Evaluating Adaptive, Computer-Based Mathematics Tutoring Systems: A Math Improvement and Feasibility Study

Angela L. Barrus, Kent E. Sabo, Stacey S. Joseph and Robert K. Atkinson

Arizona State University

Ray S. Perez

Office of Naval Research

Author Note

Angela L. Barrus, Division of Educational Leadership and Innovation, Arizona State University; Kent Sabo, Division of Educational Leadership and Innovation, Arizona State University; Stacey S. Joseph, Division of Educational Leadership and Innovation, Arizona State University; Robert K. Atkinson, The School of Computing, Informatics, and Decision Systems Engineering, Arizona State University; Ray S. Perez, Applied Instructional Research Program, Office of Naval Research.

This research was supported by Office of Naval Research under Grant N00014-10-1-0143 with Dr. Robert Atkinson as Principal Investigator.

Correspondence concerning this article should be addressed to Angela L. Barrus, Division of Educational Leadership and Innovation, Arizona State University, Tempe, AZ 85287.

E-mail: albarrus@asu.edu
Abstract

This study evaluated two off-the-shelf, adaptive and computer-based mathematics tutoring systems that teach algebra and other foundational math skills. Thirty high school algebra students that failed to pass algebra in the previous semester were randomly assigned in equal proportions to work with Carnegie Learning’s Cognitive Tutor or ALEKS algebra course product. Using the tutoring system exclusively, the students completed a 4-hour-a-day, 14 day summer school high school algebra class for credit. The results revealed that both tutoring systems produced very large learning gains, on the magnitude of a two-sigma effect, on measures of arithmetic and algebra knowledge.

Keywords: intelligent tutor, algebra, evaluation, mathematics, high school, summer school.
Evaluating Adaptive, Computer-Based Mathematics Tutoring Systems: A Math Improvement and Feasibility Study

This study was part of a larger multiyear research project designed to evaluate the effectiveness of off-the-shelf, adaptive and intelligent mathematics tutoring systems that teach pre-algebra, algebra and other foundational math skills. The results of this study were used to inform ongoing studies within this evaluation project tasked with determining the most effective intelligent tutoring system for math remediation in high school and military instructional settings. The two mathematics intelligent tutors selected for use and comparison in this study are Carnegie Learning’s Cognitive Tutor and the ALEKS algebra course product. These tutors are based on the current advances in artificial intelligence and cognitive science. The goal of this field evaluation was to compare the effectiveness of the two intelligent tutors and examine for trends in learning gains. Based on the results of this study we made recommendations concerning the ongoing use of intelligent tutoring systems as curriculum replacement in a summer school algebra course.

Now more than ever, math skills are fundamental to successful job performance. Although scientific work has always required a high level of mathematical ability, an increasing number of lower level jobs require math skills to operate high-tech equipment (Agondi, Harris, Atkins-Burnett, Heaviside, Novak, & Murphy, 2009). In response to a math achievement gap and the need for math skills in a competitive job market, the No Child Left Behind Act (2001) required schools to make adequate yearly progress in math with the goal that all students meet or exceed proficiency by 2014. Schools across the country are falling far behind this goal as the 2007 National Assessment of Educational Progress showed that many students demonstrate only basic mathematics mastery (Lee, Grigg, & Dion, 2007). Not surprisingly, many curricular
approaches are implemented in school math classes but little rigorous research exists to prove their effectiveness (Slavin & Lake, 2007).

One curricular approach is the use of intelligent tutoring systems in order to leverage advances in artificial intelligence and cognitive science as well as the evolving power of the Internet. Several features differentiate intelligent tutors from more traditional computer-based instruction including the power to contextually track a student’s performance and carefully adjust the teaching approach based on a student’s learning needs (Woolf, 2009). Numerous studies have concluded that computer based systems (including ITS) designed to deliver math instruction and assessments provide positive learning effects (Murphy, Penuel, Means, Korbak, Whaley, & Allen, 2001; Beal, Arroyo, Cohen & Woolf, 2010). However, some researchers have discovered design flaws in many computer-based instruction studies. Few studies employed a randomized, experimental design; many were only descriptive studies, and many lacked relevant data and specificity (Waxman, Lin, & Michko, 2003).

