WWC Intervention Report

A summary of findings from a systematic review of the evidence

INSTITUTE OF EDUCATION SCIENCES

Secondary Mathematics June 2016 Revised*

Cognitive Tutor®

Report Contents

Overview	p. 1
Program Information	p. 3
Research Summary	p. 4
Effectiveness Summary	p. 6
References	p. 9
Research Details for Each Study	p. 19
Outcome Measures for Each Domain	p. 29
Findings Included in the Rating for Each Outcome Domain	p. 30
Supplemental Findings for Each Outcome Domain	p. 33
Endnotes	p. 34
Rating Criteria	p. 36
Glossary of Terms	p. 37

This intervention report presents findings from a systematic review of *Cognitive Tutor®*, conducted using the WWC Procedures and Standards Handbook, version 3.0, and the Secondary Mathematics review protocol, version 3.1.

Program Description[1]

Cognitive Tutor® is a secondary mathematics curriculum developed by Carnegie Learning that focuses on how students think about and learn mathematics. Teachers facilitate student learning as students acquire and apply new information and discuss their work. The curriculum can be implemented using a textbook, adaptive software, or combination of textbook and software activities. This WWC intervention report focuses on studies of all *Cognitive Tutor®* secondary courses, which include: *Algebra I*, *Algebra II*, and *Geometry*, as well as *Integrated Math I*, *II*, and *III*, a three-course series that integrates numeric, algebraic, geometric, and statistical content.

Research[2]

The What Works Clearinghouse (WWC) identified six studies of *Cognitive Tutor® Algebra I* and one study of *Cognitive Tutor® Geometry* that both fall within the scope of the Secondary Mathematics topic area and meet WWC group design standards.[3] Two studies of *Cognitive Tutor® Algebra I* meet WWC group design standards without reservations, and four studies of *Cognitive Tutor® Algebra I* meet WWC group design standards with reservations. Together, these six studies included 12,840 students in grades 8–13 in 118 locations.[4,5] The one study of *Cognitive Tutor® Geometry* also meets WWC group design standards with reservations. This study included 669 students in grades 9–12 in eight locations.[6]

The WWC considers the extent of evidence for *Cognitive Tutor® Algebra I* on the mathematics achievement of secondary students to be medium to large for one outcome domain—algebra—and small for one outcome domain—general mathematics achievement. The WWC considers the extent of evidence for *Cognitive Tutor® Geometry* on the mathematics achievement of secondary students to be small for one outcome domain—geometry. There were no studies that meet WWC group design standards in the three other domains, so this intervention report does not report on the effectiveness of *Cognitive Tutor®* for those domains. (See the Effectiveness Summary on p. 6 for more details of effectiveness by domain.)

The findings in this report pertain only to *Cognitive Tutor® Algebra I* and *Cognitive Tutor® Geometry*. No studies that examine *Cognitive Tutor® Algebra II* or *Cognitive Tutor® Integrated Math I*, *II*, and *III* fall within the scope of the Secondary Mathematics review protocol and meet WWC group design standards.

Effectiveness[7]

Cognitive Tutor® Algebra I was found to have mixed effects on algebra and no discernible effects on general mathematics achievement for secondary students.

Cognitive Tutor® Geometry was found to have potentially negative effects on geometry for secondary students.

Table 1. Summary of findings[8]

Course and outcome domain	Rating of effectiveness	Improvement index (percentile points): Average	Improvement index (percentile points): Range	Number of studies	Number of students	Extent of evidence
Cognitive Tutor® Algebra I						
Algebra	Mixed effects	+4	–7 to +19	5	12,182	Medium to large
General mathematics achievement	No discernible effects	+2	na	1	658	Small
Cognitive Tutor® Geometry						
Geometry	Potentially negative effects	–8	na	1	669	Small
Cognitive Tutor® Algebra II	No evidence					
Cognitive Tutor® Integrated Math I	No evidence					
Cognitive Tutor® Integrated Math II	No evidence					
Cognitive Tutor® Integrated Math III	No evidence					

na = not applicable

Program Information

Background

Cognitive Tutor® was developed and is published by Carnegie Learning, Inc. Address: 437 Grant Street, Frick Building, Suite 1906, Pittsburgh, PA 15219. Email: info@carnegielearning.com. Web: www.carnegielearning.com. Telephone: (888) 851-7094. Fax: (412) 690-2444.

Program details

Cognitive Tutor® can be implemented using a textbook, adaptive software, or combination of textbook and software activities. In a blended implementation, three periods per week are spent using the textbook 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 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.

The curriculum uses a pedagogical approach focused on how students acquire and apply new information, and promotes a student-centered classroom where teachers facilitate student learning. Student discussions about mathematics are encouraged as students explain their solutions and reasoning to one another. Multiple learning styles are accommodated through a variety of learning opportunities including real world problems, peer review of student work, step-by-step demonstrations on how to solve example problems, hands-on tools including manipulatives and technology, graphic representations of mathematical concepts, and classroom discussions and explanations about mathematical understandings and key concepts.

Teacher curriculum materials contain print and electronic materials, including teacher implementation guides, as well as teacher resources, including assessments, professional development, and classroom management tools. Student curriculum materials include texts, personalized instruction using the software system, and supplemental materials such as student assignments and skills practice exercises.

Cost

As of February 2016, the cost of *Cognitive Tutor*® varies depending on the number of students who use the program and the type of implementation (textbook only, software only, or blended). Cost information for *Cognitive Tutor*® is available from the publisher.

Research Summary

This research summary includes information from studies of all available *Cognitive Tutor*® courses for secondary students.

The WWC identified 22 studies that were eligible for review:

- Eighteen eligible group design studies and one eligible regression discontinuity design study investigated the effects of *Cognitive Tutor® Algebra I* on the mathematics achievement of secondary students.
- Two eligible group design studies investigated the effects of *Cognitive Tutor® Geometry* on the mathematics achievement of secondary students.
- One eligible group design study investigated the effects of *Cognitive Tutor® Algebra II* on the mathematics achievement of secondary students.[9]

Table 2. Scope of reviewed research

Grades	8–13
Delivery method	Whole class
Program type	Curriculum

The WWC reviewed 21 of the eligible studies against group design standards. Two studies are randomized controlled trials that meet WWC group design standards without reservations, and five studies are randomized controlled trials or quasi-experimental design studies that meet WWC group design standards with reservations. Those seven studies focus on *Cognitive Tutor® Algebra I* and *Cognitive Tutor® Geometry* and are summarized in this report. Fourteen of the 21 studies do not meet WWC group design standards.

The WWC reviewed one eligible study against regression discontinuity design standards. The evidence from the regression discontinuity design study on *Cognitive Tutor® Algebra I* does not meet WWC pilot regression discontinuity design standards and therefore is not included in the effectiveness ratings in this report.

An additional 93 studies were identified but do not meet WWC eligibility criteria for review in this topic area. Citations for all 115 studies are in the References section, which begins on p. 9.

Summary of *Cognitive Tutor® Algebra I* studies meeting WWC group design standards without reservations

Cabalo et al. (2007) randomly assigned 22 classrooms within teachers' assigned sections to receive either *Cognitive Tutor® Algebra I* or the business-as-usual curriculum. The analysis sample included nine intervention teachers with 182 students and nine comparison teachers with 162 students across five Maui School District schools and Maui Community College. The pretest was administered in fall 2005 and the posttest in May 2006.

Ritter et al. (2007) conducted a randomized controlled trial in which ninth-grade algebra course sections for each of the six study teachers were randomly assigned to use either *Cognitive Tutor® Algebra I* or McDougal Littell's *Heath Algebra I*, a traditional teacher-directed curriculum. The study took place in three suburban junior high schools in Oklahoma during the 2000–01 school year. The analysis sample included 153 students within the ten *Cognitive Tutor® Algebra I* classrooms and 102 students within the six comparison classrooms. Each of the six study teachers taught sections using *Cognitive Tutor® Algebra I* and *Heath Algebra I*.

Summary of *Cognitive Tutor® Algebra I* studies meeting WWC group design standards with reservations

Campuzano et al. (2009) conducted a randomized controlled trial in which teachers in high-poverty schools were randomly assigned within each school to intervention and comparison groups as part of a national study of education technology. During the second year of the study, *Cognitive Tutor® Algebra I* was implemented in nine schools in four districts. Teachers were randomly assigned to either use the intervention or continue using their standard curriculum within each study school. Fall and spring tests were administered to 142 intervention students and 128 comparison students in grades 8 and 9. The integrity of the study's random assignment was jeopardized because the student sample was defined after teacher random assignment. Students who enter classrooms after random assignment jeopardize the random assignment because the authors discuss the effects of the intervention on students and not only effects on classrooms. The study demonstrated equivalence on the analytic sample and therefore, *meets WWC group design standards with reservations*.

Pane et al. (2014) conducted a randomized controlled trial in 73 high schools within 51 school districts in seven US states. The authors randomly assigned pairs of schools to either *Cognitive Tutor® Algebra I* or a comparison group that continued using the school's existing Algebra I curriculum. *Cognitive Tutor® Algebra I* was delivered to two cohorts of students in two consecutive school years and included a teacher-directed classroom instruction component and a computer-guided instruction component. The analytic sample of students contained 11,066 high school students. The integrity of the random assignment was jeopardized because students enrolled in the study several months after school random assignment was conducted. Students who enter classrooms after random assignment jeopardize the random assignment because the authors discuss the effects of the intervention on students and not only effects on classrooms. The study demonstrated equivalence on the analytic sample of high school students and therefore, *meets WWC group design standards with reservations*.

Shneyderman (2001) conducted a quasi-experimental study in six senior high schools in Miami–Dade County Public Schools. The schools were selected because they had an operational computer lab during the 2000–01 school year and contained some teachers already implementing *Cognitive Tutor® Algebra I*. Within each study school, two teachers were randomly selected from all algebra teachers that were implementing *Cognitive Tutor® Algebra I*, and one class was selected from each teacher to form the intervention sample. The comparison group was formed by selecting two Algebra I classrooms not using *Cognitive Tutor® Algebra I* from each of the six study schools. The analysis sample was comprised of 276 intervention students within 12 classrooms and 382 comparison students within 12 classrooms. All study students were in grades 9 and 10.

