A Study of the Effect of *Cognitive Tutor* on Long-Term Course Selections and Performance in the Maui School District

December 2008

Overview

The Maui Hawaii Educational Consortium sought evidence for the effectiveness of the *Cognitive Tutor® (CT)* program, published by Carnegie Learning, to help inform adoption decisions. Funded through the Math Science Partnership (MSP), Empirical Education was contracted during the 2007-2008 school year to conduct a follow-on study to experiments conducted previously by Empirical Education in the Maui School District. These randomized control trials were conducted during the 2005-2006 and 2006-2007 school years. In the first experiment, we studied the effectiveness of *CT's* program in Algebra I, and in the second experiment, we studied the effectiveness of Carnegie Learning's *Bridge to Algebra* program, a pre-algebra curriculum (Cabalo, Jaciw, & Vu, 2007 and Cabalo, Ma, & Jaciw, 2007). The question being addressed specifically by the research in both years of the previous studies was whether students in classes that use *CT* materials achieve higher standardized mathematics assessment scores, as measured by the Northwest Evaluation Association (NWEA) math test, than they would if they had been in a control classroom using the algebra or pre-algebra materials that were already in place. In all three years, our goal was to provide the Maui School District with useful evidence for determining the impact of *CT* within the local setting.

This report describes follow-on research to the algebra and pre-algebra experiments. This study focused on student level math course progression and grade outcomes in the school year following their involvement in the algebra or pre-algebra studies. The main research questions are:

1. Did a higher percentage of *CT* students than control students take a math course in the school year following their involvement in the previous studies?
2. Did students who had exposure to *CT* choose courses that were of a higher course progression level than students who did not have exposure to *CT*?
3. Did students who had exposure to *CT* obtain higher grades in their subsequent math course than students who did not have exposure to *CT*?
4. Did the impact of *CT* on course selection and performance vary with ethnicity?

This research was conducted as a MeasureResults™ service by Empirical Education Inc.

www.empiricaleducation.com

Methods

We examined the subsequent course-taking of students involved in the algebra and pre-algebra experiments. Researchers coded each course on a scale from 1 to 5, as shown in Table 1 and described in more detail in Appendix A.

Table 1. Course Level Coding

Course Level	General Category
1	Pre-algebra, Problem-solving, Business math, Applications
2	Algebra I
3	Geometry
4	Algebra II
5	AP courses

Due to inconsistencies in the data for middle-school students, we removed all 7th and 8th grade data from the analysis, as explained in Appendix B. The remaining high school students are drawn from the same randomized classes and this study takes advantage of that initial randomization. The sample used in this study is shown in Table 2.

Table 2. High School Students in Control and *CT* Groups

		Total number randomized	Students who enrolled in subsequent math course
Algebra I Experiment	**Control**	213	154
	CT	225	170
Pre-Algebra Experiment	**Control**	206	175
	CT	194	172
Totals		**838**	**671**

Note. Nine students who took a subsequent course did not receive a grade.

To answer research question 1, we looked at the percentage of *CT* and control students who took a math course in the school year following their involvement in the previous studies. We examined these results separately for students in the algebra and pre-algebra experiments. In answering all the other questions, we combined the results for the algebra and pre-algebra groups.

Next, we examined the effect of *CT* on students' course level selections (question 2). The students' highest ranking course in the year following their involvement in the previous studies serves as our outcome measure for question 2. This allowed us to test whether or not the *CT* group, following their experience with the program, took courses that are further along in the course progression than the students who did not use *CT.*

We then measured the impact of *CT* on the grades students received in their highest ranking subsequent math course (question 3). Appendix C explains our conversion of course grades into integer values. We recognized that if exposure to *CT* affects course-taking, then course grades may reflect not only student proficiency but also selection of more difficult or easier courses. To make a fair comparison between students who did and did not receive *CT* in terms of their subsequent performance, we adjusted for the differences in course choices between conditions (*CT* or comparison group). That is, we measured the average difference in performance while holding constant the level of course progression (see Appendix D).