Because algebra is a prerequisite for higher-level math and algebra proficiency is correlated with students’ success in college and in obtaining jobs, focus should be placed on the effectiveness of algebra curriculum. (Adelman, 1999; Carnevale & Desrochers, 2003). However, few studies exist that evaluate the effectiveness of computer based algebra instruction and even fewer yet that investigate intelligent tutoring systems. Additionally, of the algebra intelligent tutoring system studies, most are conducted by the system developers which increases the potential for experimenter bias.

Hannafin & Foshay (2006) evaluated a PLATO Learning's computer-based algebra product as part of a new high school remedial math program. The goal of the larger remedial
program was to increase scores on the math portion of the high-stakes state test. The treatment group included 87 students while 39 students were in the control. The treatment group was scheduled to work with the computer-based system for four of the five instructional days per week. Both the treatment and control made significant gains on the state exam. The mean score for the control was significantly higher than the mean score for the treatment group. However, the treatment group gain scores were significantly higher than the control group's gain scores (Hannafin & Foshay, 2006). This may be a statistical artifact that indicates regression to the mean as the treatment group's scores on the first state exam were significantly lower than the control group's scores.

In an early intelligent tutor study, Koedinger, Anderson, Hadley, & Mark (1997) evaluated a new algebra curriculum called PUMP and an intelligent tutoring system called PAT. Their experimental group included 470 students in 20 algebra classes that worked with the new algebra curriculum and the intelligent tutor. The control group included 120 students in five algebra classes who received a traditional curriculum and did not use the intelligent tutor. The experimental group worked with the intelligent tutor approximately 25 out of the 180 class meetings which lasted 44 minutes a piece. The researchers reported an a significant difference between the two groups on two standardized tests and two researcher created tests with the experimental group performing better on all the tests. The sigma effect on both standardized test was 0.3 and was 0.7 and 1.2 on each researcher created test. Percent correct on each of the four tests for the experimental group ranged from 52% to 32% (Koedinger et al., 1997).

As part of a large-scale study on reading and mathematics software, Campuzano, Dynarski, Agodini and Rall (2009) evaluated the computer based instructional program Larson
Algebra I and the intelligent tutoring system Cognitive Tutor Algebra I. The Larson program was a supplement to traditional instruction while Cognitive Tutor was used as the core algebra I curriculum. Larson students were logged on to the system for an average of 313 minutes per year with the usage occurring over six weeks. Cognitive Tutor students were logged into the system an average of 2,149 minutes a year with usage occurring over 24 weeks. Researchers reported no significant difference between treatment and control groups using computer-based and intelligent tutoring, algebra software products. Overall, the scores for the treatment students on Educational Testing Service's End-of-Course Algebra Assessment was 37.3% correct (Campuzano et al., 2009).

No independent, peer-reviewed study conclusively demonstrates the effectiveness of intelligent algebra tutors used as the exclusive method of instruction. This study sought to determine the effectiveness of two intelligent tutoring systems designed to provide algebra practice and instruction. Specifically, the goal of this intelligent tutoring implementation was to evaluate the effectiveness and learning gains provided when the systems were implemented as curriculum replacement during the intensified and concentrated time period allocated to a summer school session. This evaluation contributed evidence, independent of the software developers, on the effectiveness of algebra intelligent tutoring systems.

**Intelligent Tutoring System Foundations**

Intelligent tutoring systems are architected using information technology systems and student learning models based on ongoing research advances in cognitive science and artificial intelligence. They have been evolving in their ability to customize the learning experience to a student’s ability and simulate the efficiencies of human tutoring since the early 1970’s. Currently,
there are several commercial software companies developing intelligent tutoring systems as well as many educational research institutions (Woolf, 2009). A wide-spread adoption of these intelligent tutors in school systems, the military and other training venues has been slow in the past due to cost and complexity. However, the advances of the Internet now allow intelligent tutoring vendors to provide access to powerful and useable programs via web-based clients anytime, anywhere at a much lower cost. The Internet is moving a system traditionally constrained by space, time and cost into a more efficient knowledge and learning-driven industry (Woolf, 2009).

