Is There a Relationship Between Interacting with a Mathematical Intelligent Tutoring System and Students Performance on Standardized High-Stake Tests?[1]

Jeremiah Sullins, Rob Meister, Scotty D. Craig, William M. Wilson, Anna Bargagliotti, and Xiangen Hu[*]

The University of Memphis

[1] The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090528 to the University of Memphis. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The Authors would like to thank Memphis City Schools and Shelby County Schools. Without their support, we will not be able to obtain data necessary for this report.

[*] Contacting Author: Xiangen Hu, The University of Memphis, Memphis, TN 38152. xhu@memphis.edu

Abstract

This chapter presents findings from two observational studies conducted in Memphis City School District and Shelby County School District exploring the effectiveness of a mathematical intelligent tutoring system called ALEKS (Assessment and LEarning in Knowledge Spaces). Specifically, these studies were designed to assess the relationship between students' interaction with ALEKS and mathematics achievement test scores. The two studies involved retroactively comparing students' TCAP (Tennessee Comprehensive Assessment Program) mathematics scores with their proficiency level scores in ALEKS. The findings reveal that students who show a higher level of proficiency in ALEKS scored significantly higher on the TCAP.

Introduction

Recent research has suggested that standardized high-stakes tests, such as the SAT (Scholastic Aptitude Test) have become increasingly important to policy makers, school districts, and society in general. Scoring well on these tests has the potential to determine access to future educational opportunities that are available to learners after high school. Unfortunately, recent reports have shown that Americans are falling behind their peers in other nations on comparable assessments (Gollub, Bertenthal, Labov, & Curtis, 2002). Additionally, schools in the U.S. must adhere to the demands of the No Child Left Behind Act (NCLB). The policy instates that federal district funding is dependent on student overall performance on standardized tests in mathematics, reading and other content areas. In order to alleviate the problem of U.S. students underachieving on standardized tests, it is necessary for educators to explore areas of pedagogy that have been empirically shown to be effective.

A potential alternative method is one-to-one human tutoring, which has often been put forth as a gold standard in instruction (Bloom, 1984; Cohen, Kulik, & Kulik, 1982). A meta-analysis conducted by Cohen et al. (1984) revealed that non-expert human tutors increased learning gains by .4 standard deviation units (this translates to approximately half a letter grade). Unfortunately, teachers to not have the time and districts to do not have the resources to provide individualized tutoring to every student in a classroom. However, if we could apply technology that has the capability of being effective and personalized, then we could potentially improve teaching by making the transfer of knowledge and the process of student learning and development more efficient.

Technology in the Classroom

One area of research that has shown promise in the facilitation of math achievement is in the field of computer-based instruction and intelligent tutoring systems (ITS). Loosely defined, an ITS is a computer-based learning environment that provides individualized instruction to the learner. There is available research that has shown ALEKS (one example of a mathematical ITS) to be effective in identifying and closing gaps in math knowledge and skills at the college level (Hagerty, et al, 2005), no research at the secondary and middle school levels using ALEKS currently exists. However, there are numerous researchers that have explored the possible influence of technology on achievement scores in varying domains (e.g., reading comprehension and mathematics).

Some researchers postulate that technology provides no advantages over traditional classroom instruction. For example, Angrist and Lavy (2002) compared computer skills training and computer aided instruction in relation to mathematical achievement among fourth graders. They define computer skills training as teaching students how to use computers and computer aided instruction as using computers to teach. Results revealed that computer aided instruction provided no benefit (at least in the short term) on academic achievement in math among fourth grade students.

Additionally, Baker, Gearhart and Herman (1994) evaluated the Apple Classrooms of Tomorrow and examined the impact of interactive technologies across five school sites across the nation (e.g., California, Tennessee, Minnesota, and Ohio). The Apple Classrooms of Tomorrow were designed to encourage instructional innovation, and to emphasize to teachers the potential of computers to support student initiative, long term projects, access to multiple resources, and cooperative learning. Although there were some positive outcomes as a result of

the initiative, it was found that on standardized tests in various contents (e.g., mathematics, reading comprehension, and vocabulary) students performed no better than control groups or national norms.