Finally, we examined if the effect of *CT* on course selection and course grade varied among different ethnicity groups (question 4).

Levels of confidence in our results: *We report results based on statistical tests that give us a measure of confidence expressed as a probability. We often see a difference between, for example, the heights of two bars in a graph, but have no confidence that it is any more than a chance difference. We use* p *values where a low value indicates a low probability that we would detect a difference like the one found in the experiment if no difference actually existed. A* p *value less than .05 gives us strong confidence in the result (a level conventionally called statistically significant), while a* p *value greater than .20 gives no confidence. Between the two we may have some or limited confidence. We also provide the 80% confidence intervals for some graphs, denoting our interpretation that an 80% probability exists that the tops of the bars simultaneously fall somewhere within their respective interval. Where the intervals overlap, the* p *value is greater than .20.*

Results

Question 1: Did a higher percentage of *CT* students than control students take a math course in the school year following their involvement in the previous studies?

No, there is no discernible difference between the two groups.

Figure 1 shows the percentage of students from the algebra experiment who took a subsequent math course. Of these students, 72.3% (154 out of 213) of the control students took a subsequent math course compared to 75.6% (170 out of 225) of the *CT* students. However, this small difference could easily be due to chance (p=.45).

Figure 2 shows the percentage of students from the pre-algebra experiment who took a subsequent math course. Of these students, 85.0% (175 out of 206) of control students took a subsequent math course compared to 88.7% (172 out of 194) of *CT* students. However, this small difference could easily be due to chance (p=.30).

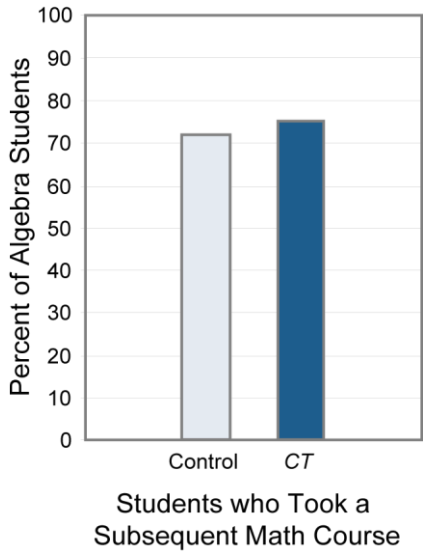

Figure 1. Percent of Students From the Algebra Experiment Who Took A Subsequent Math Course

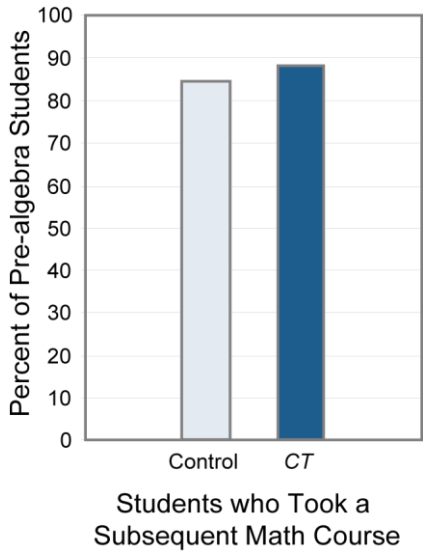

Figure 2. Percent of Students From the Pre-Algebra Experiment Who Took A Subsequent Math Course

Question 2: Did students who had exposure to *CT* choose courses that were of a higher course progression level than students who did not have exposure to *CT*?

No, there was no difference between students who had exposure to CT and those who did not in their subsequent course progression level.

We obtained a *p* value of .60 for the statistical test of the mean difference between the two groups, indicating that the difference in course progression levels is small enough to be easily explained by chance. See Appendix E for greater detail.

The following two graphs show the math course level progressions of the students in the year after they participated in the *CT* experiments. These are provided for descriptive purposes to illustrate the general course progressions for the students starting in Algebra and those starting in Pre-algebra. The graphs show the *CT* and control group students combined and illustrate that most students who are recorded as taking a subsequent math course, progressed from their initial course to a higher-level course. Students in the algebra experiment started in a level 2 algebra course. Students in the pre-algebra experiment started in a level 1 pre-algebra course.