Wolfson et al. (2008) conducted a quasi-experimental design study during the 1993–94 school year with 26 Algebra I classrooms across three high schools in the Pittsburgh Public School District. Students in the intervention classrooms used an early version of *Cognitive Tutor®*, which at the time was referred to as the *Pittsburgh Urban Mathematics Project curriculum plus Practical Algebra Tutor program*. 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.

Summary of *Cognitive Tutor® Geometry* studies meeting WWC group design standards without reservations

No studies of *Cognitive Tutor® Geometry* met WWC group design standards without reservations.

Summary of *Cognitive Tutor® Geometry* study meeting WWC group design standards with reservations

Pane et al. (2010) randomly assigned students in eight high schools in the Baltimore County Public School District to receive either *Cognitive Tutor® Geometry* or the standard geometry curriculum. The study took place over 3 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 both the intervention and comparison curricula across the two study class periods. The analytic sample included 60 classrooms (30 intervention and 30 comparison) and 669 students (329 intervention and 340 comparison). The WWC determined that the study experienced high attrition, but the authors demonstrated equivalence on the analytic sample; therefore, the study *meets WWC group design standards with reservations*.

Effectiveness Summary

The WWC review of *Cognitive Tutor® Algebra I* and *Cognitive Tutor® Geometry* for the Secondary Mathematics topic area includes student outcomes in six domains: algebra, geometry, statistics and probability, trigonometry/precalculus, calculus, and general mathematics achievement. The six studies of *Cognitive Tutor® Algebra I* that meet WWC group design standards reported findings in two of the six domains: (a) algebra and (b) general mathematics achievement. The one study of *Cognitive Tutor® Geometry* that meets WWC group design standards reported findings in one of the six domains: geometry. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Cognitive Tutor® Algebra I* and *Cognitive Tutor® Geometry* on secondary students. Additional comparisons are presented as supplemental findings for *Cognitive Tutor® Algebra I* in Appendix D. These supplemental findings do not factor into *Cognitive Tutor® Algebra I*'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. 36.

Summary of *Cognitive Tutor® Algebra I* effectiveness for the algebra domain

Table 3. Rating of effectiveness of *Cognitive Tutor® Algebra I* and extent of evidence for the algebra domain

Rating of effectiveness	Criteria met
Mixed effects *Evidence of inconsistent effects.*	In the five studies that reported findings, the estimated impact of the intervention on outcomes in the *algebra* domain was positive and statistically significant in one study, positive and substantively important in a second study, and indeterminate in three other studies.
Extent of evidence	**Criteria met**
Medium to large	Five studies that included 12,182 students[10] in 94 schools reported evidence of effectiveness in the *algebra* domain.

Five studies of *Cognitive Tutor® Algebra I* that meet WWC group design standards with or without reservations reported findings in the algebra domain.

Cabalo et al. (2007) reported, and the WWC confirmed, a negative but not statistically significant difference between the *Cognitive Tutor® Algebra I* and comparison groups on outcomes in the algebra domain. 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) reported, and the WWC confirmed, no statistically significant difference between the *Cognitive Tutor® Algebra I* and comparison groups on outcomes in the algebra domain. 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. (2014) reported, and the WWC confirmed, no statistically significant difference between the *Cognitive Tutor® Algebra I* and comparison groups on outcomes in the algebra domain for the first study cohort. Study authors also reported a statistically significant difference between the *Cognitive Tutor® Algebra I* and comparison groups on outcomes in the algebra domain for the second study cohort. However, when this result was adjusted for multiple comparisons, the WWC found that this difference was no longer statistically significant. The effect sizes for both differences were 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.

Ritter et al. (2007) reported, and the WWC confirmed, a positive but not statistically significant effect of *Cognitive Tutor® Algebra I* on outcomes in the algebra domain. The effect size was 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 substantively important positive effect.

Wolfson et al. (2008) reported, and the WWC confirmed, a positive and statistically significant difference between the *Cognitive Tutor® Algebra I* and comparison groups on the Multiple Representations Test. The authors also reported a positive and statistically significant difference between the *Cognitive Tutor® Algebra I* and comparison groups on the Iowa Algebra Aptitude Test. However, when this result was adjusted for multiple comparisons, the WWC found that this difference was no longer statistically significant. The effect sizes of both differences were not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). Taken together, the WWC characterizes this study finding as a statistically significant positive effect.

Thus, for the algebra domain, one study of *Cognitive Tutor® Algebra I* showed a statistically significant positive effect, one study showed a substantively important positive effect, and three studies showed an indeterminate effect. This results in a rating of mixed effects, with a medium to large extent of evidence.

Summary of *Cognitive Tutor® Algebra I* effectiveness for the general mathematics achievement domain

Table 4. Rating of effectiveness of *Cognitive Tutor® Algebra I* and extent of evidence for the general mathematics achievement 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 *general mathematics achievement* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 658 students in six schools reported evidence of effectiveness in the *general mathematics achievement* domain.

One study of *Cognitive Tutor® Algebra I* that meets WWC group design standards with reservations reported findings in the general mathematics achievement domain.

Shneyderman (2001) reported, and the WWC confirmed, no statistically significant difference between the *Cognitive Tutor® Algebra I* and comparison groups on the general mathematics achievement domain. 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.

Thus, for the general mathematics achievement domain, one study of *Cognitive Tutor® Algebra I* did not show a statistically significant or substantively important effect. This results in a rating of no discernible effects, with a small extent of evidence.

Summary of *Cognitive Tutor® Geometry* effectiveness for the geometry domain

Table 5. Rating of effectiveness of *Cognitive Tutor® Geometry* and extent of evidence for the geometry domain

Rating of effectiveness	Criteria met
Potentially negative effects *Evidence of a negative effect with no overriding contrary evidence.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *geometry* domain was negative and statistically significant.
Extent of evidence	**Criteria met**
Small	One study that included 669 students in eight schools reported evidence of effectiveness in the *geometry* domain.

One study of *Cognitive Tutor® Geometry* that meets WWC group design standards with reservations reported findings in the geometry domain.

Pane et al. (2010) reported, and the WWC confirmed, a statistically significant negative difference between the *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.

Thus, for the geometry domain, one study of *Cognitive Tutor® Geometry* showed a statistically significant negative effect. This results in a rating of potentially negative effects, with a small extent of evidence.

References

Studies of *Cognitive Tutor® Algebra I* that meet WWC group design 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.

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.

Additional sources:

Carnegie Learning, Inc. (2004). *Carnegie Learning research report: Moore Independent School District*. Pittsburgh, PA: Author.

Morgan, P., & Ritter, S. (2002). *An experimental study of the effects of Cognitive Tutor® Algebra I on student knowledge and attitude*. Pittsburgh, PA: Carnegie Learning, Inc. Retrieved from http://www.carnegielearning.com

Studies of *Cognitive Tutor® Algebra I* that meet WWC group design standards with reservations

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts* (NCEE 2009-4041). Washington, DC: U.S. Department of Education, Institute of Education Sciences. http://files.eric.ed.gov/fulltext/ED504657.pdf.

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: Report to Congress*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. http://files.eric.ed.gov/fulltext/ED496015.pdf.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale (high school experiment). *Educational Evaluation and Policy Analysis, 36*(2), 127–144. doi:10.3102/0162373713507480

Additional sources:

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). *Addendum to effectiveness of Cognitive Tutor Algebra I at scale* (Working Paper WR-1050-DEIES) (high school experiment). Santa Monica, CA: RAND Corporation.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., Karam, R., Daugherty, L., & Phillips, A. (2013). *Does an algebra course with tutoring software improve student learning?* (high school experiment). Santa Monica, CA: RAND Corporation.

Sales, A. C., & Pane, J. F. (2015, June). *Exploring causal mechanisms in a randomized effectiveness trial of the Cognitive Tutor* (high school experiment). Paper presented at the 8th International Conference on Educational Data Mining, Madrid, Spain.

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.

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

Additional sources:

Carnegie Learning, Inc. (2001). *Report of results from Pittsburgh, Pennsylvania* (Cognitive Tutor research report PA-91-01). Pittsburgh, PA: Author.

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 of *Cognitive Tutor® Algebra I* that do not meet WWC group design standards

Arbuckle, W. J. (2005). *Conceptual understanding in a computer-assisted Algebra 1 classroom* (Unpublished doctoral dissertation). University of Oklahoma, Norman. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Carnegie Learning, Inc. (2001). *Results from El Paso, Texas* (Cognitive Tutor research report TX-00-03). Pittsburgh, PA: Author. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Carnegie Learning, Inc. (2001). *Results from Lewisville, TX* (Cognitive Tutor research report TX-00-01). Pittsburgh, PA: Author. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Carnegie Learning, Inc. (2001). *Results from Milwaukee, Wisconsin* (Cognitive Tutor research report WI-97-01). Pittsburgh, PA: Author. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Carnegie Learning, Inc. (2001). *Results from The Colony, TX* (Cognitive Tutor research report TX-00-02). Pittsburgh, PA: Author. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Carnegie Learning, Inc. (2002). *Results from El Paso, Texas* (Cognitive Tutor research report TX-01-01). Pittsburgh, PA: Author. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Hobbs, R. M. (2012). *Improving problem-solving techniques for students in low-performing schools* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3493910) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale (middle school experiment). *Educational Evaluation and Policy Analysis, 36*(2), 127–144. doi:10.3102/0162373713507480 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:

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). *Addendum to effectiveness of Cognitive Tutor Algebra I at scale* (Working Paper WR-1050-DEIES) (middle school experiment). Santa Monica, CA: RAND Corporation.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., Karam, R., Daugherty, L., & Phillips, A. (2013). *Does an algebra course with tutoring software improve student learning?* (middle school experiment). Santa Monica, CA: RAND Corporation.