Wenglinsky (1998) explored the effects of simulation and higher order thinking technologies on mathematics achievement on the National Assessment of Educational Progress. The sample size consisted of 6,227 fourth graders and 7,146 eight graders. While controlling for socioeconomic status, class size, and teacher characteristics, the results revealed that students who used drill and practice technologies performed worse on the National Assessment of Educational Progress than students who did not use drill and practice technology.

However, this view may not be the full story. More recent evidence has been reported in favor of technology. This research found that supplementing instruction with technology (i.e., ITSs) students mathematical achievement can be increased significantly (Woolf, Arroyo, Beal, & Murray, 2006; Melis & Siekmann, 2004). Woolf et al. (2006) have developed a web-based mathematical ITS called Wayang Outpost. Wayang Outpost was designed to prepare students for the mathematics section of the SAT. Results from numerous studies have shown that Wayang Outpost has been beneficial for students in general with high improvements from pretest to post test (Woolf et al., 2006).

A meta-analysis conducted by Kulik (1994) examined over 500 studies exploring the effects of computer-based instruction. Results of the meta-analysis revealed several positive benefits of using computer based instruction. For example, students who used computer-based instruction scored in the 64^{th} percentile on tests of achievement while students not using computer-based instruction scored in the 50^{th} percentile. Additionally, students learned more in less time when using computer-based instruction.

Sivin-Kachala (1998) reviewed 219 studies to explore the effect of technology on learning on achievement across various domains and ages. Students in environments that utilized technology showed increased achievement from preschool through higher education in all major subject areas.

Although there have been a large number of studies exploring the benefits of online learning and computer-based learning environments, there is a general the lack of studies that have focused on K-12 students (barring a few examples). This research aims to fill this gap in the literature by answering the following important question: does a statistically significant and positive correlation exist between middle school students' mastery of topics available within ALEKS as measured by percent of topics mastered and achievement of middle school students' on equivalent items found within the TCAP test as measured by the student's scaled score? That is, do the middle school students who demonstrate mastery of topics in ALEKS that have been associated with Tennessee standards also perform well, i.e., demonstrate mastery, on the corresponding standards assessed with the TCAP?

ALEKS (Assessment Learning in Knowledge Spaces)

ALEKS is a Web-based, artificially intelligent assessment and learning system that uses adaptive questioning to quickly and accurately determine exactly what a student knows and does not know in a course. Subsequently, ALEKS instructs the student on the topics she is most ready to learn. As a student works through a course, ALEKS periodically reassesses the student to ensure that topics learned are also retained. ALEKS courses are very complete in their topic coverage and the system avoids multiple-choice questions. A student who shows a high level of mastery of an ALEKS course has the potential for increase in the actual course being taken.

ALEKS in Memphis Area Schools

ALEKS was introduced to Memphis Area Schools in 2004. The first group of users were college students taking a behavioral statistics course at the University of Memphis. However, ALEKS was implemented in elementary and middle schools for the first time in 2005. Since 2005, there have been a total 11 schools that have employed the use of ALEKS during the academic year. Students from over 170 schools used ALEKS during the Summer Online Learning Experience (SOLE) program. There have been over 5000 students that have benefited from using ALEKS. The current report will only analyze some preliminary findings of the data. We only report student data from schools where an institutional review board (IRB) has approved our data collection protocol.

In this chapter, we focus on two studies that we have conducted in an attempt to assess whether providing individualized instruction to the learner in the form of a mathematical Intelligent Tutoring Systems (ITS) will increase student scores on a standardized high-stake test. The purpose of Study 1 is to determine whether mastery of topics in an online math tutorial system called Assessment and LEarning in Knowledge Spaces (ALEKS) is correlated with the success in the corresponding overall math component of the Tennessee Comprehensive Assessment Program (TCAP). Furthermore it explores the possible relationship between mastery of topics in ALEKS and the Standard Performance Indicator (SPI) level. SPIs were developed by the Tennessee Department of Education to ensure that students in K-12 will learn the reading and language skills needed to succeed in the classroom, workplace with the ultimate goal being lifelong learning. The purpose of Study 2 is to partially replicate and expand on the results of Study 1 with a different sample (i.e., different grade level) in a different district. The two reports

we present here focus on the correlation between ALEKS assessment score and TCAP scared score.

