What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

High School Math August 2010

Carnegie Learning Curricula and Cognitive Tutor® Software

Program Description[1]

The combination of *Carnegie Learning Curricula and Cognitive Tutor® Software* merges algebra textbooks with interactive software developed around an artificial intelligence model that identifies strengths and weaknesses in an individual student's mastery of mathematical concepts. The software customizes prompts to focus on areas in which the student is struggling and routes the student to problems that address those specific concepts.

Research[2]

Two studies of the combination of *Carnegie Learning Curricula and Cognitive Tutor® Software* that fall within the scope of the High School Math review protocol meet What Works Clearinghouse (WWC) evidence standards, and two studies meet WWC evidence standards with reservations. The four studies included 1,723 high school students in 27 schools across 7 districts.[3]

Based on these four studies, the WWC considers the extent of evidence for the combination of *Carnegie Learning Curricula and Cognitive Tutor® Software* on high school students to be medium to large for mathematics achievement.

Effectiveness

Carnegie Learning Curricula and Cognitive Tutor® Software was found to have no discernible effects on mathematics achievement for high school students.

	Mathematics achievement	
Rating of effectiveness	No discernible effects	
Improvement index[4]	Average: –4 percentile points	Range: –7 to +2 percentile points

1. The descriptive information for this program was obtained from a publicly available source: the program's website (http://carnegielearning.com/secondary-curricula/, downloaded April 2010). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by January 2010.
2. The studies in this report were reviewed using WWC Evidence Standards, Version 2.0 (see the WWC Procedures and Standards Handbook, Chapter III), as described in protocol Version 2.0.
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
4. These numbers show the average and range of student-level improvement indices for all findings across the studies.

Absence of conflict of interest

The Campuzano et al. (2009) study summarized in this intervention report was prepared by staff of Mathematica Policy Research. Because the principal investigator for the WWC review of High School Math also is a Mathematica staff member, the study was rated by staff members from RAND and SRI. The report was then reviewed by the principal investigator, a WWC Quality Assurance reviewer, and an external peer reviewer.

Additional program information

Developer and contact

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

Scope of use

Pilot implementation of the program began in 1992 with 84 students in one school. As of January 2010, components of *Carnegie Learning Curricula and Cognitive Tutor® Software*—Bridge to Algebra, Algebra I, Geometry, Algebra II, and Integrated Math—have been used by more than 500,000 students in approximately 2,600 schools across the United States.[5]

Teaching

Carnegie Learning Curricula and Cognitive Tutor® Software addresses both mathematical content and process standards. Generally, three periods per week are spent using *Carnegie Learning Curricula* text for classroom activities, and two periods per week are spent in the computer lab using the *Cognitive Tutor® Software*. 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. Students learn with the adaptive software at their own pace. The math problems are designed to emphasize connections among verbal, numeric, graphic, and algebraic representations.

Cost

The Carnegie Learning website sells the following products related to Algebra I as of 2009–2010: Curriculum Kit (Software/Text) for $99, Teacher Text Set for $85, Software only for $84, and Student Text Set for $22. Volume discounts are available for school districts. The Campuzano et al. (2009) study estimates the annualized cost per student to be $69, of which 43% is the license fee and the remaining 57% is for teacher training and support, technical support, and printed materials and supplies.

Research

Twenty-four studies reviewed by the WWC investigated the effects of *Carnegie Learning Curricula and Cognitive Tutor® Software* on high school students. Two studies (Cabalo, Jaciw, & Vu, 2007; Campuzano et al., 2009) are randomized controlled trials that meet WWC evidence standards. Two studies (Shneyderman, 2001; Smith, 2001) are randomized controlled trials or quasi-experimental designs (QED) that meet WWC evidence standards with reservations. The remaining 20 studies do not meet either WWC evidence standards or eligibility screens.

Meets WWC evidence standards

Cabalo, Jaciw, and Vu (2007) randomly assigned 22 classrooms to receive either the *Carnegie Learning Curricula and Cognitive Tutor® Software* Algebra I program or the standard curriculum. Eight teachers taught at least one intervention class, and nine teachers taught at least one comparison class, at one of five Maui School District schools or at Maui Community College. The analysis sample consisted of 182 intervention and 162 comparison students who had taken both the pretest (in fall 2005) and the posttest (in May 2006).