Sales, A. C., & Pane, J. F. (2015, June). *Exploring causal mechanisms in a randomized effectiveness trial of the Cognitive Tutor* (middle school experiment). Paper presented at the 8th International Conference on Educational Data Mining, Madrid, Spain.

Sabo, K. E., Atkinson, R. K., Barrus, A. L., Joseph, S. S., & Perez, R. S. (2013). Searching for the two sigma advantage: Evaluating algebra intelligent tutors. *Computers in Human Behavior, 29*(4), 1833–1840. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Sarkis, H. (2004). *Cognitive Tutor® Algebra 1: Miami-Dade County Public Schools*. Lighthouse Point, FL: The Reliability Group. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

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. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

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 group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Study of *Cognitive Tutor® Algebra I* that does not meet WWC pilot regression discontinuity design standards

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 does not meet WWC pilot regression discontinuity design standards because neither institutional nor statistical integrity of the forcing variable is established.

Additional sources:

Plano, G. (2003). *Carnegie Learning research report: Kent School District*. Pittsburgh, PA: Carnegie Learning, Inc.

Plano, G. S. (2004). *The effects of the Cognitive Tutor® Algebra on student attitudes and achievement in a 9th grade algebra course* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3130130)

Studies of *Cognitive Tutor® Geometry* that meet WWC group design standards without reservations

None.

Study of *Cognitive Tutor® Geometry* that meets WWC group design standards with reservations

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.

Study of *Cognitive Tutor® Geometry* that does not meet WWC group design standards

Bibi, T. (2010). *Analysis of Cognitive Tutor Geometry curriculum*. Ames: Iowa State University. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Study of *Cognitive Tutor® Algebra II* that does not meet WWC group design standards

Wills, C. (2014). *Man versus machine: The impact of instructional delivery method on the ACT math scores of African American students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3630401) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Studies of *Cognitive Tutor® Integrated Math I* that meet WWC group design standards

None.

Studies of *Cognitive Tutor® Integrated Math II* that meet WWC group design standards

None.

Studies of *Cognitive Tutor® Integrated Math III* that meet WWC group design standards

None.

Studies of *Cognitive Tutor®* that are ineligible for review using the Secondary 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 is out of scope of the protocol.

Aleven, V., Roll, I., McLaren, B., Ryu, E. J., & Koedinger, K. (2005). An architecture to combine meta-cognitive and cognitive tutoring: Pilot testing the Help Tutor. In C. G. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Proceedings of the 12th International Conference on Artificial Intelligence in Education* (pp. 17–24). Amsterdam: IOS Press. The study is ineligible for review because it is out of scope of the protocol.

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist, 45*(4), 224–233. doi:10.1080/00461520.2010.517740 The study is ineligible for review because it is out of scope of the protocol.

Anderson, J. R. (1992). Intelligent tutoring and high school mathematics. *Lecture Notes in Computer Science, 608*, 1–10. The study is ineligible for review because it does not use an eligible design.

Arevalillo-Herraez, M., Arnau, D., & Marco-Gimenez, L. (2013). Domain-specific knowledge representation and inference engine for an intelligent tutoring system. *Knowledge-Based Systems, 49*, 97–105. The study is ineligible for review because it is out of scope of the protocol.

Arroyo, I., Burleson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013). Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology, 105*(4), 957–969. The study is ineligible for review because it is out of scope of the protocol.

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 use an eligible design.

Baker, R. S. J. d. (2009). Differences between intelligent tutor lessons, and the choice to go off-task. *Proceedings of the 2nd International Conference on Educational Data Mining*, 11–20. http://files.eric.ed.gov/fulltext/ED539066.pdf. The study is ineligible for review because it is out of scope of the protocol.

Bakia, M., Shear, L., Toyama, Y., & Lasseter, A. (2012). *Understanding the implications of online learning for educational productivity*. Washington, DC: U.S. Department of Education, Office of Educational Technology. http://files.eric.ed.gov/fulltext/ED532492.pdf. This study is ineligible for review because it does not use a sample aligned with the protocol.

Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 431–440. The study is ineligible for review because it is out of scope of the protocol.

Bernacki, M., & Walkington, C. (2014). *The impact of a personalization intervention for mathematics on learning and non-cognitive factors*. Submitted to the 2014 International Conference of Educational Data Mining, London. The study is ineligible for review because it is out of scope of the protocol.

Bibi, T. (2011). Analysis of Cognitive Tutor Geometry curriculum. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 71*(10-A), 3582. This study is ineligible for review because it does not use a sample aligned with the protocol.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25*, 24–34. doi:10.1016/j.learninstruc.2012.11.002 http://files.eric.ed.gov/fulltext/ED543090.pdf. This study is ineligible for review because it does not use a sample aligned with the protocol.

Brown, E. A. (2013). Algebra program shows promise in large-scale study. *Education Daily, 46*(118), 1–2. The study is ineligible for review because it is out of scope of the protocol.

Carnegie Learning, Inc. (2001). *Report of results from Canton, Ohio* (Cognitive Tutor research report OH-01-01). Pittsburgh, PA: Author. The study is ineligible for review because it is out of scope of the protocol.

Carnegie Learning, Inc. (2001). *Results from Denver, CO* (Cognitive Tutor research report CO-00-01). Pittsburgh, PA: Author. The study is ineligible for review because it does not use an eligible design.

Carnegie Learning, Inc. (2001). *Results from San Francisco Unified School District summer school 2000* (Cognitive Tutor research report CA-00-01). Pittsburgh, PA: Author. The study is ineligible for review because it does not use an eligible design.

Carnegie Learning, Inc. (2010). *Cognitive Tutor effectiveness: Executive summary of research*. Pittsburgh, PA: Author. Retrieved from http://www.carnegielearning.com The study is ineligible for review because it does not use an eligible design.

Casas, I., Imbrogno, J., Ochoa, S. F., & Vergara, A. (2014, June). *Adapting a cognitive tutoring strategy for mathematics in Latin America*. Paper presented at the Fifth International Workshop on Culturally-Aware Tutoring Systems (CATS2014), Manoa, HI. The study is ineligible for review because it is out of scope of the protocol.

Cen, H., Koedinger, K. R., & Junker, B. (2007). Is over practice necessary?—Improving learning efficiency with the Cognitive Tutor through educational data mining. *Frontiers in Artificial Intelligence and Applications, 158*, 511–520. The study is ineligible for review because it is out of scope of the protocol.

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. The study is ineligible for review because it does not use an eligible design.

Clewell, B. C., de Cohen, C. C., Campbell, P. B., & Perlman, L. (2005). *Review of evaluation studies of mathematics and science curricula and professional development models*. Washington, DC: The Urban Institute. The study is ineligible for review because it does not use an eligible design.

Corbett, A. (2002). Cognitive Tutor Algebra I: Adaptive student modeling in widespread classroom use. In National Research Council (Ed.), *Technology and assessment: Thinking ahead—Proceedings from a workshop* (pp. 50–62). Washington, DC: The National Academies Press. The study is ineligible for review because it does not use an eligible design.

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

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

Corbett, A. T., Koedinger, K. R., & Hadley, W. H. (2001). Cognitive Tutors: From the research classroom to all classrooms. In P. S. Goodman (Ed.), *Technology enhanced learning: Opportunities for change* (pp. 253–263). Mahwah, NJ: Erlbaum. The study is ineligible for review because it does not use an eligible design.

Cortes, K., Goodman, J., & Nomi, T. (2013). A double dose of algebra. *Education Next, 13*(1), 70–76. This study is ineligible for review because it does not use a sample aligned with the protocol.

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 an eligible design.

Fancsali, S. E. (2014, July). *Causal discovery with models: Behavior, affect, and learning in Cognitive Tutor Algebra*. Paper presented at the 7th International Conference of Educational Data Mining, London. The study is ineligible for review because it does not use an eligible design.

Gleason, C. (2012). *Exploring the use of online tutoring software with students in a remedial math class*. New Wilmington, PA: Westminster College. This study is ineligible for review because it does not use a sample aligned with the protocol.

Graesser, A. C., Conley, M. W., & Olney, A. M. (2012). Intelligent tutoring systems. In S. Graham & K. Harris (Eds.), *APA educational psychology handbook: volume 3. Applications to learning and teaching* (pp. 451–473). Washington, DC: American Psychological Association. The study is ineligible for review because it is out of scope of the protocol.

Hausmann, R. G. M., Ritter, S., Towle, B., Murray, R. C., & Connelly, J. (2010). Incorporating interactive examples into the Cognitive Tutor. *Lecture Notes in Computer Science, 6095*, 446. The study is ineligible for review because it is out of scope of the protocol.

Hausmann, R. G. M., Vuong, A., Towle, B., Fraundorf, S. H., Murray, R. C., & Connelly, J. (2013). An evaluation of the effectiveness of just-in-time hints. *Lecture Notes in Computer Science, 7926*, 791–794. The study is ineligible for review because it is out of scope of the protocol.

Joshi, A., Fancsali, S. E., Ritter, S., Nixon, T., & Berman, S. R. (2014, July). *Generalizing and extending a predictive model for standardized test scores based on Cognitive Tutor interactions*. Paper presented at the 7th International Conference of Educational Data Mining, London. The study is ineligible for review because it does not use an eligible design.

Karabenick, S. A. (2011). Methodological and assessment issues in research on help seeking. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 267–281). New York: Routledge. The study is ineligible for review because it is out of scope of the protocol.

Kessler, A. M., Boston, M. D., & Stein, M. K. (2014). Conceptualizing teacher's practices in supporting students' mathematical learning in computer-directed learning environments. *Proceedings of International Conference of the Learning Sciences, ICLS, 2*(January), 997–1001. The study is ineligible for review because it is out of scope of the protocol.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1995). Intelligent tutoring goes to school in the big city. *Proceedings of the 7th World Conference on Artificial Intelligence in Education*. The study is ineligible for review because it is out of scope of the protocol.

Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. J. (2002). *Carnegie Learning's Cognitive Tutor: Summary research results*. Pittsburgh, PA: Carnegie Learning, Inc. The study is ineligible for review because it is out of scope of the protocol.

Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013). Using data-driven discovery of better student models to improve student learning. *Artificial Intelligence in Education*, 421–430. The study is ineligible for review because it does not use an eligible design.

Lepenioti, D., Vosniadou, S., & Alexandris, C. (2014). Designing an interactive tutoring tool for improving mathematical skills. *HCI International 2014-Posters' Extended Abstracts* (pp. 106-111). New York: Springer International Publishing. This study is ineligible for review because it does not use a sample aligned with the protocol.

Long, Y., & Aleven, V. (2012). Skill diaries: Can periodic self-assessment improve students' learning with an intelligent tutoring system? *Lecture Notes in Computer Science, 7315*, 673–674. This study is ineligible for review because it does not use a sample aligned with the protocol.

Long, Y., & Aleven, V. (2013). Skill diaries: Improve student learning in an intelligent tutoring system with periodic self-assessment. *Lecture Notes in Computer Science, 7926*, 249–258. This study is ineligible for review because it does not use a sample aligned with the protocol.

Mäkitalo-Siegla, K., & Fischer, F. (2011). Stretching the limits in help-seeking research: Theoretical, methodological, and technological advances. *Learning and Instruction, 21*(2), 243–246. The study is ineligible for review because it is out of scope of the protocol.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., Cohen, W. W., & Koedinger, K. R. (2011). Learning by teaching SimStudent—an initial classroom baseline study comparing with Cognitive Tutor. *Lecture Notes in Computer Science, 6738*, 213–221. The study is ineligible for review because it is out of scope of the protocol.

Meigs, J. G. (2012). *A study of the effects of the Carnegie Cognitive Tutor program on students' math achievement scores on the Alabama Reading and Math Test*. Troy, AL: Troy University. The study is ineligible for review because it does not use an eligible design.

Nye, B. D., Graesser, A. C., & Hu, X. (2014). Multimedia learning with intelligent tutoring systems. *Multimedia Learning*, 703–728. The study is ineligible for review because it is out of scope of the protocol.

Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction, 23*, 98–114. The study is ineligible for review because it is out of scope of the protocol.

Rau, M. A., & Scheines, R. (2012). Searching for variables and models to investigate mediators of learning from multiple representations. In K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 110–117). http://files.eric.ed.gov/fulltext/ED537203.pdf. The study is ineligible for review because it is out of scope of the protocol.

Reed, S. K., Corbett, A., Hoffman, B., Wagner, A., & MacLaren, B. (2013). Effect of worked examples and Cognitive Tutor training on constructing equations. *Instructional Science, 41*(1), 1–24. doi:10.1007/s11251-012-9205-x The study is ineligible for review because it does not use an eligible design.

Rigeman, S., & McIntire, N. (2005). Enhancing curriculum and instruction through technology: Mississippi Bend Districts implement a technology-rich, research-based math program to meet EETT Goals. *THE (Technological Horizons In Education) Journal, 32*(12), 31. The study is ineligible for review because it does not use an eligible design.

Ritter, S. (2011). *The research behind the Carnegie Learning Math Series*. Pittsburgh, PA: Carnegie Learning, Inc. The study is ineligible for review because it does not use an eligible design.

Ritter, S., & Anderson, J. (2006). *Cognitive Tutor: Tracking learning in real time*. Testimony to the National Mathematics Panel. The study is ineligible for review because it does not use an eligible design.

Ritter, S., Anderson, J., Koedinger, K., & Corbett, A. (2007). Cognitive Tutor®: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14*(2), 249–255. The study is ineligible for review because it does not use an eligible design.

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 does not use an eligible design.

Ritter, S., Joshi, A., Fancsali, S. E., & Nixon, T. (2013). Predicting standardized test scores from cognitive tutor interactions. *Proceedings of the 6th International Conference on Educational Data Mining*. The study is ineligible for review because it does not use an eligible design.

Ritter, S., Towle, B., Murray, R. C., Hausmann, R. G. M., & Connelly, J. (2010). A cognitive tutor for geometric proof. *Lecture Notes in Computer Science, 6095*, 453. The study is ineligible for review because it is out of scope of the protocol.

Ritter, S., Towle, B., Murray, R. C., Hausmann, R. G. M., & Connelly, J. (2010). Research-based improvements in Cognitive Tutor Geometry. *Lecture Notes in Computer Science, 6095*, 452. The study is ineligible for review because it is out of scope of the protocol.

Rittle-Johnson, B., & Koedinger, K. (2005). Designing knowledge scaffolds to support mathematical problem solving. *Cognition and Instruction, 23*(3), 313–349. The study is ineligible for review because it does not use an eligible design.

Roberge, D., Rojas, A., & Baker, R. (2012). Does the length of time off-task matter? *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 234–237. The study is ineligible for review because it is out of scope of the protocol.

Rodrigo, M. M. T., Baker, R. S. J. d., Agapito, J., Nabos, J., Repalam, M. C., & Reyes Jr., S. S. (2010). Comparing disengaged behavior within a cognitive tutor in the USA and Philippines. *Lecture Notes in Computer Science, 6095*, 263–265. The study is ineligible for review because it is out of scope of the protocol.

Roll, I., Aleven, V., Koedinger, K. R., Berland, M., Martin, T., Benton, T., … Pea, R. (2012). Building (timely) bridges between learning analytics, educational data mining and core learning sciences perspectives. *Proceedings* *from the 10th International Conference of the Learning Sciences: The Future of Learning, 2*, 134–141. The study is ineligible for review because it is out of scope of the protocol.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition—applying Cognitive Tutor principles to the tutoring of help seeking. *Metacognition and Learning, 2*(2), 125–140. The study is ineligible for review because it is out of scope of the protocol.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267–280. doi:10.1016/j.learninstruc.2010.07.004 The study is ineligible for review because it is out of scope of the protocol.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Metacognitive practice makes perfect: Improving students' self-assessment skills with an intelligent tutoring system. *Lecture Notes in Computer Science, 6738*, 288–295. The study is ineligible for review because it is out of scope of the protocol.

Roll, I., Baker, R. S. J. d., Aleven, V., & Koedinger, K. R. (2014). On the benefits of seeking (and avoiding) help in online problem-solving environments. *Journal of the Learning Sciences*, 23(4), 537–560. The study is ineligible for review because it is out of scope of the protocol.

Rummel, N., Mullins, D., & Spada, H. (2012). Scripted collaborative learning with the Cognitive Tutor Algebra. *International Journal of Computer-Supported Collaborative Learning, 7*(2), 307–339. doi:10.1007/s11412-012-9146-z This study is ineligible for review because it does not use a sample aligned with the protocol.

Salden, R. J. C. M., 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 is ineligible for review because it is out of scope of the protocol.

Salden, R. J. C. M., Aleven, V. A., 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 is out of scope of the protocol.

Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review, 22*(4), 379–392. The study is ineligible for review because it does not use an eligible design.

Schofield, J. W., Evans-Rhodes, D., & Huber, B. R. (1990). Artificial intelligence in the classroom: The impact of a computer-based tutor on teachers and students. *Social Science Computer Review, 8*(1), 24–41. The study is ineligible for review because it does not use an eligible design.

Schwonke, R., Ertelt, A., Otieno, C., Renkl, A., Aleven, V., & Salden, R. J. C. M. (2013). Metacognitive support promotes an effective use of instructional resources in intelligent tutoring. *Learning and Instruction, 23*, 136–150. This study is ineligible for review because it does not use a sample aligned with the protocol.

Schwonke, R., Renkl, A., Salden, R., & Aleven, V. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior, 27*(1), 58–62. doi:10.1016/j.chb.2010.03.037 The study is ineligible for review because it is out of scope of the protocol.

Shapiro, L. J., Sueker, E., & Koedinger, K. (1998, April). *Quantitative literacy development algebra*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA. The study is ineligible for review because it is out of scope of the protocol.

Slavin, R. E. (2005). *Evidence-based reform: Advancing the education of students at risk*. Report prepared for Renewing Our Schools, Securing Our Future: A National Task Force on Public Education (A joint initiative of the Center for American Progress and the Institute for America's Future). The study is ineligible for review because it does not use an eligible design.

Soliman, M., & Guetl, C. (2011). A survey of pedagogical functions of intelligent agents in virtual learning environments. *Journal of Internet Technology, 12*(6), 995–1005. The study is ineligible for review because it does not use an eligible design.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K-12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987. The study is ineligible for review because it does not use an eligible design.

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. This study is ineligible for review because it does not use a sample aligned with the protocol.

Tienken, C. H., & Maher, J. A. (2008). The influence of computer-assisted instruction on eighth grade mathematics achievement. *Research in Middle Level Education Online, 32*(3), 1–13. http://files.eric.ed.gov/fulltext/EJ827007.pdf. The study is ineligible for review because it is out of scope of the protocol.

Uzunboylu, H., Bicen, H., & Cavus, N. (2011). The efficient virtual learning environment: A case study of Web 2.0 tools and Windows Live spaces. *Computers and Education, 56*(3), 720–726. The study is ineligible for review because it does not use an eligible design.

VanHoudnos, N. (2014). *The efficacy of the Hedges correction for unmodeled clustering, and its generalizations in practical settings* (Unpublished doctoral dissertation). Carnegie Mellon University, Pittsburgh. The study is ineligible for review because it does not use an eligible design.

VanLehn, K. (2013). Model construction as a learning activity: A design space and review. *Interactive Learning Environments, 21*(4), 371–413. The study is ineligible for review because it is out of scope of the protocol.

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 does not use an eligible design.