Study 1

Purpose of the Study

Studies have shown that ALEKS is efficient at providing individualized instruction to students at the college level (Haggerty & Smith, 1997). However, a question that warrants further investigation is whether the same effect will transfer to younger students (i.e., middle school students). Additionally, the primary significance of the study concerns the association of ALEKS topics with the Tennessee state standards or SPIs. If significant positive correlations exists between a students' success in mastering ALEKS topics and a students' demonstration of mastery of the corresponding SPIs, then ALEKS may prove to be a potential predictor of future success on the Tennessee state standardized mathematics assessment (e.g., TCAP). If such a relationship exists, ALEKS could provide educators with the means by which to regularly, effectively, and efficiently determine each individual student's strengths, weaknesses, and needs for further study and practice to improve his or her performance on TCAP well in advance of the administration of the TCAP test. Therefore, educators could take immediate corrective actions to adjust lesson plans well in advance of the standardized test in order to better serve and educate students in precisely those areas where additional attention is needed. This technological advantage could save educators time and effort that could then be more effectively redistributed to students according to their individual needs.

Method

Procedures

Students had ALEKS available to them for approximately three fourths of the school year (or about 20 weeks) from the beginning of the 2006-2007 school up until the TCAP was administered. Coinciding with their standardized TCAP assessment, an assessment was administered online via the ALEKS system. Therefore, two mathematics performance measures were collected for each student: 1) achievement as measured by the percent of ALEKS topics mastered in the ALEKS assessment and 2) achievement in the TCAP as measured by the student's mathematics scaled score. Percentage scores at the individual student level in ALEKS and scaled scores in TCAP were compared to determine the degree to which student performance on the two tests is related.

Participants

The study included students enrolled in the middle grades (grades six, seven, and eight) of the Memphis City and Shelby County School Systems. Students participated in their normal curriculum as determined by their respective school and school system. In addition, they were given access to the ALEKS system. One middle school from the Memphis City School system and one school from the Shelby County School system volunteered and agreed to use ALEKS as a significant part of their curriculum. The Memphis City Schools had total of 267 students in 6th, 7th, and 8th grade. One school's mathematics achievement was based on a "report card" from the Tennessee Department of Education in which the school received a mathematics category grade of "D". This grade represents that students were underperforming in mathematics achievement. The Shelby City Schools has total of 1073 students in 3th to 7th grade. Based on

the "report card" from Tennessee Department of Education, the mathematics category grade was an "A" representing that students were over-performing mathematics achievement.

In order to join the ALEKS research project, among other things, participating schools agreed to weight a student's participation in ALEKS at thirty percent of his or her overall course grade. Between these two schools, 218 students participated in the study in the spring of 2007.

Hypotheses

Based on the available research there are two different possible hypotheses. According to an *equivalency prediction* (Angrist, Lavy, 2002; Baker, Gearhart & Herman, 1994; Wenglinsky, 1998) we would expect no significant relationships to be found between students' proficiency in ALEKS and their achievement scores on the TCAP.

According to an *instructional supplement prediction* (Woolf, Arroyo, Beal, & Murray, 2006; Melis & Siekmann, 2004; Kulik, 1994; Sivin-Kachala, 1998) a significant positive relationship between students ALEKS proficiency and their TCAP scores would be expected.

Results

Results of a correlation analysis indicate that a positive and statistically significant correlation ($r = .842$, $p < .01$, $n = 216$) exists between TCAP scaled scores and Assessment performance goal percentage after in ALEKS. These results suggest that as students become more proficient in ALEKS they are more likely to score higher on the math portion of the TCAP. Figure 1 presents a scatter plot relating the two measures.