5. The only available studies that meet WWC standards, with or without reservations, are those that cover *Cognitive Tutor®* for Algebra I.

Research *(continued)*

Campuzano et al. (2009) randomly assigned teachers in high-poverty schools to intervention and comparison groups as part of a national study of software products. During the second year of the study (presented in this report), *Carnegie Learning Curricula and Cognitive Tutor® Software* was implemented in nine schools in four districts. Nine teachers were randomly assigned to use the Algebra intervention, and nine were assigned to the comparison condition and used traditional instructional methods, with a pair of intervention and comparison teachers in each school. The fall and spring tests were administered to 145 intervention and 131 comparison students in 8th and 9th grades.

Meets WWC evidence standards with reservations

Shneyderman (2001) conducted a quasi-experiment in six senior high schools in Miami-Dade County, Florida, that had a computer lab by October 2000. For each school, two teachers were randomly selected from all teachers using the *Carnegie Learning Curricula and Cognitive Tutor® Software* Algebra I program. One class for each teacher was randomly selected into an intervention sample of 12 classrooms; the comparison sample was composed of 12 randomly selected nonintervention Algebra I classrooms in the same six schools. The analyses were conducted on 276 intervention and 382 comparison students in 9th and 10th grades.

Smith (2001) was a randomized controlled trial that was compromised by restrictions placed on the analysis sample after random assignment. Therefore, it was treated as a QED that demonstrated baseline equivalence of the analysis sample on a pretest and made the necessary statistical adjustments, allowing it to meet WWC evidence standards with reservations. The study involved all students in seven high schools in Virginia Beach City Public Schools who completed a three-semester Algebra I sequence during the 1999–2000 and 2000–01 school years. Students were randomly assigned to the sequence in which the math teacher was willing to implement the *Carnegie Learning Curricula and Cognitive Tutor® Software* program (229 students) or the sequence with the traditional curriculum (216 students).

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.[6]

The WWC considers the extent of evidence for *Carnegie Learning Curricula and Cognitive Tutor® Software* to be medium to large for high school students.

Effectiveness

Findings

The WWC review of interventions for High School Math addresses student outcomes in one domain: mathematics achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Carnegie Learning Curricula and Cognitive Tutor® Software* on high school students.[7]

Mathematics achievement. Cabalo, Jaciw, and Vu (2007) reported a negative but not statistically significant effect of *Carnegie Learning Curricula and Cognitive Tutor® Software* on the Northwest Evaluation Association (NWEA) Algebra End-of-Course Achievement Level Test/Measures of Academic Progress. Campuzano et al. (2009) reported a negative but not statistically significant effect of *Carnegie Learning Curricula and*

6. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Carnegie Learning Curricula and Cognitive Tutor® Software* is in Appendix A6.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. For the *Carnegie Learning Curricula and Cognitive Tutor® Software* studies summarized here, no corrections for clustering or multiple comparisons were needed.

Effectiveness *(continued)*

Cognitive Tutor® Software on the Educational Testing Service (ETS) Algebra I End-of-Course Assessment. Shneyderman (2001) reported a positive but not statistically significant effect of *Carnegie Learning Curricula and Cognitive Tutor® Software* on the Florida Comprehensive Assessment Test (FCAT) Norm-Referenced Component. Smith (2001) reported a negative but not statistically significant effect of *Carnegie Learning Curricula and Cognitive Tutor® Software* on the Virginia Standards of Learning (SOL) Algebra Assessment. None of the findings were large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25).

In sum, four studies showed indeterminate effects in the mathematics achievement domain.

Rating of effectiveness

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

The WWC found *Carnegie Learning Curricula and Cognitive Tutor® Software* to have no discernible effects on mathematics achievement for high school students

Improvement index

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

The average improvement index for mathematics achievement is –4 percentile points across the four studies, with a range of –7 to +2 percentile points across findings.

Summary

The WWC reviewed 24 studies on *Carnegie Learning Curricula and Cognitive Tutor® Software* for high school students. Two of these studies meet WWC evidence standards; two studies meet WWC evidence standards with reservations; the remaining 20 studies do not meet either WWC evidence standards or eligibility screens. Based on the four studies, the WWC found no discernible effects on mathematics achievement for high school students. The conclusions presented in this report may change as new research emerges.