Walker, E. (2010). *Automated adaptive support for peer tutoring* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 857079814) The study is ineligible for review because it is out of scope of the protocol.

Walker, E., McLaren, B. M., Rummel, N., & Koedinger, K. (2007). Who says three's a crowd? Using a cognitive tutor to support peer tutoring. *Frontiers in Artificial Intelligence and Applications, 158*, 399–408. The study is ineligible for review because it is out of scope of the protocol.

Walker, E., Ogan, A., Baker, R. S. J. d., de Carvalho, A., Laurentino, T., Rebolledo-Mendez, G., & Castro, M. J. (2011). Observations of collaboration in Cognitive Tutor use in Latin America. *Lecture Notes in Computer Science, 6738*, 575–577. The study is ineligible for review because it is out of scope of the protocol.

Walker, E., Rummel, N., & Koedinger, K. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning, 6*(2), 279–306. doi:10.1007/s11412-011-9111-2 The study is ineligible for review because it is out of scope of the protocol.

Walkington, C., Clinton, V., Ritter, S. N., & Nathan, M. J. (2015, November). How readability and topic incidence relate to performance on mathematics story problems in computer-based curricula. *Journal of Educational Psychology, 7*(4), 1051–1074. The study is ineligible for review because it does not use an eligible design.

Walkington, C., Clinton, V., Ritter, S., Nathan, M., & Fancsali, S. E. (2014, July). *The impact of cognitive and non-cognitive text-based factors on solving mathematics story problems*. Paper presented at the 7th International Conference of Educational Data Mining, London. The study is ineligible for review because it does not use an eligible design.

Walkington, C., Petrosino, A., & Sherman, M. (2013). Supporting algebraic reasoning through personalized story scenarios: How situational understanding mediates performance. *Mathematical Thinking and Learning, 15*(2), 89–120. The study is ineligible for review because it is out of scope of the protocol.

Walkington, C., & Sherman, M. (2012). Using adaptive learning technologies to personalize instruction: The impact of interest-based scenarios on performance in algebra. *Proceedings of the 10th International Conference of the Learning Sciences*, 80–87. The study is ineligible for review because it is out of scope of the protocol.

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology, 105*(4), 932–945. doi:10.1037/a0031882 The study is ineligible for review because it is out of scope of the protocol.

Xu, Y., & Mostow, J. (2011, July). *Logistic regression in a dynamic bayes net models multiple subskills better!* Poster presented at the 4th International Conference on Educational Data Mining Eindhoven, Netherlands. The study is ineligible for review because it is out of scope of the protocol.

Zimmerman, J. E. (2004). *The impact of Cognitive Tutor software on student performance in college intermediate algebra* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3103843) This study is ineligible for review because it does not use a sample aligned with the protocol.

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

Cabalo, J. V., Jaciw, A., & Vu, M. (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 *Cognitive Tutor® Algebra I* findings

Meets WWC group design standards without reservations

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

Setting

The study took place at Maui Community College and in five schools (grades 8–12) within the Maui School District, both located in Maui County, Hawaii. A total of nine teachers and 22 Algebra I classrooms participated in the study. 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.

Study sample

After giving an informational session to a group of teachers in the Maui School District, nine teachers volunteered to participate in a study of the effectiveness of Carnegie Learning's *Cognitive Tutor® Algebra I*. When possible, each teacher's classes were paired, and one class per pair was assigned to the intervention. When pairing was not possible, a class was randomly assigned to the intervention based on a coin toss. Pre-intervention math achievement data were collected in fall 2005, and a posttest evaluation was administered in May 2006. The authors included only students with both a pretest and a posttest in their analysis. The analytic sample included a total of 344 students (182 intervention, 162 comparison) from the 22 study classrooms (11 intervention, 11 comparison).

Intervention group

As reported by the authors, the curriculum is designed for 40% computer lab time and 60% classroom activities, combining software-based individualized lessons with collaborative problem-solving activities. In the intervention classes, teachers implemented the Cognitive Tutor® Algebra I classroom activities in fall and added lab activities to classroom activities in January. Each teacher taught both an intervention class and a comparison class.

Comparison group

In the comparison classrooms, teachers continued to use the curriculum in use at the time of study implementation. The authors indicated a variety of Algebra I curricula were used, including those published by Addison-Wesley, Bittinger and Beecher, Holt, McDougal Littell, Merrill, and Prentice Hall.

Outcomes and measurement

The eligible outcome measure was student scores on the Northwest Evaluation Association (NWEA) Algebra End-of-Course Achievement Level Test (a paper test administered to participating students enrolled in the Maui School District) or the Measure of Academic Progress (a computer-adapted version of the paper assessment administered to participating students enrolled at Maui Community College). The primary outcome was the overall score. The authors also analyzed substrand scores from the test, including Quadratic Equations, Algebraic Operations, Linear Equations, and Problem Solving. The analytic sample differed on each of the substrands, since students needed to respond to enough items in a substrand to receive a score. Two of the substrand scores meet WWC standards and are reported in Appendix D. These supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Teachers implementing *Cognitive Tutor® Algebra I* received 3 days of professional development led by a consultant from the curriculum developer. Teachers received a brief classroom observation and had the opportunity to ask questions of a developer representative early in the implementation period, but were given flexibility in actual classroom implementation. No ongoing technical assistance was provided.

Appendix A.2: Research details for Ritter et al. (2007)

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.

Additional sources:

Carnegie Learning, Inc. (2004). *Carnegie Learning research report: Moore Independent School District*. Pittsburgh, PA: Author.

Morgan, P., & Ritter, S. (2002). *An experimental study of the effects of Cognitive Tutor® Algebra I on student knowledge and attitude*. Pittsburgh, PA: Carnegie Learning, Inc. Retrieved from http://www.carnegielearning.com

Table A2. Summary of *Cognitive Tutor® Algebra I* findings — **Meets WWC group design standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Algebra	16 classrooms/255 students	+15	No

Setting

The study took place in three junior high schools in the Moore Independent School District, a suburban school district located near Oklahoma City.

Study sample

The study began with 426 ninth-grade students (206 intervention, 220 comparison) who were assigned to one of six algebra teachers in three study schools.[11] Algebra course sections for each teacher were randomly assigned to a curriculum. To reduce the cost of the Algebra I assessment, only one comparison class was randomly selected for testing for each teacher participating in the study; thus, the baseline sample of students within tested classrooms included 206 intervention students and 142 comparison students. The analytic sample included a total of 255 students (153 intervention, 102 comparison) from the 16 study classrooms (10 intervention, six comparison).

Intervention group

Cognitive Tutor® Algebra I was implemented during the 2000–01 school year, the first year the six study teachers implemented the curriculum. Students spent three class periods per week in group activities and classroom discussions using the *Cognitive Tutor® Algebra I* text and two class periods per week working on problem-solving skills with the *Cognitive Tutor® Algebra I* software.

Comparison group

Students in the comparison group were taught using *Heath Algebra I* published by McDougal Littell, which the authors refer to as a traditional algebra curriculum. The six study teachers taught both intervention and comparison classrooms in each of the three schools. At the start of the study, teachers had several years of experience using *Heath Algebra I*.

Outcomes and measurement

The primary outcome measure was the Educational Testing Service (ETS) End-of-Course Algebra Test. The authors also examined three other outcomes that were not taken into account in the effectiveness rating: first semester grades, second semester (final) grades, and student attitudes. These outcomes are not eligible for this review. For a more detailed description of the eligible outcome measure, see Appendix B.

Support for implementation

All study teachers implemented *Cognitive Tutor® Algebra I* for the first time. During the summer prior to the start of the school year, teachers attended a 4-day training course to familiarize themselves with the *Cognitive Tutor®* software, and to learn pedagogical approaches to the classroom instruction and techniques for integrating the software and classroom instruction.

Appendix A.3: 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.

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: Report to Congress*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Table A3. Summary of *Cognitive Tutor® Algebra I* findings — **Meets WWC group design standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Algebra	9 schools/270 students	–9	No

Setting

This 2-year congressionally-mandated study of education technology included several algebra interventions, including *Cognitive Tutor® Algebra I*. Findings from the first year of the study pooled all of the algebra interventions together. Therefore, this WWC intervention report focuses on findings from the second year, in which results are reported separately by intervention. In the second year of the study, *Cognitive Tutor® Algebra I* was implemented in nine schools in four districts. Districts were located in urban and urban fringe areas. Within each of the nine study schools, teachers were randomly assigned to the intervention and comparison groups prior to the first year of the study. Teachers maintained their assignment in the second year of the study.

Study sample

Schools were eligible to be in the study if they were in high-poverty areas, had no prior education technology use, and had at least two algebra teachers. The 18 study teachers averaged 14 years of teaching experience, and approximately half of them had a master's degree. Of the 270 study students (142 intervention, 128 comparison), approximately 50% were female, and the average age was 15.

Intervention group

Students in the intervention group were taught using *Cognitive Tutor® Algebra I* as their core math curriculum. The curriculum covered proportional reasoning, solving linear equations and inequalities, solving systems of linear equations, analyzing data, and using polynomial functions, powers, and exponents. Teachers were in their second year of implementing *Cognitive Tutor® Algebra I*, while students were experiencing the curriculum for the first time. According to the study authors, students in the study used the computer portion of the curriculum for an average of 1,840 minutes during 18 weeks.

Comparison group

Students in the comparison classes received traditional algebra instruction using their district's standard algebra curriculum. The comparison curricula were not specified and may have varied across districts and schools.

Outcomes and measurement

The study team administered the ETS End-of-Course Algebra I Test to all study students at the beginning and end of the school year. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

Within each district, teachers in the intervention group received 4 days of initial training from the publisher in the summer of 2004. They were trained on classroom management and the curriculum, and provided with opportunities to practice using the product. An unspecified amount of phone and email support was provided throughout the study.