The delivery power of the Internet combined with the scientific advances in artificial intelligence and cognitive science have created a fertile environment for intelligent tutoring systems associated with many different content areas and learning tasks to become relevant, useful and accessible for students of all ages learning a variety of content areas and skills. “The implication is that personalized tutoring can be made widely and inexpensively available just as printed materials and books were made widely and inexpensively available by the invention of the printing press” (Woolf, 2009).

**Individualized Instruction**

Education is no longer a one size fits all issue since learning populations have undergone major demographic shifts and teachers in a variety of settings are struggling with the need to address their students’ individual learning needs (Woolf, 2009). One-on-one tutoring provided by human experts has been shown to be an extremely effective method of teaching and learning. In one of several seminal studies conducted to try to discover pedagogies that are as effective as one-to-one human tutoring, students tutored individually by master teachers performed better
than 98% of students who received classroom instruction (Bloom, 1984). These results have set
the standard for intelligent tutoring evaluation because intelligent tutoring systems seek to
provide a form of individualized learning using artificial intelligence to define a student’s
learning gaps and provide the practice, scaffolding and processes necessary to produce learning
gains similar or greater than those observed by Bloom with human tutors (Woolf, 2009).

Method

Evaluation Framework

Summative evaluation is a process implemented for the “purpose of determining the
merit, worth, or value of the evaluand in a way that leads to a final evaluative judgment” (Russ-
Eft & Preskill, 2009). A summative evaluation is “conducted after completion of the program
and for the benefit of some external audience or decision maker (for example, funding agency,
oversight office, historian or future possible users)” (Scriven, 1991). A summative evaluation
usually addresses how well a program or system met its intended goals, if and how the learning
outcomes were met, if the results were worth the associated costs, if participants benefitted from
their participation in the program and if and how the program or system is reproducible in other,
similar contexts. (Russ-Eft, 2009).

A specific type of summative evaluation known as an impact evaluation is used to “assess
the effects of a settled program” (Owen, 2006, p. 47) and focuses on what happens to participants
as a result of the intervention or program as they typically try to make causal connections
between an evaluand and an outcome (Russ-Eft, 2009). In order to conduct an impact,
summative evaluation of an intelligent tutoring system, an evaluation logic and framework must
be defined to “understand how one determines the processes and impact of a program”. (Russ-
Eft, 2009). This logic and framework typically includes establishing criteria, constructing standards, measuring performance and comparing with standards, synthesizing and integrating evidence into a judgment of worth and developing recommendations. (Russ-Eft, 2009).

This process of developing an evaluation framework is applied to educational technologies, including intelligent tutoring systems, but these systems are also evaluated differently than traditional classroom instruction or software. Educational technology evaluations seek to show that learning outcomes have been achieved and that the software or tool works appropriately but they usually go beyond laboratory or developer evaluations to demonstrate that they “are useful with real students in real settings” (Woolf, 2009).

Participants

Participants were 30 remedial high school algebra students (16 male and 14 female) who chose to enroll in a district-organized 14 day summer school high school algebra class. Students were not enrolled in any other summer courses and did not receive any other formal math instruction other than that provided by the summer school course. Each class meeting was four hours and started at 8AM. Tuition was $175 per session per ½ credit. Three high school math teachers were assigned by the district to supervise the summer school sessions.