References

Meets WWC evidence standards

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

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

Additional source:

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

References *(continued)*

Meets WWC evidence standards with reservations

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

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

Studies that fall outside the High School Math review protocol or do not meet WWC evidence standards

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

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

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

Baker, R., Corbett, A., Roll, I., & Koedinger, K. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction, 18*(3), 287–314. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Blessing, S. B., Gilbert, S. G., Oureda, S., & Ritter, S. (2009). Authoring model-tracing cognitive tutors. *International Journal of Artificial Intelligence in Education, 19*, 189–210. The study is ineligible for review because it does not examine the effectiveness of an intervention.

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

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

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science, 32*(2), 366–397. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Mac Iver, M. A., & Mac Iver, D. J. (2009). Urban middle-grade student mathematics achievement growth under comprehensive school reform. *Journal of Educational Research, 102*(3), 223–236. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

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

Additional source:

Plano, G. S. (2004). The effects of the *Cognitive Tutor* Algebra on student attitudes and achievement in a 9th-grade algebra course. (Doctoral dissertation, Seton Hall University,

References *(continued)*

2004). *Dissertation Abstracts International 65*(04A), 47–291. (AAI3130130).

Ritter, S. (2005). Authoring model-tracing tutors. *Technology, Instruction, Cognition and Learning, 2*(3), 231–247. The study is ineligible for review because it does not examine the effectiveness of an intervention.

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

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

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

Additional source:

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

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

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

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

Vinogradova, E., King, C., & Rhoades, T. (2008). *Success for all students: What works? Best practices in Maryland public schools.* Paper presented at the American Sociological Association Annual Meeting, Boston, MA. Retrieved from http://www.allacademic.com/meta/p_mla_apa_research_citation/2/4/1/0/5/p241050_index.html. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Wolfson, M., Koedinger, K., Ritter, S., & McGuire, C. (2008). Cognitive Tutor *Algebra I: Evaluation of results (1993–1994).* Pittsburgh, PA: Carnegie Learning, Inc. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Additional source:

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

Appendix

Appendix A1.1 Study characteristics: Cabalo, Jaciw, & Vu, 2007

Characteristic	Description
Study citation	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.
Participants	After an informational session with a group of teachers in the Maui School District, nine teachers volunteered to participate in a study of the effectiveness of the *Carnegie Learning Curricula and Cognitive Tutor® Software* Algebra I program. When possible, classes were paired based on class size and achievement level, with a coin toss determining which one of the pair would be assigned to the intervention group. Classes that were unable to be paired (when a teacher had an odd number of classes) were assigned to the intervention or comparison group by coin toss. Pre-intervention math achievement data were collected in fall 2005, and a posttest evaluation was administered in May 2006; only students with both tests were included in the analysis. Of the initial sample of 541 students (281 intervention and 260 comparison), 344 (182 intervention and 162 comparison) had both pre- and posttest scores. At the beginning of the study, students in grades 9–12 comprised 73% of the sample, with 19% in grade 8 and 7% enrolled at Maui Community College.[1]
Setting	The study took place in five schools within the Maui School District, and in Maui Community College, all located in Maui County, Hawaii. The Maui School District includes schools on two other islands, but only schools on Maui itself were part of this study. According to the authors, Maui County is a mixed suburban and rural community located on one of the seven islands of Hawaii. Nine teachers and 22 classrooms participated in the study. Within the participating Maui School District schools, students were 32% Filipino, 28% part-Hawaiian, 11% White, 8% Japanese, 5% Hawaiian, 3% Hispanic, and 14% other; the distribution of ethnicities at Maui Community College was similar. Approximately 27% of students participated in the National School Lunch Program, and approximately 6% were designated as limited English proficient.
Intervention	Classrooms selected for the intervention group implemented the *Carnegie Learning Curricula and Cognitive Tutor ® Software* Algebra I program. Selected classrooms utilized the intervention for six months, from October/November through the end of the 2005–06 school year.
Comparison	For the comparison classrooms, teachers continued to follow the textbook program in use at the time of study implementation, one of several branded Algebra I textbooks.
Primary outcomes and measurement	Student math achievement was measured by 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 Measure of Academic Progress (a computer-adapted version of the paper assessment administered to participating students enrolled at Maui Community College). For a more detailed description of these outcome measures, see Appendix A2. Results from both tests were combined by the authors and in the results presented in Appendix A3; the disaggregated results by subscale are presented in Appendix A4.
Staff/teacher training	Teachers utilizing the *Carnegie Learning Curricula and Cognitive Tutor® Software* Algebra I program received three days of professional development led by a consultant from the developer. Teachers were observed briefly in the classroom and given an opportunity to ask questions of a developer representative early in the implementation period. No ongoing technical assistance was provided.

