Appendix

Appendix A1.1 Study characteristics: Morgan & Ritter, 2002 (randomized controlled trial)

Characteristic	Description
Study citation	Morgan, P., & Ritter, S. (2002). *An experimental study of the effects of Cognitive Tutor Algebra I on student knowledge and attitude.* Retrieved November 22, 2006, from http://www.carnegielearning.com/research/research_reports/morgan_ritter_2002.pdf.
Participants[1]	Three hundred and sixty-nine ninth-grade students participated in the part of the study that used random assignment at the section level and included teachers who taught both conditions.[2] The study authors eliminated from the data set students who transferred from one classroom to another or who transferred in or out of the school (or district) during the course of the school year, resulting in 342 students in 19 sections in the end-of-semester grades analysis and 255 students in 16 sections in the ETS total scores analysis. About two-thirds of the students in the study were Caucasian, with the other third Asian, African-American, Hispanic, Native American, and other. None of the students were in special education.
Setting	Three suburban junior high schools in the Moore Independent School District in Oklahoma.
Intervention	The intervention group used *Cognitive Tutor® Algebra I*. Students in this group spent three class periods each week in group activities and classroom discussions using the *Cognitive Tutor® Algebra I* text and two class periods working on problem-solving skills with the *Cognitive Tutor® Algebra I* software. The same teacher taught the intervention and comparison classroom in each of the three schools. Intervention group teachers taught *Cognitive Tutor®* for the first time.
Comparison	The comparison group used Heath Algebra I, published by McDougal Littell. The study does not provide further information on this curriculum. The same teachers taught the intervention and comparison classroom in each of the three schools. Comparison group teachers had several years of experience with the comparison curriculum.
Primary outcomes and measurement	This study used the Algebra End-of-Course Assessment, developed by ETS, consisting of 25 multiple-choice and 15 constructed-response questions. The other two outcomes were first semester grades and second semester grades. (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	All intervention group teachers implemented the curriculum for the first time. According to the study author, during the summer prior to the start of the program the teachers received the standard four-day course required for all teachers beginning the program.

1. The WWC requested and received from the second study author a description of the random assignment process and the results of an HLM analysis for students in classrooms that were randomly assigned within teachers. According to the study author, because of the availability of computers in the schools the study was limited to two teachers per school, and those teachers' classes were randomly assigned to intervention or comparison. Therefore, this review focused on six teachers in three schools.
2. Students from two additonal schools participated in the study. However, because random assignment within teachers did not occur in these schools, these data were not reviewed for rating purposes.

Appendix A1.2 Study characteristics: Shneyderman, 2001 (quasi-experimental design)

Characteristic	Description
Study citation	Shneyderman, A. (2001, September). *Evaluation of the Cognitive Tutor Algebra 1 Program.* Unpublished manuscript. (Miami-Dade County Public Schools Office of Evaluation and Research, 1500 Biscayne Boulevard, Miami, FL 33132.)
Participants[1]	The quasi-experimental study included 439 ninth-grade students (191 in the intervention and 248 in the comparison group) who were enrolled in algebra I during the 2000–01 school year.[1] In each of the participating schools, two teachers of predominantly regular education students were randomly selected. One class per teacher was included in the intervention group. The comparison group comprised an equal number of students within the same schools. About 47% of the students were female, and 54% of the students participated in the free or reduced-price lunch program. About 9% of the students in the intervention group and 20% of the students in the comparison group were English language learners.
Setting	The participating students were from six Miami-Dade County public senior high schools that used *Cognitive Tutor® Algebra I* during the 2000–01 school year. These schools were selected from the nine schools in the district that implemented the curriculum based on the availability of computer labs for student use at the time of the study, because the curriculum involves extensive computer use.
Intervention	Students spent about 40% of their mathematics instructional time using the *Cognitive Tutor® Algebra I* software in their algebra classes during the 2000–01 school year. Students spent the other 60% of their mathematics instructional time working in small groups in the classroom setting using the *Cognitive Tutor® Algebra I* textbook.
Comparison	Comparison students took algebra I but did not use the *Cognitive Tutor® Algebra I* software and textbook. The author does not report what curricula the comparison students used.
Primary outcomes and measurement	The primary outcome measures were the norm referenced component of the Florida Comprehensive Assessment Test (FCAT-NRT) and the ETS Algebra End-of-Course Assessment (EOCT).[2] (See Appendix A2 for detailed descriptions of outcome measures.)
Teacher training	The study did not provide information on teacher training.