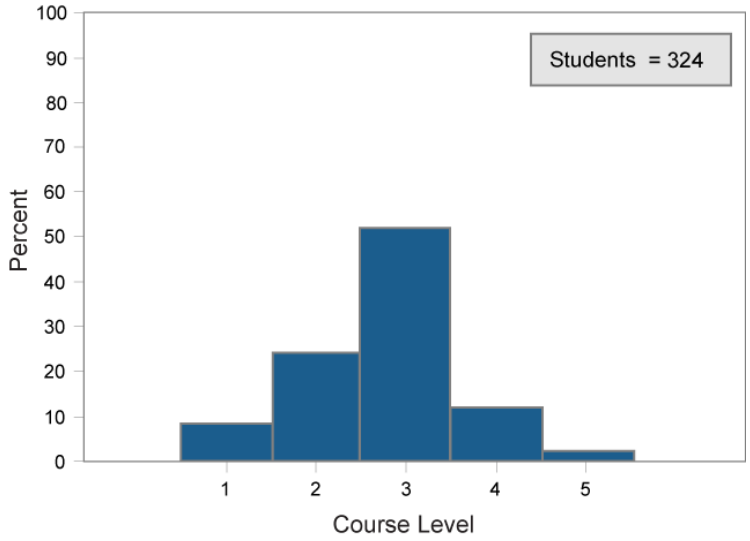

Figure 3 shows the results for students in the algebra experiment. The bars show the percentage of students in each of the course rankings (1-5) in the school year following their involvement in the study.

Figure 3. Course Levels for Students From the Algebra Experiment

Figure 4 shows the percentage of students in each of the course rankings (1-5) in the school year following their involvement in the pre-algebra study.

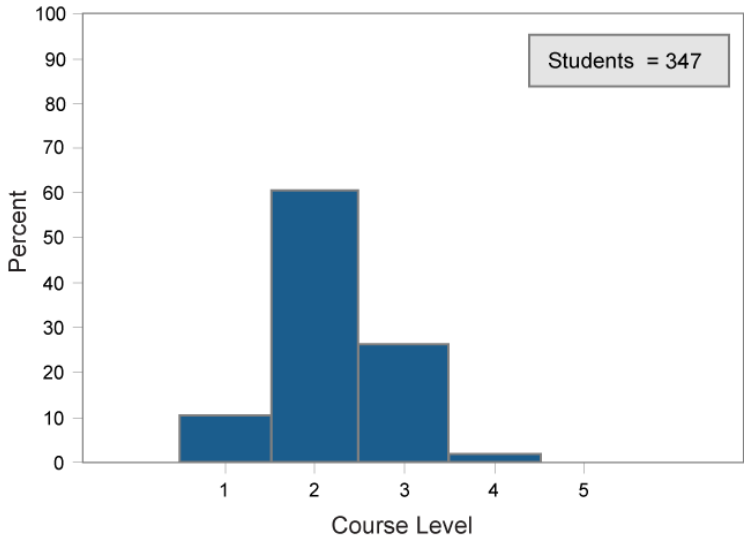

Figure 4. Course Levels for Students From the Pre-Algebra Experiment

Question 3: Did students who had exposure to *CT* obtain higher grades in their subsequent math course than students who did not have exposure to *CT*?

No, there was no discernible difference between the two groups.

We did not find a difference in average course grade between students who received *CT* and those who did not. The *p* value of .65 gives us no confidence that the effect of *CT* on course grade, controlling for the difference between conditions in course progression level, is different from zero. See Appendix F for the table of results.

Question 4: Did the impact of *CT* on course selection and grade performance vary with ethnicity?

No, we did not find that the impact of CT *on course selection or grade performance varies with ethnicity.*

Course Selection

Our statistical test revealed that the effect of *CT* on course progression does not vary with ethnicity. The *p* value of .89 indicates that any observed variation among ethnicities in this effect is easily due to chance. Appendix G contains further details about the results.