1. As noted in the protocol, students in grades outside of high school were included in the review if they were included in the study analysis sample along with students in grades 9 through 12.

Appendix A1.2 Study characteristics: Campuzano, Dynarski, Agodini, & Rall, 2009

Characteristic	Description
Study citation	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.
Participants	This national study of software products included an examination of algebra products. Schools were eligible to be in the study if they were in high-poverty areas, had no prior software product use, and had enough teachers in each grade. Teachers in the participating schools were randomly assigned to intervention and comparison groups, and students were allocated to classrooms based on conventional school methods. The fall and spring tests were administered to 276 students, who were age 14 on average and 51% female. Eighteen percent of students were in 8th grade and 82% were in 9th grade.[1]
Setting	During the second year of the study (presented in this report), *Carnegie Learning Curricula and Cognitive Tutor ® Software* was implemented in nine schools in four districts; results from the first year were not disaggregated by intervention. Districts were located in urban and urban fringe areas, averaging 230 schools and 133,000 students. Nine teachers were randomly assigned to use the intervention, and nine were assigned to the comparison condition, with at least a pair of intervention and comparison teachers in each school. Teachers averaged 16 years of experience and 47% had a master's degree.
Intervention	The intervention group consisted of nine teachers from nine schools in four school districts. The intervention was delivered as a full curriculum that included proportional reasoning; solving linear equations and inequalities; solving systems of linear equations; analyzing data; and using polynomial functions, powers, and exponents.
Comparison	The comparison group consisted of nine other teachers from the same nine schools in the four school districts. The students in these classes received traditional algebra instruction using standard district materials.
Primary outcomes and measurement	The study team administered the Educational Testing Service (ETS) Algebra I End-of-Course Assessment. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	Teachers in the intervention group received four days of initial training in the summer of 2004, conducted by a qualified trainer at a school or district location. They were given information on classroom management and curriculum, along with opportunities to practice using the product. Phone and email support was available.

1. As noted in the protocol, students in grades outside of high school were included in the review if they were included in the study analysis sample along with students in grades 9 through 12.

Appendix A1.3 Study characteristics: Shneyderman, 2001

Characteristic	Description
Study citation	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.
Participants	For each of six schools, two teachers were randomly selected from all teachers participating in the program (excluding those working with classes of predominantly exceptional education students). One class for each teacher was randomly selected, creating an intervention sample of 12 classrooms with 325 students. The comparison sample was composed of 12 classrooms with 452 students, randomly selected from a pool of classrooms not implementing the program in the same six schools. Initial proportions of student recipients of free and reduced-price lunch were identical (54%) for the two groups, and ethnic (30% Black, 56% Hispanic, and 13% White for intervention; 27% Black, 62% Hispanic, and 10% White for comparison) and gender (46% and 48% female for intervention and comparison, respectively) distributions were similar. Most of the students in both groups were in 9th and 10th grades: 79% and 18% for the intervention group and 88% and 11% for the comparison group. The analyses were conducted on 276 intervention and 382 comparison students in 9th and 10th grades.
Setting	Within Miami-Dade County Public Schools, nine senior high schools used the *Carnegie Learning Curricula and Cognitive Tutor® Software* Algebra I program during the 2000–01 school year. Of those, six schools that had a computer lab as of October 2000 were examined in the study.
Intervention	*Carnegie Learning Curricula and Cognitive Tutor® Software* Algebra I program covering a full year Algebra I course.
Comparison	Comparison group students took Algebra I.
Primary outcomes and measurement	Algebra performance was measured using the Florida Comprehensive Assessment Test (FCAT) Norm-Referenced Component and the Educational Testing Service (ETS) Algebra I End-of-Course Assessment. However, based on data received by the WWC in response to a query, the intervention and comparison groups were too dissimilar at baseline on the ETS assessment (0.14), and the analysis did not adjust for the pretest differences, so only the FCAT is included in the findings presented in Appendix A3. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	Nothing specified.