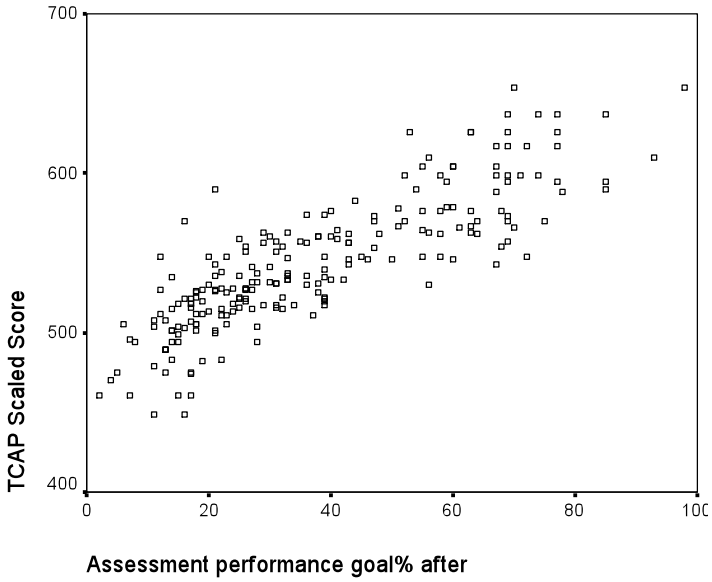

Figure I. Scatter plot showing the relationship between TCAP scores and Assessment Performance.

Also, the results of the correlation analysis indicate that a positive and statistically significant correlation exists between the overall TCAP and ALEKS measures at each school ($r = .716$, $p < .01$, $n = 146$ and $r = .697$, $p < .01$, $n = 72$).

Additionally, correlations were performed on each individual category within ALEKS and SPI. The results of the correlation analysis indicate that a positive and statistically significant correlation ($r = .687$, $p < .01$, $n = 128$) exists between the TCAP Numbers and Operations category measure and ALEKS Number and Operations measure.

A significant positive correlation ($r = .522$, $p < .01$, $n = 128$) was found between the TCAP Algebraic Thinking category measure and ALEKS Algebraic Thinking measure. A significant positive correlation($r = .408$, $p < .01$, $n = 128$) was discovered between the TCAP Graphs and Graphing measure category and ALEKS Graphs and Graphing category. Additionally, a significant positive correlation ($r = .523$, $p < .01$, $n = 128$) was found between the TCAP Data Analysis and Processing category measure and ALEKS Data Analysis

and Processing measure. A significant positive correlation ($r = .657$, $p < .01$, $n = 128$) was also discovered between the TCAP Measurement category measure and ALEKS Measurement measure. Lastly, a significant positive correlation ($r = .597$, $p < .01$, $n = 128$) was detected between the TCAP Geometry category measure and ALEKS Geometry measure. Correlations can be found in Table 1.

Table 1.
Correlations Between ALEKS and Corresponding TCAP section

Grade	Pearson Correlation
Numbers and Operations	$r = .687$, $p < .000$, $n = 128$
Algebraic Thinking	$r = .522$, $p < .000$, $n = 128$
Graphs and Graphing	$r = .408$, $p < .000$, $n = 128$
Data Analysis and Processing	$r = .523$, $p < .000$, $n = 128$
Measurement	$r = .657$, $p < .000$, $n = 128$
Geometry	$r = .597$, $p < .000$, $n = 128$

Note: There are only 128 records available for each of the TCAP sections.

Limitations

The developers of ALEKS have associated ALEKS' topics with the Tennessee state standards. Unfortunately, this association was based on face validity rather than rigorous testing and scientific methodology. Therefore, the correlations found in the current study may be threatened.

Study 2

Purpose of the study

Study 2 was designed in order to partially replicate (using a different school district and sample) the results found in Study 1. The results of Study 1 suggest that performance in ALEKS is significantly correlated with performance on the standardized high-stakes tests TCAP. Study 2

examines the relationship between students mastery of topics in ALEKS with their overall TCAP score while ignoring SPI. In this study, the relationship between ALEKS and TCAP scores is examined across all grades (fifth through seventh) as well as each grade independently.

Methods

Procedures

Shortly after the start of the school year, students were given access to ALEKS and were allowed to interact with the system up until the TCAP was administered between mid March and early April. As with the previous study, teachers used ALEKS as part of their regular mathematics instruction. Teachers allocated one day (usually Friday) a week as an "ALEKS day". Because most of the students participating in the study had access to an Internet connection at home, participants could additionally engage with the system outside of ALEKS day as well. Similar to the procedure in Study 1, two outcome measures were collected per student: 1) achievement as measured by the percent of ALEKS topics mastered from this point on referred to as Goal After in the ALEKS assessment, and 2) achievement in the TCAP from this point on referred to as TCAP as measured by the student's mathematics scaled score.

