewed are provided in the references. The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1), along with those described in the High School Mathematics review protocol (version 2.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] One of the studies summarized in this intervention report, Campuzano et al. (2009), was prepared by staff of one of the WWC contractors. Because the principal investigator for the WWC review of high school mathematics is also a staff member of that contractor and an author of this study, the study was rated by staff members from a different organization, who also prepared the intervention report. The report was then reviewed by the principal investigator, a WWC Quality Assurance reviewer, and an external peer reviewer.

[4] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 23. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

[5] Grade, delivery method, and program type refer to the studies that meet WWC evidence standards without or with reservations.

[6] The WWC has determined that the Pittsburgh Urban Mathematics Project curriculum plus Practical Algebra Tutor program is sufficiently similar to the *Carnegie Learning Curricula and Cognitive Tutor®* to be included in this review.

[7] Consistent with the High School Mathematics review protocol, students in grades outside of high school were included in the review if they were included in the study analysis sample along with students in grades 9–12.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, January). *High School Mathematics intervention report:* Carnegie Learning Curricula and Cognitive Tutor®. Retrieved from http://whatworks.ed.gov.

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence 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 evidence 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 evidence 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. 23.

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

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 subjects 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 investigators randomly assign eligible participants into 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. 23.

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.

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% ($p < 0.05$).

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

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.