Randomized Controlled Trial (RCT) Evaluation of a Tutoring System for Algebra Readiness

Carole Beal	Shandy Hauk	Steven A. Schneider	Weiling Li	Christopher Harrison
University of Arizona	*WestEd*			

Abstract: This paper reports on results of a randomized controlled trial evaluating the impact of an online tutoring system on Grade 6 students' mathematics learning. Starting with 71 teachers randomly assigned to use or not use the AnimalWatch program to replace 15% of mathematics class lessons, the study ended with 58 teachers' classes (35 Treatment, 23 Control). Data from over 1200 students are included in the analysis (795 Treatement, 496 Control). Small but significant effects were found. Treatment group gains were larger than Control group gains on a major standardized test and on project-based topic quizzes (Cohen's d=0.3). Exploration of fidelity of implementation indicated several factors at work and that the minimal effective implementation of this type was in classes where students completed at least 8 of 14 lessons spaced across several months. Software use records indicated that 75% of student-item interactions resulted in students persisting to a correct answer.

Objectives

The study goal was to evaluate the impact of an online tutoring system on Grade 6 students' mathematical proficiency relative to business-as-usual classroom instruction. In particular, the research question driving the work: *Do Grade 6 students who use the program AnimalWatch as replacement for about 15% of mathematics instructional time show greater learning gains than those who do not use AnimalWatch?* The project responds to the clear need to improve the mathematics proficiency in the United States. In international comparisons, U.S. students score in the average range overall, and much lower than students from other nations that are comparable in terms of economic development (Gonzales, Williams, Jocelyn, Roey, Kastberg, & Brenwald, 2008; Organization for Economic Cooperation and Development [OECD], 2010). Although there is some evidence that mathematics scores have improved over the last decade the growth remains slow, and in many states there has been no improvement at all. There is special concern about algebra as a gatekeeper course for subsequent success, pointing to the need to ensure that middle school students are prepared to succeed in algebra with thorough proficiency in basic computation, fractions, ratios and proportions and other algebra readiness topics (Stein, Kaufman, Sherman & Hillen, 2011; Wimberly & Noeth, 2005). Thus, educators, researchers and policy-makers are searching for approaches that can help to improve students' learning outcomes in math.

One possibility for addressing the identified need is the use of tutoring software for mathematics learning (Kulik, 2003). Such software is designed with the goal of providing individualized instruction to the learner via the use of features such as integrated explanations, videos and other multimedia resources to help them master concepts. In contrast, standard paper-based worksheets and end-of-chapter practice sets do not provide interactive, customized learning options. The opportunity to direct their own learning while using software may also provide a motivational advantage for some students. For example, struggling learners can choose to look at a multimedia lesson or explanation on the computer in private, whereas they might be quite reluctant to do so in the "live" classroom. Software also provides integrated tracking (e.g., problems completed, number completed correctly and so on) that can inform both students and teachers about progress in real time. These features are argued to lead to more effective learning (Bloom, 1984; Cohen, Kulik, & Kulik, 1982; Graesser, Conley & Olney, 2011; vanLehn, 2011; Woolf, 2009).

Theoretical framework

A considerable body of work suggests that the ideal learning context is one-on-one instruction with an experienced human tutor (Bloom, 1984). Human tutors present problems that help diagnose the student's sources of difficulty, choose problems within the student's Zone of Proximal Development, scaffold the student to a successful solution, and then attribute the success to the student's effort and enhanced understanding (Brown, Ellery, & Campione, 1994; Lepper, Woolverton, Mumme, & Gurtner, 1993). Human tutors balance the goal of providing challenge to ensure cognitive conflict, resolution and improved understanding, with the goal of sustaining the learner's motivation and persistence (Merrill, Reiser, Ranney, & Trafton, 1992). Unfortunately, it is very difficult for math teachers to provide much individual help in large classes. Many Grade 6 math teachers have four to five classes a day with 25-30 students or more per class. It is physically impossible for even the most dedicated and skilled teacher to diagnose and assist every student. Nor is it necessary when technology-based systems can perform some of these functions, providing students with individualized practice opportunities and, in turn, providing teachers with real-time reports of students' progress.