Appendix A1.4 Study characteristics: Smith, 2001

Characteristic	Description
Study citation	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.
Participants	The target population included all students who completed the Introduction to Algebra course during the 1999/2000 school year, and then finished their Algebra I requirement by passing Algebra X during the fall semester of the 2000/01 school year. These two courses are part of the district's core curriculum and cover the standard Algebra I material at a slower pace than the traditional math sequence; students are recommended for this sequence by previous math teachers because they have struggled with lower-level math courses. Thus, the sample population consisted of 445 students (229 intervention and 216 comparison) who followed this course progression in one of the seven schools included in the study. Students were randomly assigned to available classes through a computer-scheduling program. As the sample was limited to only those students who completed the three-semester sequence, the randomization process was compromised; therefore, the study is treated as a QED. It does demonstrate equivalence on a pretest and makes the necessary statistical adjustments, so it meets WWC evidence standards with reservations.
Setting	The study involved high schools in Virginia Beach City Public Schools, a large, urban, K–12 school district in Virginia. Of the 10 high schools, one opted not to participate in the program, and two did not keep students in the intervention program together for all three semesters of the study; therefore, seven schools were used for the analysis. The student population was 33.5% minority, including 25% Black.
Intervention	Each high school secured a volunteer mathematics teacher who was willing to implement the intervention rather than the traditional curriculum. Each teacher agreed to spend 40% of class time on the computer and 60% of class time receiving instruction outside the computer lab. The author uses the term *Carnegie Algebra Tutor Software* throughout the report.
Comparison	Comparison classes used traditional instruction based on the city curriculum and textbook, without use of computers or the tutoring software.
Primary outcomes and measurement	At the conclusion of Algebra X, students took the Virginia Standards of Learning (SOL) Assessment for Algebra I. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	Each teacher participated in a three-day training program on how to implement the intervention. Two-thirds of the intervention group teachers were replaced in the second year, and the new teachers did not receive training.

Appendix A2 Outcome measures for the mathematics achievement domain

Outcome measure	Description
Educational Testing Service (ETS) Algebra I End-of-Course Assessment	This 50-question multiple-choice test is based on the Algebra I standards of the National Council of Teachers of Mathematics (as cited in Campuzano et al., 2009).
Florida Comprehensive Assessment Test (FCAT) Norm-Referenced Component	This 48-question multiple-choice test has questions ranging from problem solving to pre-calculus (as cited in Shneyderman, 2001).
Northwest Evaluation Association (NWEA) Algebra End-of-Course Achievement Level Test/Measures of Academic Progress	The two adaptive tests are 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. These results are combined by the authors (as cited in Cabalo, Jaciw, & Vu, 2007).
Virginia Standards of Learning (SOL) Algebra Assessment	This high-stakes assessment, which students need to pass to graduate from high school, consists of 50 questions that contribute to the student's score: 12 on expressions and operations, 12 on relations and functions, 18 on equations and inequalities, and 8 on statistics (as cited in Smith, 2001).

Appendix A3 Summary of study findings included in the rating for the mathematics achievement domain[1]

			Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation)[2]					
Outcome measure	Study sample	Sample size (schools/ students)	*CLC & CT®S* group[3]	Comparison group	Mean difference[4] (*CLC & CT®S* – comparison)	Effect size[5]	Statistical significance[6] (at α = 0.05)	Improvement index[7]
			Cabalo, Jaciw, & Vu, 2007[8]					
NWEA	Grade 8+	6/344	243.4 (7.67)	244.7 (7.47)	–1.34	–0.18	ns	–7
Average for mathematics achievement (Cabalo, Jaciw, & Vu, 2007)[9]						**–0.18**	**ns**	**–7**
			Campuzano et al., 2009[8]					
ETS	Grade 8+	9/276	29.78 (11.04)	31.88 (14.52)	–2.10	–0.16	ns	–6
Average for mathematics achievement (Campuzano et al., 2009)[8]						**–0.16**	**ns**	**–6**
			Shneyderman, 2001[8]					
FCAT	Grades 9 & 10	6/658	683.7 (29.8)	682.5 (27.8)	1.19	0.04	ns	+2
Average for mathematics achievement (Shneyderman, 2001)[8]						**0.04**	**ns**	**+2**
			Smith, 2001[8]					
SOL	Grade 9+	6/445	397.9 (32.9)	400.0 (29.1)	–2.10	–0.07	ns	–3
Average for mathematics achievement (Smith, 2001)[9]						**–0.07**	**ns**	**–3**
Domain average for mathematics achievement across all studies[9]						**–0.09**	**na**	**–4**

