Empirical Education®

RESEARCH SUMMARY

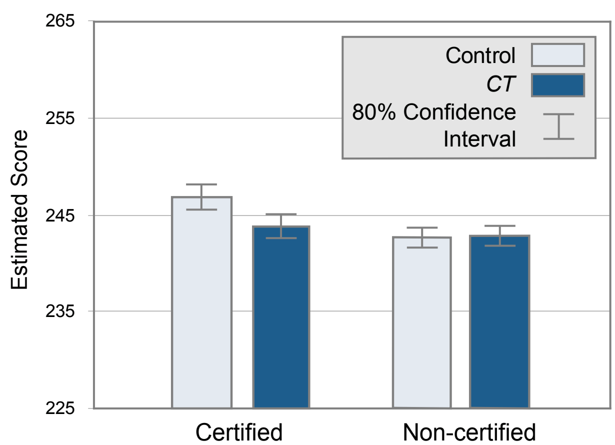
Comparative Effectiveness of Carnegie Learning's *Cognitive Tutor*® Algebra I Curriculum:

A Report of a Randomized Experiment in the Maui School District

Introduction. The Maui Hawaii Educational consortium (the Maui School District and Maui Community College) sought scientifically based evidence for the effectiveness of the *Cognitive Tutor (CT)* Algebra I Curriculum to inform adoption decisions. Decision makers were particularly interested in whether the use of the *CT* program affects achievement of their students and is a useful tool for teachers. The participating teachers used the program in their randomly selected *CT* classes for six months during the 2005-2006 school year while teaching their other classes with their standard materials. The experiment involved nine teachers, 22 classes, and 541 students.

Findings. Our experiment for the most part was unable to detect any impact of *Cognitive Tutor* on student achievement. Of the four sub-strands, the analysis found negative results for two of them: Quadratic Equations and Algebraic Operations. The results must be interpreted in the context of the particular resources for the implementation of the program on Maui, which were not favorable. Nonetheless, the size of the negative impact of the two sub-strands was considerable. The effect sizes were -0.33 and -0.25. These are comparable in magnitude to the positive effects found in other studies of this product.

One explanation for the negative outcome is shown in the analysis of teacher certification. While the certified teachers out performed the uncertified teachers with their control class, they performed the same as the uncertified teachers with their *CT* classes. This is shown in the following figure of the moderating effect of teacher certification on NWEA overall math outcomes.



Perhaps, because both groups were equally unfamiliar with *CT* their results were the same. With the existing program, however, the more experienced teachers were far more familiar with the methods and content of the program than the uncertified teachers (who were equally unfamiliar with both programs). It is possible that over time, with more familiarity with *CT*, both groups may improve. Tracking these teachers into their second year with *CT* may provide some indication as to whether the experienced teachers can overcome the initial deficit.

While teachers gave a generally positive view of *CT*, they reported difficulties with implementation—particularly access to computers. However, our exploratory analysis of student usage data did not support lack of computer access as an explanation for the poor results. Aspects of the implementation other than computer lab time may have influenced the results—the late delivery of the product and lack of familiarity, for example.

We conducted an additional exploratory analysis of the data provided by the *CT* intervention on units completed. This indicated that there is an overall positive relationship between the number of units completed and student outcomes on the NWEA test. This also suggested that the NWEA test was a good measure of what was covered in the *Cognitive Tutor* curriculum. We must caution, however that the nature of the outcome measure (the paper version of the NWEA end-of-course Algebra I test) makes us less confident in our conclusions about lower achieving students. First, it appears that lower achieving students were less likely to take or complete the posttest. Second, the test itself is not sensitive to students scoring at the low end and appears to distort their scores upward due to a floor effect in the paper-version of the NWEA test.

If the implementation and start-up problems were part of the reason for the poor performance, the results suggest that school districts should avoid undertaking this program without adequate resources and preparation time. Our recommendation is to continue the experiment into an additional year, providing adequate resources for implementation, before undertaking wider deployment.

This small study illustrates a general caution in interpreting findings from isolated experiments. Previous research has indicated that the Cognitive Tutor program is effective (Morgan & Ritter, 2002). This experiment demonstrates the importance of conducting multiple replication trials of any application in varying contexts and conditions.

Setting. The district and community college are located in a mixed suburban and rural community. The community is ethnically diverse and the average ethnic breakdown for the participating schools includes approximately 32% Filipino, 28% Part-Hawaiian, 11% White, 8% Japanese, 5% Hawaiian, 3% Hispanic, and 14% Other. The district's existing math program consisted of a variety of Algebra I textbooks from several publishers.

The intervention. *Cognitive Tutor* is a full Algebra I curriculum, combining software-based, individualized computer lessons with collaborative, real-world problem-solving activities. The publisher, Carnegie Learning, describes *CT* as a research-based approach to improving student understanding of mathematical concepts that includes a simple and straightforward design, research-based pedagogy, multiple representations of word problems, just-in-time feedback, a skillometer, and a blended curriculum of 40% computer lab and 60% classroom activities. *CT* also uses verbal, numerical, algebraic, and graphical representations as different modalities to ease problem-solving. A Carnegie Learning consultant led all eight *CT* teachers in three days of training, after which they could use of the materials as best suited the needs of their *CT* classes and students.

Research design. We conducted a randomized experiment which compared outcomes for groups of students taught using the *CT* program and students taught using the regular Algebra I curriculum. We randomly assigned the Algebra I classes of each participating teacher to the *CT* or control group. Randomization of participating teachers' classes was stratified according to class size and achievement level. Using the tests obtained from the Northwest Evaluation Association, we collected pretest measures of general math and posttest measures of algebra achievement. The posttest provided sub-strands for Algebra Operations, Linear Equations, Quadratic Equations, and Problem-Solving. We also gathered demographic information on students and teachers. We collected three types of qualitative data: 1) classroom observations, to document implementation; 2) phone interviews, to gauge teachers' attitudes and opinions about the *CT* program as well as the kinds of challenges and difficulties encountered; and 3) web-based surveys, to learn about factors that may influence the results. In addition, as a feature of the *CT* Teacher's Toolkit, we were able to gather information on student time spent on the software and the number of problems and units each student completed.

Analysis. The basic question for the statistical analysis was whether, following the intervention, students in the *CT* classrooms had higher math scores than those in control classrooms. We used multi-level models that account for the clustering of students in classes, providing a more accurate, and often more conservative, assessment of the confidence we should have in the findings. To increase the precision of our estimate, we included students' pretest scores in the analysis and examined the interaction between this covariate and the experimental condition. Finally, to better understand unexpected results, we use other data such as that from the *CT* Teacher's Toolkit in exploratory analyses.

To read the complete report, contact:

Empirical Education Inc.
www.empiricaleducation.com
425 Sherman Avenue, Suite 210
Palo Alto, CA 94306
(650) 328-1734