Appendix A.4: Research details for Pane et al. (2014)

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale (high school experiment). *Educational Evaluation and Policy Analysis, 36*(2), 127–144. doi:10.3102/0162373713507480

Additional sources:

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). *Addendum to effectiveness of Cognitive Tutor Algebra I at scale* (Working Paper WR-1050-DEIES) (high school experiment). Santa Monica, CA: RAND Corporation.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., Karam, R., Daugherty, L., & Phillips, A. (2013). *Does an algebra course with tutoring software improve student learning?* (high school experiment). Santa Monica, CA: RAND Corporation.

Sales, A. C., & Pane, J. F. (2015, June). *Exploring causal mechanisms in a randomized effectiveness trial of the Cognitive Tutor* (high school experiment). Paper presented at the 8th International Conference on Educational Data Mining, Madrid, Spain.

Table A4. Summary of *Cognitive Tutor® Algebra I* findings — **Meets WWC group design standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Algebra	73 schools/11,066 students	+3	No

Setting The study took place in 51 school districts across seven states, including urban districts in Alabama, Connecticut, New Jersey, and Texas; suburban districts in Michigan; and rural districts in Kentucky and Louisiana.

Study sample Within the participating districts, the study included 73 high schools that participated for 2 consecutive school years. The analytic sample included 11,066 high school students: 5,328 students in the first year of the study and 5,738 in the second, without imputed data.[12] The analysis was conducted separately by cohort, as the researchers hypothesized that implementation of the program could be better during the second year due to teachers gaining experience with the curriculum. Participating students were enrolled in Algebra I; most were in ninth grade.[13] Of the 11,066 students in the sample, 5,072 were in the intervention group, and 5,994 were in the comparison group. Of the 73 study schools, 36 were in the intervention group, and 37 were in the comparison group.

Intervention group *Cognitive Tutor® Algebra I* was implemented for 2 consecutive school years. The curriculum involved teacher-directed classroom instruction 3 days a week and computer-guided instruction 2 days a week. The software was available for students to use during class and other times during the day. It was self-paced, and students' progress was based on mastery of the material. During the classroom lessons, students were exposed to topics such as solving linear equations, mathematical modeling with linear and quadratic expressions, problem solving using proportion reasoning, and analyzing data and making predictions. To apply these concepts, students completed worksheets and other activities and engaged in a variety of problem solving strategies. An implementation study indicated that teachers generally implemented all components of *Cognitive Tutor® Algebra I*, but sometimes emphasized the components differently from the publisher recommendations.

Comparison group Students in the comparison condition received each school's existing Algebra I curriculum, which included curricula published by Glencoe, McDougal Littell, and Prentice Hall.

Outcomes and measurement Student achievement was measured using the Algebra Proficiency Exam, a 32-item multiple-choice standardized test from the Acuity series created by CTB/McGraw-Hill. It is designed as an end-of-course exam for Algebra I. Students were also assessed using a 17-item researcher-created tool to measure student attitudes toward algebra, math, and computers; this outcome is not eligible for this WWC review. For a more detailed description of the eligible outcome measure, see Appendix B.

Support for implementation Teachers received 4 days of training throughout the study. During a 3-day session prior to the school year, teachers were introduced to the curriculum, software, and tools, and provided instruction on connections between the curriculum and software and how to use the data to inform instruction. Teachers received a fourth day of training during the school year, at which time professional development staff observed classrooms, offered recommendations, and helped with any problems the teachers had. In addition, teachers received training materials, an implementation guide, and a book of resources and assessments.

Appendix A.5: 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 A5. Summary of *Cognitive Tutor® Algebra I* findings — **Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings: Average improvement index *(percentile points)*	Study findings: Statistically significant
General mathematics achievement	24 schools/658 students	+2	No

Setting During the 2000–01 school year, nine senior high schools within the Miami–Dade County Public School District implemented *Cognitive Tutor® Algebra I*. Of those nine schools, six were selected to participate in the study because they had computer labs as of October 2000 (to support implementation of *Cognitive Tutor® Algebra I*) and because not all algebra classes were using *Cognitive Tutor® Algebra I*.

Study sample Within each of the six study schools, two teachers were randomly selected among teachers who were implementing *Cognitive Tutor® Algebra I*. Then one class was randomly selected from each teacher, creating an intervention sample of 12 classrooms with 325 students. The comparison sample was composed of 12 classrooms with 452 students within the same six schools, where classrooms were randomly selected from a pool of classrooms not implementing *Cognitive Tutor® Algebra I*. Among the baseline student sample, 54% were eligible for free and reduced-price meals, and the race/ethnicity was comparable across intervention and comparison groups (56% Hispanic, 30% African American, and 13% White among intervention students; 62% Hispanic, 27% African American, and 10% White among comparison students). In addition, gender was comparable across groups (46% and 48% female for intervention and comparison, respectively). Most of the students were in grades 9 and 10: 79% and 18%, respectively, for the intervention group, and 88% and 11%, respectively, for the comparison group. Each group contained a small number of students in grades 11 and 12, but they were excluded from the analyses due to a lack of baseline test scores. The analyses were conducted on 276 intervention and 382 comparison students, all of whom were in grades 9 and 10. The main results presented in this review pool together the two grades. In addition, supplemental analyses for the tenth-grade subgroup are presented in Appendix D.[14]

Intervention group

Students in the intervention group were taught using *Cognitive Tutor® Algebra I* for a full school year. Students worked with the curriculum in a computer lab 2 days per week, and they worked in the classroom on small-group activities 3 days per week. One study school had a functioning computer lab at the beginning of the school year, but the other four schools did not have operational computer labs until October, which according to the study author, could have affected the implementation of the software component of the intervention within these schools.

Comparison group

Students in the comparison group received Algebra I instruction using a curriculum other than *Cognitive Tutor® Algebra I*; the comparison curriculum was not named by the author.

Outcomes and measurement

The primary outcome measure was the Florida Comprehensive Assessment Test–Norm Referenced Test 2001 (FCAT-NRT). The study also examined student performance on the ETS End-of-Course Algebra I Test; however, based on data received from the author by the WWC in response to a query, the analysis sample used in the ETS Algebra I End-of-Course Assessment analysis was not equivalent at baseline across the intervention and comparison groups. Therefore, only findings related to the FCAT-NRT are included in this review. The author also presented subgroup findings on the FCAT-NRT by gender and grade. The results for the tenth-grade subgroup meet WWC standards and are presented in Appendix D. These supplemental findings do not factor into the intervention's rating of effectiveness. The study also used the Mathematics Attitude Comparison survey to assess student attitudes towards mathematics; this outcome is not eligible for this review. For a more detailed description of the eligible outcome, see Appendix B.

Support for implementation

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

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.

Additional sources:

Carnegie Learning, Inc. (2001). *Report of results from Pittsburgh, Pennsylvania* (Cognitive Tutor research report PA-91-01). Pittsburgh, PA: Author.

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.

Table A6. Summary of *Cognitive Tutor® Algebra I* findings

Meets WWC group design standards with reservations

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Algebra	3 schools/247 students	+19	Yes

Setting

The study took place in three high schools (Langley, Brashear, and Carrick) in the Pittsburgh Public Schools District. In the study schools, 50% of the student body were African American, 50% came from one-parent families, and 15% went on to attend college.

Study sample

The intervention group included 26 classes from all three schools, and the comparison group included five classes from two of the three schools. Intervention and comparison classes were matched on the basis of student math grades from the previous school year. Among the full study sample, 34% were African American, 56% were female, and 60% were eligible for free or reduced-price meals. Sixty-five percent of the sample were in the ninth grade, 24% were in the tenth grade, 8% were in the eleventh grade, and 2% were in the twelfth grade.

Intervention group

Students in the intervention group were taught in the 1993–94 school year using an early version of the *Cognitive Tutor®* software, then referred to as the *Pittsburgh Urban Mathematics Project curriculum plus Practical Algebra Tutor program (PUMP + PAT)*. 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; therefore, this early version varies from the current version of *Cognitive Tutor®*, which uses software for about 40% of instructional time.

Comparison group

Students in the comparison group were taught using their schools' traditional Algebra I curricula, which were not specified in the study.

Outcomes and measurement

The study utilized four outcome measures that were administered over 2 days at the end of the spring semester. Three outcomes fell within the algebra domain: (1) the Iowa Algebra Aptitude Test (IAAT); (2) the Complex Problem Solving Test, referred to as a Problem Situations Test, a researcher-designed measure; and (3) the Multiple Representations Test of algebraic concepts, a researcher-designed measure. The study also included a subset of questions from the Math subtest from the Scholastic Aptitude Test (SAT), which fell within the general mathematics achievement domain.

The IAAT was administered to all students on the first day. According to the authors, on the second day, students were administered two assessments randomly selected from the other three. 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 for the Problem Situations Test assessment sample. Therefore, only findings related to the sample of students who took the IAAT and the Multiple Representations Test are included in this review. For a more detailed description of the eligible outcome measures, see Appendix B.

Support for implementation

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

Appendix A.7: 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 A7. Summary of *Cognitive Tutor® Geometry* findings

Meets WWC group design standards with reservations

		Study findings	
Outcome domain	**Sample size**	**Average improvement index *(percentile points)***	**Statistically significant**
Geometry	8 schools/669 students	–8	Yes

Setting

The study was conducted in eight high schools in the Baltimore County Public School District. Two of the study schools participated in each of 3 academic years (2005–06 to 2007–08), three participated for 2 years, and three participated for 1 year.

Study sample

In each academic year, two geometry teachers at each participating school took part in the study, and each teacher taught a morning geometry period and an afternoon geometry period. The researchers randomly assigned the two study morning periods to use either *Cognitive Tutor® Geometry* or the school's usual geometry curriculum. Then, each teacher was assigned to teach using the opposite curriculum in the afternoon. Thus, each study teacher taught one period using *Cognitive Tutor® Geometry* and one period using the school's usual geometry curriculum.