1. The WWC requested and received from the study author sample size, means, and standard deviations for a subsample of ninth-grade students who received both the pretests and the posttests. Results for ninth-grade students with no pretest scores were not included in this review because this subsample did not meet the WWC evidence standards for quasi-experimental designs. Results for other students included in this study were not reviewed because they attended higher grade levels (see the Middle School Math Protocol for further detail on inclusion criteria).
2. The study also reported on end-of-year algebra I passing rates and final grades, but these analyses did not take into account baseline differences and so were not reviewed by the WWC.

Appendix A2 Study characteristics: Shneyderman, 2001 (quasi-experimental design)

Outcome measure	Description
ETS Algebra End-of-Course Assessment (EOCT)	The ETS Algebra End-of-Course Assessment (EOCT) included 25 multiple-choice and 15 constructed-response (or performance) items. Each of the performance items contained four parts that progressively increased in difficulty level (as cited in Morgan and Ritter, 2002; Shneyderman, 2001).
Florida Comprehensive Assessment Test—Norm Reference Test (FCAT-NRT)	The mathematics subtest of the state standardized test, the FCAT-NRT consisted of 48 multiple-choice questions in 10 areas, ranging from problem solving to precalculus. Test results were reported as scale scores, mean number of multiple-choice items correct, and mean algebra multiple-choice items correct (as cited in Shneyderman, 2001).
Math achievement grades	End-of-semester and end-of-year grades in mathematics (as cited in Morgan and Ritter, 2002).

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

Outcome measure	Study sample	Sample size (classrooms/ students)[3]	Author's findings from the study: Mean outcome (standard deviation[2]) — *Cognitive Tutor®* group	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Cognitive Tutor®* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Morgan & Ritter, 2002 (randomized controlled trial)[8]								
ETS Algebra End-of-Course Assessment	Grade 9	16/255	16.92 (5.82)	15.28 (5.33)	2.13	0.38	ns	+15
Math achievement grades (end of first semester)	Grade 9	19/342	3.06 (0.99)	2.76 (1.16)	0.45	0.42	Statistically significant	+16
Math achievement grades (end of second semester)	Grade 9	19/342	2.74 (1.11)	2.38 (1.29)	0.44	0.37	Statistically significant	+14
Average[9] for math achievement (Morgan & Ritter, 2002)						0.39	ns	+15
Shneyderman, 2001 (quasi-experimental design)[8]								
ETS Algebra End-of-Course Assessment	Grade 9	24/360	10.27 (5.14)	9.90 (4.79)	0.37	0.07	ns	+3
FCAT-NRT	Grade 9	24/439	681.85 (29.71)	682.39 (28.41)	–0.54	–0.02	ns	–1
Average[9] for math achievement (Shneyderman, 2001)						0.03	ns	+1
Domain average[9] for math achievement across all studies						0.21	na	+8

ns = not statistically significant
na = not applicable

1. This appendix reports finding considered for the effectiveness rating and the average improvement indices.
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. For Morgan and Ritter (2002) and Shneyderman (2001), means and standard deviations were received from the study authors; the WWC further estimated the combined standard deviation across all teachers, taking into account the number of students per teacher. In Morgan and Ritter (2002), the intervention group mean is the comparison group mean plus the mean difference.
3. Sample size information presented here was requested by the WWC and received from the study authors.
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 Technical Details of WWC-Conducted Computations.
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.

(continued)

Appendix A3 Summary of study findings included in the rating for the math achievement domain *(continued)*

8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Shneyderman (2001), corrections for clustering and multiple comparisons were needed. In the case of Morgan and Ritter (2002), corrections for multiple comparisons were needed; corrections for clustering in this study were not needed because the study showed statistically significant effects for two outcomes based on analysis that takes clustering into account. The significance levels may differ from those reported in the original studies.
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 size.

Appendix A4 *Cognitive Tutor® Algebra I* rating for the math achievement domain

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

For the outcome domain of math achievement, the WWC rated *Cognitive Tutor® Algebra I* as having potentially positive effects. It did not meet the criteria for positive effects because only one study showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Cognitive Tutor® Algebra I* was assigned the highest applicable rating.

Rating received

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

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

  Met. One study of *Cognitive Tutor® Algebra I* showed statistically significant positive effects.

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

  Met. One study of *Cognitive Tutor® Algebra I* showed an indeterminate effect and no studies showed statistically significant or substantively important negative effects.

Other ratings considered

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

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

  Not met. One study of *Cognitive Tutor® Algebra I* showed statistically significant positive effects. The second study of *Cognitive Tutor® Algebra I* showed indeterminate effects.

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

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

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Math achievement	2	9	781	Moderate to large

1. A rating of "moderate 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."