Participants

The initial sample consisted of a total of 1106 students (grade 5-8) from an urban school in the Memphis area that had used ALEKS since 2005. From this base sample, this study selected only those students that satisfied the following two conditions: 1) students enrolled in grades five, six, and seven and 2) students whom had 2009 TCAP scores available at the time of the study. Using these criteria, this study included 124 fifth graders, 98 sixth graders, and 99 seventh graders for a total sample of 321 students. Due to the lack of availability of TCAP

scores, the final sample did not contain any individuals in grade eight. Sample statistics for the grades are in Table 2.

Table 2.
Sample Statistics

Grade	*N*	Variable	Mean	Std Dev
5	124	TCAP	535.33	34.1276
5	124	Goal After*	0.45	0.1706
6	98	TCAP	556.28	36.5016
6	98	Goal After*	0.71	0.19
7	99	TCAP	567.69	32.6961
7	99	Goal After*	0.67	0.1919

* Percentage score

Data Coding

The data obtained from grades five through seven, was first evaluated by combining individual classes within grade level. The criteria for combining individual classes within grade level were based on similarity of ALEKS topics covered by each individual class. To merge individual class levels standardization of their respective TCAP scaled scores (ZTCAP), and Goal After percentage scores (ZGoal After) was conducted. Finally, to evaluate the strength of the relationship between the two variables ZGoal After and ZTCAP a correlation was calculated for each individual grade level.

In order to carry out an analysis between the TCAP and Goal After percentage across all grade levels, each of the grades' data were combined using their respective raw TCAP and Goal After percentage scores. Once the sample was combined the TCAP and Goal After percentage scores were normalized into z scores for further analysis of the strength of the relationship between the two variables.

Results

The standardized ZTCAP scores across all grade levels ranged from a minimum -3.05 to a positive 3.33. In addition, the standardized variable ZGoal After had a range of -2.30 to 1.87. Results of the correlational analysis, as can be visually observed in the scatter plot presented in Figure 2., revealed a statistically significant positive correlation between the standardized ZGoal After and the standardized ZTCAP when all of the grades were combined ($r = .74$, $p < .0001$, $n = 321$).

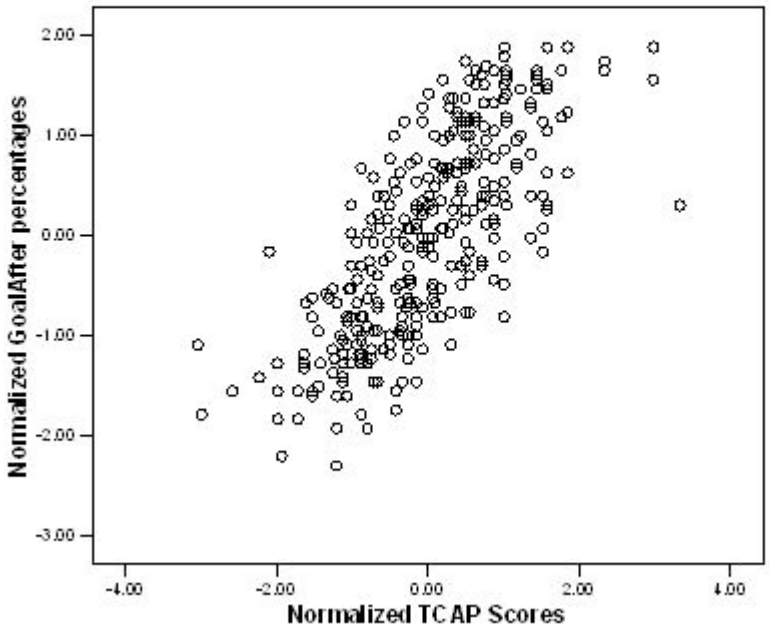

Figure 2. Scatter plot showing the relationship between TCAP scores and Assessment Performance in ALEKS

In addition, the analysis at the individual grade levels revealed a statistically significant positive correlation for the relationship between ZGoal After and ZTCAP (Table 3).

The results also revealed a statistical statistically significant positive correlation between the standardized ZGoal After and the standardized ZTCAP. However, it may be important to

determine if the mastery of ALEKS topics differed by grade level. Due to differences in topic mastery, it is possible that the correlation between the variable Goal After and TCAP scores would also differ.