Grade Performance

Our statistical test revealed that the effect of *CT* on course grade (after controlling for the imbalance between conditions on course progression) does not vary with ethnicity. The *p* value of .84 shows that any observed variation among ethnicities in this effect is easily due to chance. Appendix H contains further details about the results.

Cautions for Interpreting These Results:

- We undertook this study knowing that, although there would be limitations, through this initial exploration of patterns of course-taking we can better understand future opportunities for research with larger samples.
- Although the number of students was large the number of randomization units (classes) is relatively small. Since the statistical calculations are largely based on the number of classes and secondarily on the number of students, the experiments did not have the statistical power to detect small effects.

References

Cabalo, J.V., Ma, B., & Jaciw, A. (2007, October). *Comparative effectiveness of Carnegie Learning's Cognitive Tutor Bridge to Algebra Curriculum*. (Empirical Education Rep. No. EEI_EdCT2-06-FR-Y2-0.2). Palo Alto, CA: Empirical Education Inc.

Cabalo, J.V., & Vu, M. (2007, May). *Comparative effectiveness of Carnegie Learning's Cognitive Tutor Algebra I curriculum: A report of a randomized experiment in Maui School District*. (Empirical Education Rep. No. EEI_EdCT-05-FR-Y1-O.2). Palo Alto, CA: Empirical Education Inc.

Appendix

A. Course Ranking Development

To establish a scale with course progression levels as the outcome, we worked with course rankings provided by members of the Maui Educational Consortium. In order to maintain a continuous ranking, we only included courses in this ranking that appeared in the dataset (for example, no students had a precalculus or calculus course as their highest ranking course, so we did not include those courses in the ranking). For cases in which students had multiple math courses with different rankings, we selected the highest ranking course and corresponding grade as the outcomes for this study. Table A-1 summarizes the categorizations that we used to establish the course ranking scale. Note that Business Math and Applications are not considered preparatory for more advanced math although some number of students took these courses after having taken Algebra 1.

Table A- 1.

Course Level	General Category
1	Pre-algebra, Problem-solving, Business math, Applications
2	Algebra I
3	Geometry
4	Algebra II
5	AP courses

B. Data Sample

Due to inconsistencies in the data, we removed all data for students who were in 7th and 8th grade in the algebra and pre-algebra experiments. Upon examining the data, we found that 7th graders did not universally move into an 8th grade math in the subsequent school year. Many moved into pre-Algebra or Algebra and were coded as in 9th or 10th grade. Furthermore, we found that more than half of the students who began in 8th grade in the pre-algebra experiment, were retained in 8th grade math and coded as 8th graders in the subsequent school years. Because we were unclear about these patterns, we decided to run the analysis and report only on the high school progression.

C. Course Grade Scale

Standard letter grades were used as measures of performance. These were converted to integer values (F=0, D=1, C=2, B=3, A=4). There were no + or – grades in the source data. A grade of “S” or “N” were mapped to 2, 0 respectively (pass/fail). Course grade was also treated as a continuous outcome measure.

D. Relationship between Course Level and Grade

A challenge to estimating the long term impact of *CT* on course grade is that *CT* may influence both course selection and course performance. For example, if students with exposure to *CT* are more likely to take courses higher in the course sequence pattern, then their letter-grade performance may end up being lower than the comparison group’s, not because they are lower in proficiency, but because it may be harder to obtain high scores on the courses that they are selecting as a consequence of having been exposed to *CT*. On the other hand, students who tend to achieve higher grades might also tend to take courses that are farther along in the progression as a result of their high

achievement. For this reason we include the outcome, level of the course progression, as a covariate in the analysis of the relationship between *CT* and course performance.

E. Impact of *CT* on Course Level (for statistical reviewer)

Table A- 2.