ns = not statistically significant
na = not applicable
CLC & CT®S = Carnegie Learning Curricula and Cognitive Tutor® Software
NWEA = Northwest Evaluation Association Algebra End-of-Course Achievement Level Test/Measures of Academic Progress
ETS = Educational Testing Service Algebra I End-of-Course Assessment
FCAT = Florida Comprehensive Assessment Test Norm-Referenced Component
SOL = Virginia Standards of Learning Algebra Assessment

(continued)

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the mathematics achievement domain. Subscale findings from the same studies are not included in these ratings but are reported in Appendix A4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For Cabalo, Jaciw, and Vu (2007) and Campuzano et al. (2009), the intervention group value is the comparison score plus the program coefficient from the hierarchial linear modeling (HLM) analysis. For Campuzano et al. (2009), the standard deviations were obtained from the study authors. For Shneyderman (2001), means and standard deviations for both the intervention and comparison groups were computed using data on 9th- and 10th-grade samples obtained from the study author.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. For the *Carnegie Learning Curricula and Cognitive Tutor® Software* studies summarized here, no corrections for clustering or multiple comparisons were needed.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4 Summary of subscale findings for the mathematics achievement domain[1]

			Authors' findings from the study: Mean outcome (standard deviation)[2]		WWC calculations			
Outcome measure	Study sample	Sample size (schools/ students)	*CLC & CT®S* group[3]	Comparison group	Mean difference[4] (*CLC & CT®S* – comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Cabalo, Jaciw, & Vu, 2007[8]								
NWEA—Quadratic Equations	Grade 8+	6/333	238.96 (11.24)	242.40 (9.98)	–3.44	–0.32	Statistically significant	–13
NWEA—Algebraic Operations	Grade 8+	6/345	241.03 (9.99)	243.50 (10.18)	–2.47	–0.24	ns	–10
NWEA—Linear Equations	Grade 8+	6/335	244.81 (9.57)	245.24 (7.94)	–0.43	–0.04	ns	–2
NWEA—Problem Solving	Grade 8+	6/338	246.67 (11.90)	246.38 (10.69)	0.29	0.03	ns	+1

ns = not statistically significant

CLC & CT®S = Carnegie Learning Curricula and Cognitive Tutor® Software

NWEA = Northwest Evaluation Association Algebra End-of-Course Achievement Level Test/Measures of Academic Progress

1. This appendix presents subscale findings for measures that fall in the mathematics achievement domain. Total scale scores were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For Cabalo, Jaciw, and Vu (2007), the intervention group value is the comparison score plus the program coefficient from the HLM analysis.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Cabal, Jaciw, and Vu (2007), no corrections for clustering or multiple comparisons were needed.

Appendix A5 *Carnegie Learning Curricula and Cognitive Tutor® Software* rating for the mathematics achievement domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of mathematics achievement, the WWC rated *Carnegie Learning Curricula and Cognitive Tutor® Software* as having no discernible effects for high school students.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. None of the four studies showed a statistically significant or substantively important effect.

Other ratings considered

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

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

  Not met. No study showed a statistically significant positive effect.

AND

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

  Met. No studies showed a statistically significant or substantively important negative effect.

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

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

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

AND

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

  Not met. All four studies showed an indeterminate effect.

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

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

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

OR

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

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

(continued)

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

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important *positive* effect.

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

OR

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

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

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

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

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

AND

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

  Met. No studies showed a statistically significant or substantively important positive effect.

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

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size: Schools	Sample size: Students	Extent of evidence[1]
Mathematics achievement	4	27	1,723	Medium to large

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