Prior to the start of each study school year, students who signed up for geometry were randomly assigned to the intervention and comparison classrooms. After excluding the students who never attended the classes, the non-randomized students who entered study classrooms after the start of school, and the randomized students with missing posttest scores, the analytic sample included 669 students (329 intervention, 340 comparison). Among the analytic student sample, 76% were minorities, and 36% were eligible for free or reduced-price meals.

Intervention group

Students in the intervention classrooms were taught using *Cognitive Tutor® Geometry* for an entire academic school year. The curriculum included teacher-directed classroom instruction (60% of classroom time) and computer-guided individual instruction (40% of classroom time). Both components focused on inductive problem solving. During classroom instruction, the teacher led students through math problems and assisted them as they worked in groups to solve additional problems, after which students presented their groups' work to the class. During computer-guided instruction, students worked through math problems matched to their current ability. The software has several interactive, feedback-driven features, so it tailors the difficulty of the problem to the student's demonstrated mastery. The student is able to ask the software for hints when solving problems.

Comparison group

Students in the comparison classrooms were taught using the school's standard geometry curriculum, which was not specified in the study.

Outcomes and measurement

The primary outcome measure used by study authors was the Baltimore County Public School District Geometry Assessment. The study also included the Survey of High School Math Students, a measure of students' mathematics confidence and attitudes. This outcome is not eligible for this review. For a more detailed description of the eligible outcome measure, see Appendix B.

Support for implementation

Each study teacher received 3 days of training on *Cognitive Tutor® Geometry* prior to using the curriculum. In addition, each teacher received 1 day of follow-up training during the school year. The curriculum included a pacing guide designed by the district to ensure that the district's required geometry content would be covered in the intervention classrooms.

Appendix B: Outcome measures for each domain

Algebra	
CTB/McGraw-Hill Algebra Proficiency Exam	The Algebra Proficiency Exam is a standardized test from the CTB/McGraw-Hill Acuity series (as cited in Pane et al., 2014). This is a 32-item multiple-choice assessment designed to measure mastery of Algebra I content knowledge at the end of the course. The exams were scored using a three-parameter item response theory (IRT) model, and scores were standardized within the population to have a mean of 0 and standard deviation of 1.
Educational Testing Service (ETS) End-of-Course Algebra Test	The ETS End-of-Course Algebra Test is a commercially available standardized test based on Algebra I standards of the National Council of Teachers of Mathematics (as cited in Ritter et al., 2007 and Campuzano et al., 2009). The questions are designed to assess students' understanding of algebraic concepts, processes, and skills and are reported on a 0–50 scale. As reported in Ritter et al. (2007), the test included 25 multiple-choice items and 15 constructed-response items, with each type of question accounting for 50% of the student's score. As reported in Campuzano et al. (2009), the study team worked with ETS to develop two separate versions of the test (for the pre- and posttest assessments) that had equal levels of difficulty, and scores were reported as a percent correct.
Iowa Algebra Aptitude Test (IAAT)	The IAAT is a standardized algebra test consisting of 60 items (as cited in Wolfson et al., 2008). This commercially available test was developed by the University of Iowa to measure algebra readiness and includes questions on pre-algebra number skills and concepts, interpreting mathematical information, representing relationships, and using symbols. The test takes approximately 50 minutes to complete.
Multiple Representations Test	The Multiple Representations Test is a nine-item researcher-designed measure created to assess students' abilities to translate between representations of algebraic content including verbal descriptions, graphs, and symbolic equations (as cited in Wolfson et al., 2008). The reported Cronbach's alpha of the assessment was .64.
Northwest Evaluation Association (NWEA) Algebra End-of-Course Achievement Level Test	The NWEA Algebra End-of-Course Achievement Level Test is a comprehensive test that measures students' instructional level. The test is scored on a Rasch 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. The study administered a paper version of the NWEA Algebra End-of-Course Achievement Level Test to study students enrolled in the Maui School District. A computer-adapted version of the NWEA Algebra End-of-Course assessment, called the Measure of Academic Progress, was administered to study students enrolled at Maui Community College. These results on the two versions of the test were combined by the authors (as cited in Cabalo et al., 2007).
General mathematics achievement	
Florida Comprehensive Assessment Test–Norm Referenced Test 2001 (FCAT-NRT)	The FCAT-NRT includes 48 multiple choice questions in 10 areas that range from problem solving to precalculus (as cited in Shneyderman, 2001). The Kuder-Richardson reliability was .92 for the ninth-grade test and .93 for the tenth-grade test.
Geometry	
Baltimore County Public School District (BCPS) Geometry Assessment	The BCPS Geometry Assessment includes 30 multiple choice items and 11 extended response items administered as a final exam with a potential score that ranges from 0–50 (as cited in Pane et al., 2010). The authors reported interrater reliability statistics from the test scores for the extended response questions in the range of .97 to .99. Carnegie Learning rated the degree to which each item on the BCPS Geometry Assessment was covered by *Cognitive Tutor*® and found that 80% of multiple choice items were covered "thoroughly" by the curriculum, while only 3% of items were not covered.

Appendix C.1: Findings included in the rating for studies of *Cognitive Tutor® Algebra I* for the algebra 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]								
Northwest Evaluation Association (NWEA) Algebra End-of-Course Achievement Level Test	Grades 8–13	6 schools/ 344 students	243.37 (7.67)	244.71 (7.47)	–1.34	–0.18	–7	.23
Domain average for algebra (Cabalo et al., 2007)						**–0.18**	**–7**	**Not statistically significant**
Campuzano et al. (2009)[b]								
Educational Testing Service (ETS) End-of-Course Algebra Test	Grades 8 and 9	9 schools/ 270 students	32.39 (11.13)	35.31 (14.33)	–2.92	–0.23	–9	.30
Domain average for algebra (Campuzano et al., 2009)						**–0.23**	**–9**	**Not statistically significant**
Pane et al. (2014)[c]								
CTB/McGraw-Hill Algebra Proficiency Exam	Grades 8–12 (Year 1)	73 schools/ 5,328 students	–0.02 (0.75)	0.03 (0.78)	–0.05	–0.06	–3	.46
CTB/McGraw-Hill Algebra Proficiency Exam	Grades 8–12 (Year 2)	73 schools/ 5,738 students	0.08 (0.81)	–0.08 (0.74)	0.16	0.21	+8	.04
Domain average for algebra (Pane et al., 2014)						**0.08**	**+3**	**Not statistically significant**
Ritter et al. (2007)[d]								
ETS End-of-Course Algebra Test	Grade 9	16 classrooms/ 255 students	17.41 (5.82)	15.28 (5.33)	2.13	0.38	+15	.09
Domain average for algebra (Ritter et al., 2007)						**0.38**	**+15**	**Not statistically significant**
Wolfson et al. (2008)[e]								
Iowa Algebra Aptitude Test	Grades 9–12	3 schools/ 247 students	0.50 (0.17)	0.45 (0.17)	0.04	0.26	+10	.02
Multiple Representations Test	Grades 9–12	3 schools/ 117 students	3.54 (1.68)	2.34 (1.68)	1.20	0.71	+26	< .001
Domain average for algebra (Wolfson et al., 2008)						**0.49**	**+19**	**Statistically significant**
Domain average for algebra across all studies						**0.11**	**+4**	**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 each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Cabalo et al. (2007), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. 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 hierarchical linear model (HLM) coefficient for the difference between the two groups in the study. The standard deviations are based on the unadjusted group means. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

[b] For Campuzano et al. (2009), means and standard deviations displayed in this table were obtained through an author query. The values in this table differ from the results reported in an earlier WWC intervention report, which included cases with imputed pretest values. A correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-value presented here was reported in the original study. The WWC calculated the program group mean using a difference-in-differences approach 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. By agreement between the authors and publishers, interventions were not named in the first year of the study, but were in the second. Therefore, this report provides results of the second cohort of students in the study; findings from the first year of the study are ineligible for review because they pooled multiple algebra interventions together. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

[c] For Pane et al. (2014), means and standard deviations displayed in this table were obtained through an author query. A correction for multiple comparisons was needed and resulted in a WWC-computed critical *p*-value of .025 for the Algebra Proficiency Exam in Cohort B; therefore, the WWC does not find the result to be statistically significant. The *p*-values presented here were reported in the original study. The analysis was conducted separately by cohort, as the researchers hypothesized that implementation of the program could be better during the second year due to teachers gaining experience with the curriculum. The study was also conducted with middle school students, but it could not demonstrate equivalence for the middle school student analytic sample, so the results are not included in this report. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

[d] For Ritter et al. (2007), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-value presented here was reported in the original study. Means are the unadjusted means. The means reported in this table differ slightly from those reported in an earlier WWC intervention report, in which the mean for the intervention group was computed as the unadjusted comparison group mean plus the program coefficient from the HLM analyses. This study is characterized as having a substantively important positive effect because the estimated effect is positive and not statistically significant but is larger than 0.25 standard deviations. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

[e] For Wolfson et al. (2008), sample sizes, means, and standard deviations displayed in this table were obtained through an author query. The outcomes and values in this table differ from those reported in an earlier WWC intervention report; the data reported here were obtained through an updated author query. Corrections for clustering and multiple comparisons were needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-values presented here were reported in the original study. 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 Procedures and Standards Handbook (version 3.0), p. 26.