Table 3.
Normalized Goal After with TCAP Correlation

Grade	Pearson Correlation
5	$r = .63$, $p < .0001$, $n = 124$
6	$r = .81$, $p < .0001$, $n = 98$
7	$r = .72$, $p < .0001$, $n = 99$

Limitations

The current analyses for this study were unable to distinguish whether students who had exposure to the ALEKS system outside of the classroom did better on mastery of topics than those who only used ALEKS in the classroom. Therefore, the study may be biased towards students who have with internet access at home.

General Discussion

Both studies discussed in this chapter revealed significant positive correlations between students' mastery of topics in ALEKS and their scores on the TCAP in the mathematics content domain. These studies appear to support the instructional supplement hypothesis that states that adding a computer-based learning environment could facilitate and possibly increase student achievement scores in mathematics. This finding is consistent with a number of other studies that have shown significant learning gains for students using learning technologies (e.g., Jonassen, Peck, & Wilson, 1999).

Given the magnitude of the problem of improving mathematics achievement in the United States, researchers and instructors should look at these results are encouraging. These

results may reveal the existence of possible cost-effective and time-effective ways to increase student mathematical achievement and success. This study provides encouraging results that presents a possible solution that could prevent the potential loss of funding through the NCLB.

As mentioned earlier, one-to-one human tutoring appears to be the gold-standard but unfortunately teachers do not have the time or resources to provide that level of individualized instruction to each student. These studies demonstrate a cost-effective efficient solution that provides individualized instruction to each student. Also, ALEKS appears to provide instructors with a reliable measure to monitor student progress throughout the academic year. In a traditional classroom, teachers give periodic assessments using such tools as paper and pencil tests and teacher prepared paper quizzes. The problem with these types of traditional assessments is that the reliability, effectiveness, and validity is unknown. In all but a couple of the current methods applied, the effectiveness, reliability, and validity of the measure are questionable.

If more accurate measurements of student achievement were available (e.g., ALEKS), this could potentially provide tremendous gains in the classroom and on standardized high-stake tests (e.g., TCAP). An instructor would have access to each student's ALEKS assessments, and therefore be able to accurately monitor a student's progress in mathematics proficiency. The ALEKS system provides a practical alternative providing feedback on the student's current ability. Thus, implementation of ALEKS into the classroom would allow teachers to devote more time and energy to the students that need it the most specifically in the areas where students need help while at the same time improving standardized test score such as TCAP.

References

Angrist, J. & Lavy, V. (2002). “New evidence on classroom computers and pupil learning”. *The Economic Journal, 112,* 735-765.

Baker, E. L., Gearhart, M., & Herman, J. L. (1994). Evaluating the apple classrooms of tomorrow. In E.L. Baker, and H. F. O’Neil, Jr. (Eds.) *Technology assessment in education and training.* Hillsdale, NJ: Lawrence Erlbaum.

Bloom, B. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13(6)*, 4–16.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19,* 237–248.

Gollub, J. P., Bertenthal, M. W., Labov, J. B., & Curtis, P. C. (2002). *Learning and understanding: Improving advanced study of mathematics and science in U. S. high schools.* Washington DC: National Academy Press.

Hagerty, G. W., & Smith, S. (2005). Using the Web-based interactive software ALEKS to enhance college algebra. *Mathematics and Computer Education, 39,* 183-194.

Jonassen, D., Peck., K., & Wilson, B. (1999). *Learning with technology: A constructivist perspective*. Upper Saddle River, NJ: Prentice Hall, Inc.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker, and H. F. O’Neil, Jr. (Eds.) *Technology assessment in education and training.* Hillsdale, NJ: Lawrence Erlbaum.

Melis, E., & Siekmann, J. (2004). ActiveMath: An intelligent tutoring system for mathematics. *In Proceedings of Seventh International Conference 'Artificial Intelligence and Soft Computing* (ICAISC), 91-101. Springer.

Sivin-Kachala, J. (1998). *Report on the effectiveness of technology in schools, 1990-1997.* Software Publisher's Association.

Wenglinsky, H. (1998). Does it compute? The relationship between educational technology and student achievement in mathematics. *Educational Testing Service Policy Information Center.*

Woolf, B. P., Arroyo, I., Beal, C. R., & Murray, T. (2006). Gender and cognitive differences in help effectiveness during problem solving. *Technology, Instruction, Cognition and Learning, 3,* 89-95.