Intelligent Tutors

This study evaluated two off-the-shelf, computer-based mathematics tutors that contained pre-algebra and algebra curricula (Carnegie Learning’s Cognitive Tutor and ALEKS algebra course product). Carnegie Learning’s Adaptive Intervention Solution is designed to provide mathematics instruction for struggling students by providing them with algebra I instruction. It provides interactive examples with step-by-step instructions, feedback, and recognizes common
student errors. The Cognitive Tutor is based on the Adaptive control of thought–rational (ACT-R) theory. ACT-R is a cognitive architecture and theory, the purpose of which is to model the processes of human cognition. In ACT-R theory, procedural knowledge called production rules control human cognition. These production rules take the form of If-then statements. Learning in this theory is comprised of three learning processes. First, experiences are coded into declarative knowledge called chunks. Second, the chunks are converted into the form of a production rule. Third, the production rules and chunks are strengthened through active repetition (Anderson, 1992).

ALEKS’s algebra course product is a web-based, artificially intelligent assessment and learning system. ALEKS uses adaptive questioning to determine student knowledge. ALEKS then instructs the student on the topics they are most ready to learn. As a student works through a course, ALEKS reassesses the student to ensure that topics learned are retained. ALEKS is based on knowledge space theory (KST). KST attempts to mimic the ability of an expert teacher to assess a student’s knowledge state (Doignon & Falmagne, 1985). KST is not a theory of human cognition, rather it is a theory that informed the creation of a computer-based assessment procedure that provides an accurate and continuously updated assessment of student knowledge. KST theorists define a knowledge state as a “particular subset of questions or problems that the subject is capable of solving (Falmagne, Doignon, Koppen Villano, & Johannesen, 1990, p. 201). A knowledge space for a certain topic is made up of all the knowledge states specific to that topic.

Instrument

The Accuplacer is a computer-adaptive placement test featuring ten modules three of
which are designed to measure mathematics skills and knowledge. Two modules were used for this study; arithmetic and elementary algebra. The 17-item arithmetic module measured students’ ability to perform basic arithmetic operations and to solve problems that involve fundamental arithmetic concepts. Questions were divided into three types: (a) operations with whole numbers and fractions: topics included in this category are addition, subtraction, multiplication, division, recognizing equivalent fractions and mixed numbers, and estimating; (b) operations with decimals and percents: topics include addition, subtraction, multiplication, and division with decimals (percent problems, recognition of decimals, fraction and percent equivalencies, and problems involving estimation were also given); and (c) applications and problem solving: topics included rate, percent, and measurement problems, simple geometry problems, and distribution of a quantity into its fractional parts. The 12-item elementary algebra module contained three types of questions. The first type involved operations with integers and rational numbers, and included computation with integers and negative rationals, the use of absolute values, and ordering. A second type involved operations with algebraic expressions using evaluation of simple formulas and expressions, and adding and subtracting monomials and polynomials. Questions involved multiplying and dividing monomials and polynomials, the evaluation of positive rational roots and exponents, simplifying algebraic fractions, and factoring. The third type of question involved the solution of equations, inequalities, word problems, solving linear equations and inequalities, the solution of quadratic equations by factoring, solving verbal problems presented in an algebraic context, including geometric reasoning and graphing, and the translation of written phrases into algebraic expressions (The College Board, 2011).
Procedure

Students completed the 14-day summer school session in two computer labs with PC-workstations connected to the internet. The software programs used in this evaluation were delivered through the Internet using subscription, web-based software clients that do not require individual installations on the computer workstations. Participant and parental consent was obtained through signed assent and consent forms from both the participant and the participant's parents, respectively.