Appendix C.2: Findings included in the rating for studies of *Cognitive Tutor® Algebra I* for the general 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**
Shneyderman (2001)[a]								
Florida Comprehensive Assessment Test–Norm Referenced Test 2001	Grades 9 and 10	6 schools/ 658 students	683.88 (29.80)	682.47 (27.77)	1.41	0.05	+2	> .05
Domain average for general mathematics achievement domain (Shneyderman, 2001)					**1.41**	**0.05**	**+2**	**Not statistically significant**
Domain average for general mathematics achievement domain across all studies						**0.05**	**+2**	**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. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Shneyderman (2001), means and standard deviations displayed in this table were obtained through an author query. A correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-value presented here was reported in the original study. The WWC calculated the program group mean using a difference-in-differences approach 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. The mean difference reported in this table differ from that reported in the earlier version of the report. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix C.3: Findings included in the rating for studies of *Cognitive Tutor® Geometry* in the geometry domain

			Mean (standard deviation)		WWC calculations			
Outcome measure	**Study sample**	**Sample size**	**Intervention group**	**Comparison group**	**Mean difference**	**Effect size**	**Improvement index**	***p*-value**
Pane et al. (2010)[a]								
Baltimore County Public School District Geometry Assessment	High school students	8 schools/ 669 students	17.01 (8.84)	18.79 (9.49)	–1.78	–0.19	–8	.03
Domain average for geometry (Pane et al., 2010)						**–0.19**	**–8**	**Statistically significant**
Domain average for geometry across all studies						**–0.19**	**–8**	**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 statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Pane et al. (2010), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. 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 sample sizes, means, and standard deviation for this study differ from those in a prior WWC intervention report. The data in this report were obtained from an author query in 2015. This study is characterized as having a statistically significant negative effect because the effect for at least one measure within the domain is negative and statistically significant, and no effects are positive and statistically significant, accounting for multiple comparisons. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D: Description of supplemental findings of *Cognitive Tutor® Algebra I* for the general 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]								
Northwest Evaluation Association (NWEA) Algebra Test—Algebraic Operations Subscore	Grades 8–13	6 schools/ 345 students	241.03 (9.99)	243.50 (10.18)	–2.47	–0.24	–10	.16
NWEA Algebra Test—Problem Solving Subscore	Grades 8–13	6 schools/ 338 students	246.67 (11.90)	246.38 (10.69)	0.29	0.03	+1	.86
Shneyderman (2001)[b]								
Florida Comprehensive Assessment Test–Norm Referenced Test 2001	Grade 10	6 schools/ 92 students	688.50 (24.69)	693.60 (24.23)	–5.10	–0.21	–8	> .05

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or 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 Cabalo et al. (2007), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. 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 are the unadjusted standard deviations.

[b] For Shneyderman (2001), means and standard deviations displayed in this table were obtained through an author query. A correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-value presented here was reported in the original study. The WWC calculated the program group mean using a difference-in-differences approach 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. The mean difference reported in this table differs from that reported in the earlier version of the WWC intervention report. The slight difference results from analyses using a difference-in-differences approach, which was not used in the earlier version. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

Endnotes

* Due to the 2015 restructuring of the Mathematics topic area from three areas (Elementary, Middle, and High School) to two areas (Primary and Secondary Mathematics), this report is considered a new report rather than an updated report. The information in this report includes reviews of some, but not all, of the studies in the prior Middle and High School Mathematics reports, since not all studies in the prior reports are eligible for review under the Secondary Mathematics review protocol. Endnote 2 explains which studies from the prior reports are treated differently in this report.

On December 21, 2016, the WWC modified this report in response to an independent review by a quality review team. Based on the review, the WWC changed the description of intervention implementation for one study, clarified the reason for excluding findings from another study, and deleted an erroneous citation from the reference list. The WWC has not added studies to the evidence base, updated the literature search, changed any study ratings, or changed values presented in tables since the June 2016 report.

[1] The descriptive information for this program was obtained from a publicly available source: the publisher's website (www.carnegielearning.com, downloaded October 14, 2014). The WWC requests publishers review the program description sections for accuracy from their perspective. The program description was provided to the publisher in October 2014, and the WWC incorporated feedback from the publisher. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.

[2] The WWC previously released two reports for *Cognitive Tutor®*: *Cognitive Tutor® Algebra I* was reviewed under the Middle School Mathematics (MSM) review protocol and released in July 2009, and *Carnegie Learning Curricula and Cognitive Tutor®* was reviewed under the High School Mathematics (HSM) review protocol and released in January 2013. The literature search reflects documents publicly available by September 2015. This report includes reviews of 78 studies that were not included in the prior reports. Of the additional studies, 69 were not within the scope of the review protocol for the Secondary Mathematics topic area, eight were within the scope of the review protocol but did not meet WWC group design standards, and one met WWC group design standards with reservations. A complete list and disposition of all studies reviewed are provided in the references.

This report includes reviews of all studies that met WWC group design standards with or without reservations in the previous math reports. The reviews resulted in a revised disposition for three studies: Campuzano et al. (2009), Pane et al. (2010), and Smith (2001).

The Campuzano et al. (2009) study received a disposition in this report of *meets WWC group design standards with reservations*, where it had previously received the rating of *meets WWC evidence standards without reservations*. The study was previously reviewed using version 2.1 standards, and is currently reviewed using version 3.0 standards, which include a clarification in guidance regarding cluster randomized controlled trials. This updated guidance indicates that if the authors of a cluster randomized controlled trial study characterize the intervention as having effects on student scores (rather than only on cluster-level scores), and some students enter clusters after random assignment, then the study must demonstrate the equivalence of the analytic sample. As this study meets those criteria, and demonstrates equivalence as required, it is now rated *meets WWC group design standards with reservations*.

The Pane et al. (2010) study received a disposition in this report of *meets WWC group design standards with reservations*, where it had previously received the rating of *meets WWC evidence standards without reservations*. The revised disposition is due to an author query response where the authors clarified the number of students in the analytic sample at the time of randomization. Based on the revised sample size provided by the authors, the study has high subcluster attrition. Studies with high subcluster attrition must demonstrate equivalence of the analytic sample. Equivalence was demonstrated on an analytic sample without imputed data using data provided by the authors, so it is now rated *meets WWC group design standards with reservations*.

The Smith (2001) study received a disposition in this report of *does not meet WWC group design standards*, whereas it previously received the rating of *meets WWC evidence standards with reservations*. The study was previously reviewed using version 2.1 standards, and is currently reviewed using version 3.0 standards which include a clarification in guidance that imputed data cannot be used to demonstrate equivalence of the analytic sample. The author did not respond to the WWC's request for data that could be used to demonstrate equivalence, so it is now rated *does not meet WWC group design standards*.

The studies in this report were reviewed using the Standards from the WWC Procedures and Standards Handbook (version 3.0) and the Secondary Mathematics review protocol (version 3.1). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] Absence of conflict of interest: This intervention report includes a study conducted by staff from Mathematica Policy Research, Inc. Because Mathematica is one of the contractors that administers the WWC, the study was reviewed by staff members from a different organization. The report was then reviewed by the lead methodologist, a WWC Quality Assurance reviewer, and an external peer reviewer.

[4] Some studies had multiple outcomes with analytic sample sizes that varied by outcome. This represents the total analytic sample size across the five studies, which includes some double-counting of students.

[5] One study includes students in community college, which are described as grade 13 students in this report. These community college students are eligible for review under the Secondary Mathematics review protocol (version 3.1).

[6] This study was conducted in high schools, but the authors do not report the specific grades or ages of students enrolled in the study.

[7] This report combines the research and presents conclusions differently than two previous WWC intervention reports on *Cognitive Tutor®*. Some of these differences arise because this report combines studies in the previous reports with additional studies not previously reviewed, and provides separate effectiveness ratings for each secondary course and in multiple achievement outcome domains. In the 2009 MSM Report, *Cognitive Tutor® Algebra I* was found to have potentially positive effects on math achievement. In the 2013 HSM Report, *Carnegie Learning Curricula and Cognitive Tutor®* was found to have mixed effects on mathematics achievement. The new disposition for *Cognitive Tutor® Algebra I* is based on multiple studies and uses multiple achievement outcome domains, whereas the 2009 MSM disposition for *Cognitive Tutor® Algebra I* was based on one study. The new disposition for *Cognitive Tutor® Geometry* is based on one study (Pane et al., 2010), which is rated identically in the HSM and current report.

[8] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 36. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies.

[9] No studies examining the effectiveness of *Cognitive Tutor® Algebra II* or *Cognitive Tutor® Integrated Math I*, *II*, and *III* fall within the scope of the Secondary Mathematics review protocol and meet WWC group design standards. Because no studies meet WWC group design standards at this time, the WWC is unable to draw any conclusions based on research about the effectiveness or ineffectiveness of *Cognitive Tutor® Algebra II* or *Cognitive Tutor® Integrated Math I*, *II*, and *III* on secondary students. Additional research that meets WWC standards is needed to determine the effectiveness or ineffectiveness of these interventions.

[10] One study included in this total (Wolfson et al., 2008) has two different analytic sample sizes for the two outcome measures used in the study. The larger of this study's two analytic sample sizes was used in computing the total number of students across all studies.

[11] The study intended to include five schools; however, one school did not have the necessary computer resources to support *Cognitive Tutor® Algebra I* implementation, and another school did not adhere to the study design (they erroneously did not assign each teacher to at least one *Cognitive Tutor®* and one traditional class). The WWC reviewed this study under the assumption that these two schools are excluded from the analysis and do not count toward attrition rates.

[12] These sample sizes are based on unimputed data. Other analyses presented by the authors, but not included in this review, contained larger samples with imputed data. In addition to the high school study, the authors reported results of a middle school study based on 5,519 students in 74 middle schools (predominantly, students in eighth grade). Because the study defined the student sample after the school random assignment was conducted, equivalence of the analytic sample is required. However, the middle school study does not demonstrate equivalence and therefore is rated *does not meet WWC group design standards*.

[13] The study also conducted analyses for subgroups of students who spent varying levels of time using the *Cognitive Tutor® Algebra I* software. Because these are not subgroups that are identified under the review protocol for the Secondary Mathematics topic area, the results for these subgroup analyses are not presented as supplemental findings in this report.

[14] The ninth-grade subgroup does not meet standards as the groups were not equivalent at baseline, so it is not presented in Appendix D.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, June). *Secondary Mathematics intervention report: 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 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. 36.

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. 36.

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.

Intervention Report

Practice Guide

Quick Review

Single Study Review

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.