Appendix

Appendix A1 Study Characteristics: Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007 (randomized controlled trial)

Characteristic	Description
Study citation	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 source:* 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	Participants included 426 ninth-grade students (206 treatment, 220 control) who were assigned to one of six algebra teachers in three study schools.[1] Algebra course sections for each teacher were randomly assigned to a curriculum. The study authors eliminated from the analysis 83 students who transferred within the district to a different section of the course, did not enroll in the district for the second semester, did not receive a grade, or whose records indicated a conflict between the curriculum and class assignment.[2] In order to reduce the cost of the Algebra I assessment, only one control class was randomly selected for each teacher involved in the study. The algebra assessment analysis sample included 255 students (153 intervention, 102 control) from 16 classrooms (10 intervention, 6 control). The analysis sample for the grades analyses included 343 students (173 intervention, 170 control) in 19 sections (10 intervention, 9 control); however, grades are a subjective measure and were not included in the effectiveness rating.
Setting	Participating students were from three junior high schools in the Moore Independent School District in Oklahoma. Moore is a suburban school district located near Oklahoma City.
Intervention	Students spent three class periods per 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 intervention occurred during the 2000–01 school year, the first year of implementation of *Cognitive Tutor® Algebra I* for the six study teachers.
Comparison	Students in the control group were taught using Heath Algebra I, a traditional textbook published by McDougal–Littell. Study authors do not provide further information on this curriculum. The six study teachers taught both intervention and control classrooms in each of the three schools. At the start of the study, teachers had several years of experience teaching Heath Algebra I.
Primary outcomes and measurement	The study used the Algebra End-of-Course Assessment, developed by the Education Testing Service (ETS). The other two outcomes, which were not taken into account in the effectiveness rating, were first semester grades and second semester (final) grades. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	All teachers implemented *Cognitive Tutor® Algebra I* for the first time. During the summer prior to the start of the intervention, teachers attended a four-day training course to familiarize themselves with the *Cognitive Tutor®* software and to learn teaching techniques.

1. The study authors excluded from the analysis two schools that did not randomly assign classrooms to a curriculum. One school did not have sufficient computer resources to implement *Cognitive Tutor® Algebra I*. Due to a scheduling error, teachers at the other school taught either *Cognitive Tutor® Algebra I* or the traditional curriculum but not both. Only the three schools that implemented the within-teacher random assignment design were analyzed by the study authors and included in this report.
2. Eleven students whose records indicated a conflict between the curriculum and class assignment were excluded by the study authors due to uncertainty about their classroom experience. The school registrar reported that these students were assigned to the control group that received the traditional curriculum but were actually enrolled in a *Cognitive Tutor® Algebra I* classroom.

Appendix A2 Outcome measures for the math achievement domain

Outcome measure	Description
ETS Algebra End-of-Course Assessment	The ETS Algebra End-of-Course Assessment included 25 multiple-choice and 15 constructed-response items, with each type of question accounting for 50% of the student's score. The questions were designed to assess students' understanding of algebraic concepts, processes, and skills (as cited in Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007).
Math achievement grades	First semester and second semester (final) grades were included as an additional measure of performance. Grades are a subjective measure and were not considered in the effectiveness rating; rather, these outcomes are presented as additional findings in Appendix A4.

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

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Cognitive Tutor®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group[4]	WWC calculations: Mean difference[5] (*Cognitive Tutor®* – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007 (randomized controlled trial)[9]								
ETS Algebra End-of-Course Assessment	Grade 9	16/255	17.41 (5.82)	15.28 (5.33)	2.13	0.38	ns	+15
Domain average for math achievement (Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007)[10]						**0.38**	**ns**	**+15**

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math achievement domain. Additional findings from the Ritter, Kulikowich, Lei, McGuire, & Morgan (2007) study 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. The *Cognitive Tutor®* value is the unadjusted control group mean plus the program coefficient from the hierarchical linear modeling (HLM) analysis. The standard deviation was obtained from the study authors.
4. The control group mean is unadjusted. The mean and standard deviation were obtained from the study authors.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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.
9. 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 an explanation about the clustering correction, see the WWC Tutorial on Mismatch. 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 Ritter, Kulikowich, Lei, McGuire, & Morgan (2007), no corrections for multiple comparisons or clustering were needed because only one outcome was considered for the effectiveness rating, and the authors accounted for clustering in their HLM analysis.
10. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4 Summary of additional findings for the math achievement domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Cognitive Tutor®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group[4]	WWC calculations: Mean difference[5] (*Cognitive Tutor®* – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007 (randomized controlled trial)[9]								
First semester grades	Grade 9	19/343	3.22 (1.00)	2.77 (1.16)	0.45	0.42	Statistically significant	+16
Second semester (final) grades	Grade 9	19/343	2.82 (1.12)	2.39 (1.29)	0.43	0.38	Statistically significant	+14

1. This appendix presents additional findings for measures that fall in the math achievement domain. ETS scale scores were used for rating purposes and are presented in Appendix A3. The protocol for the middle school math area states that subjective measures, such as student grades assigned by teachers, do not qualify as relevant outcome measures.
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. The *Cognitive Tutor®* values are the unadjusted control scores group means plus the program coefficients from the hierarchical linear modeling (HLM) analysis. The standard deviations were obtained from the study authors.
4. The control group means are unadjusted. The means and standard deviations were obtained from the study authors.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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.
9. 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 an explanation about the clustering correction, see the WWC Tutorial on Mismatch. 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 Ritter, Kulikowich, Lei, McGuire, & Morgan (2007), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study; no correction for clustering was needed because the authors accounted for clustering in their HLM analysis.

Appendix A5 *Cognitive Tutor® Algebra I* rating for the math 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 math achievement, the WWC rated *Cognitive Tutor® Algebra I* as potentially positive. The remaining ratings (mixed, no discernible effects, potentially negative, and negative) were not considered, as *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 a 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.

  Met. No studies of *Cognitive Tutor® Algebra I* showed a statistically significant or substantively important negative effect, and no studies showed indeterminate 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. No studies of *Cognitive Tutor® Algebra I* showed statistically significant positive effects.

AND

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

  Met. No studies of *Cognitive Tutor® Algebra I* 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. 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		Extent of evidence[1]
		Schools	Students	
Math achievement	1	3	255	Small

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.