Solution for Fixed Effects					
Effect	**Estimate**	**Standard error**	**DF**	***t* value**	***p* value**
Course level for controls	2.54	0.13	30	19.04	<.01
Difference (*CT* – control) in course level	-0.10	0.19	30	-0.53	.60
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Class mean achievement	0.26	0.07		3.54	<.01
Within-class variation	0.43	0.02		17.74	<.01

Note. Adding the pretest as a covariate into the analysis did not change the result (*p* value for the treatment effect is .54)

F. Impact of *CT* on Course Grade, Controlling for Course Level (for statistical reviewer)

Table A- 3.

Effect	**Estimate**	**Standard error**	**DF**	***t* value**	***p* value**
Achievement for a control with an average pretest	1.08	0.20	30	5.26	<.01
Effect of *CT*	-0. 07	0.15	30	-0.45	.65
Effect of course level	0.24	0.07	629	3.52	<.01
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Class mean achievement	0.09	0.05		1.90	.03
Within-class variation	1.69	0.10		17.70	<.01

Note. Adding the pretest as a covariate into the analysis did not change the result (*p* value for the treatment effect is .60)

G. Moderating Effect of Ethnic Background on Course Level

Table A- 4.

Fixed effects	Estimate	Standard error	DF	*t* value	*p* value
Outcome for a Hawaiian control student with an average pretest	2.55	0.15	30	16.60	<.01
Change in outcome for each unit-increase on the pretest	0.11	0.03	505	6.62	<.01
Effect of *CT* for a Native Hawaiian student	-0.06	0.22	50	-0.29	.77
Difference (Filipino student – Hawaiian student) in control outcome	0.04	0.10	75	0.42	.68
Difference (Other student – Hawaiian student) in control outcome	-0.00	0.11	75	-0.02	.99
Difference (White student – Hawaiian student) in control outcome	0.13	0.16	75	0.83	.41
Difference (Filipino student – Hawaiian student) in the effect of *CT*	-0.07	0.15	505	-0.46	.65
Difference (Other student – Hawaiian student) in the effect of *CT*	-0.04	0.15	505	-0.26	.80
Difference (White student – Hawaiian student) in the effect of *CT*	-0.15	0.20	505	-0.76	.45
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Class mean achievement	0.29	NA		NA	NA
Within-class variation	0.36	NA		NA	NA

Table A- 5. Type-3 Test of Fixed Effects

Effect	Numerator DF	Denominator DF	*f* value	*p* value
Pretest	1	505	13.13	<.01
CT	1	30	0.41	.53
Ethnicity	3	75	0.30	.82
Interaction of *CT* and ethnicity	3	505	0.21	.89

H. Moderating Effect of Ethnic Background on Course Grades

Table A- 6.

Fixed effects	Estimate	Standard error	DF	*t* value	*p* value
Outcome for a Filipino control student with an average pretest	0.90	0.26	30	3.43	<.01
Change in outcome for each unit-increase on the pretest	0.28	0.06	504	4.59	<.01
Effect of course level	0.27	0.08	504	3.45	<.01
Effect of *CT* for a Native Hawaiian student	-0.00	0.22	30	-0.02	.99
Difference (Filipino student – Hawaiian student) in control outcome	0.13	0.20	75	0.67	.51
Difference (Other student – Hawaiian student) in control outcome	0.14	0.22	75	0.64	.53
Difference (White student – Hawaiian student) in control outcome	0.01	0.31	75	0.64	.97
Difference (Filipino student – Hawaiian student) in the effect of *CT*	0.23	0.29	504	0.79	.43
Difference (Other student – Hawaiian student) in the effect of *CT*	0.01	0.30	504	0.03	.98
Difference (White student – Hawaiian student) in the effect of *CT*	0.09	0.40	504	0.23	.82
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Class mean achievement	0.05	NA		NA	NA
Within-class variation	1.61	NA		NA	NA

Table A- 7. Type-3 Test of Fixed Effects

Effect	Numerator DF	Denominator DF	*f* value	*p* value
Course level	1	504	11.90	<.01
pretest	1	504	21.09	<.01
CT	1	30	0.28	.60
ethnicity	3	75	0.22	.88
Interaction of *CT* and ethnicity	3	504	0.27	.84