On the first day of the summer school session, each student was randomly assigned to either the Aleks or Carnegie intelligent tutoring system and given a study booklet which outlined the purpose of the study, their assigned software and directions for its use, a detailed study protocol. The booklets also included a demographic questionnaire. Student were then given a brief verbal overview of the summer school course and the evaluation study. Students also completed the Accuplacer pretest modules on Day 1. On Day 2, students began working in their assigned intelligent tutoring system and practiced algebra skills and concepts using their tutoring system for the next 10 days for four hours per day. On a typical day, students would arrive at 8AM, log into their assigned tutor and begin working on the algebra content. Along with the system's help and hint functions, the supervising math teachers were available to the students to answer their algebra questions. Students were given regular breaks based on the district’s standard summer school schedule. On day 7, students completed the adaptive arithmetic reasoning and elementary algebra Accuplacer math tests as a repeated measure assessment intended to capture iterative learning gains. The final two days of the fourteen day program consisted of the students completing the Accuplacer module post tests and a demographic and
experience questionnaire as well as completing the district final exam that determined each student’s readiness to progress to the next math course in the high school sequence. By the conclusion of the fourteen day summer school course, each student had practiced and studied using his or her assigned intelligent tutoring system for approximately 35 hours. Students completed periodic assessments provided and used by the intelligent tutoring systems to regularly adapt the practice, pace and content to the individual student’s learning needs. The summer school teachers used these periodic tutor assessments as well as student progress in the tutor curriculum to assign students grades for the course.

Results

This study utilized a mixed within- and between-subjects design to evaluate the effect of the intelligent tutors on algebra test scores. The between-subjects factor was math software group with two levels (CL and AK) and the within-subjects was time with three levels (Day 1, Day 7 and Day 13). Two separate, two-way mixed ANOVAs were conducted to evaluate the effect of the two math programs on the Accuplacer tests. To evaluate the effect of the software programs on the Accuplacer algebra test, the between-subjects factor was math software group with two levels (CL and AK) and the within-subjects was time with three levels (Pretest on Day 1, Day 7 and Posttest on Day 13). For the dependent measure Accuplacer algebra score, results indicated a significant effect for time, Wilks’ Lambda = .52, F(2,27) = 12.72, p < .01, multivariate η² = .49, and a nonsignificant interaction effect, Wilks’ Lambda = .99, F(2, 27) = .15, p > .05, multivariate η² = .01. Follow-up analyses indicated that at each point in time the Accuplacer math scores differed significantly from one another and increased overtime. Students made significant gains on the Accuplacer algebra subtest from Day 1 to Day 13.
To evaluate the effect of the software programs on the Accuplacer arithmetic reasoning test, the between-subjects factor was math software group with two levels (CL and AK) and the within-subjects was time with three levels (Pretest on Day 1, Day 7 and Posttest on Day 13). For the dependent measure Accuplacer arithmetic reasoning score, results indicated a significant effect for time, Wilks’ Lambda = .44, F(2, 27) = 16.95, p < .01, multivariate $\eta^2 = .56$, and a nonsignificant interaction effect, Wilks’ Lambda = .99, F(2, 27) = .17, p > .05, multivariate $\eta^2 = .01$. Follow-up analyses indicated that at each point in time the Accuplacer arithmetic reasoning scores differed significantly from one another and increased overtime. Students made significant gains on the Accuplacer arithmetic reasoning subtest from Day 1 to Day 13.

**Discussion**

Intelligent tutoring systems attempt to replicate human one-to-one tutoring as described by Bloom (1984) to replicate the individual pacing and instruction delivery process. Other field evaluation studies. We believe our treatment intensity (time spent using the intelligent tutoring systems) contributed to the learning gains associated with this field evaluation. We also believe that a more optimal spacing effect (time between learning) contributed to learning gains. Our study suggests summer school as an effective implementation option. It is now important to investigate the optimal amount of practice time and spacing needed to achieve these learning gains within the constraints of a standard school-year schedule. Further, we intend to investigate the use of intelligent tutoring systems in regular classrooms during the standard school day and in an after-school program. And, indeed, future studies are needed. Despite the raw number of computer-based math instruction studies: 1) few are peer reviewed, 2) many exhibit design deficiencies, 3) most focus on elementary curricula, and 4) recent technologies are not studied.
EVALUATING ADAPTIVE MATHEMATICS TUTORING SYSTEMS

This study demonstrates that intelligent tutoring systems can be implemented by classroom teachers to provide an effective learning environment that may begin to simulate the live tutor and meet the needs of more students using a more efficient instructional delivery system.
References


http://professionals.collegeboard.com/higher-ed/placement/accuplacer