AnimalWatch is an intelligent tutoring system designed to give individualized instruction by selecting problems that will build a Grade 6 student's proficiency (Beal & Arroyo, 2002). Like other intelligent tutoring systems, AnimalWatch uses *student models* (representations of what the student is estimated to understand, in relation to the target domain: here a network of 30 math topics mapped to the California Grade 6 curriculum frameworks) to guide problem selection. Though not the topic of the current report, we note that the student's actions while working on problems (e.g., answers, clicks to view multimedia help resources) can be researched to update our understanding of the nature and importance of individual tracking of a model for each student. Although tutoring software is designed to include features intended to promote learning, research results in the K-12 classroom have been inconsistent. A recent meta-analysis of a wide range of computer-based instruction concluded that the use of technology was generally associated with improvements to student achievement (Tamin, Bernard, Borokhovski. Abrami & Schmid, 2011). VanLehn (2011) reviewed a battery of tutoring systems that adapt instruction in ways similar to what a human tutor would do and concluded that such software was as effective as human tutoring. Several studies suggest that there may be benefits to the use of tutoring software specifically for mathematics learning. Arroyo, Woolf, Royer, Tai and English (2010) reported that students who used a tutoring system for geometry along with modules for basic skills practice showed significant improvements in mathematics problem solving. Barrow, Markman and Rouse (2009) concluded that a computer-assisted instructional program for pre-algebra had a significant benefit for students, in comparison to business-as-usual instruction. Beal, Arroyo, Cohen and Woolf (2010) found that a tutoring system for middle school mathematics was comparable to small-group tutoring delivered by a mathematics teacher and also showed benefits in a quasi-experimental design study with a business-as-usual comparison condition. Ritter, Kulikowich, Lei, McGuire and Morgan (2007) reported positive results associated with the Cognitive Tutor, an instructional system for high school algebra that has been widely used in classrooms.

Although results from these studies are encouraging, conclusion are limited because they have typically involved fairly brief interventions, and the use of study-specific mathematics tests as outcome measures. In contrast, research using more rigorous experimental methodologies and more standardized outcome measures has not demonstrated a consistent benefit associated with tutoring systems. For example, a review of four widely-used mathematics tutoring systems for mathematics conducted by the What Works Clearinghouse concluded that there was no overall evidence that the software was associated with better student mathematics test performance (Campuzano, Dynarski, Agodini & Rall, 2009). Thus, there is a need for more systematic research on the potential benefits of tutoring systems to support mathematics instruction, and how teacher implementation may influence the effectiveness of tutoring systems. The study reported here was designed to address this need through a rigorous evaluation of the impact of one tutoring system for algebra readiness on students' mathematical proficiency.

Methods

Experimental design and participants

The study was a randomized controlled trial (RCT). Teachers were recruited and signed participation consent forms before being randomly assigned to either the Treatment or Control condition. Treatment teachers participated in a two-hour online professional development module designed to introduce them to the software and its features, and then integrated it into their instruction throughout the school year. Software use replaced approximately 15% of instructional time, usually through weekly visits to a computer lab or use of laptop carts in the classroom. Control teachers continued with business-as-usual instruction. The student sample was consenting Grade 6 students. Though begun with 71 teachers (36 Treatment group, 35 Control), due to student movement and teacher attrition, we ended the year with 1,010 in 35 Treatment classes and 636 students in 23 Control classes across 40 districts. Their districts identified about 20% of students as English Learners. In most districts, a letter sent home to caregivers notified them of study activities and offered an opt-out form to abstain from their child's scores being included in analysis; in one district, active caregiver consent was obtained for all participants.

Tutoring system

The AnimalWatch software evaluated in this study focused on word problem solving, generally considered a central component of mathematics proficiency (Koedinger & Nathan, 2004). As recommended by the National Council of Teachers of Mathematics (NCTM, 2000) and Common Core State Standards (CCSS, 2010) AnimalWatch integrates mathematics with meaningful content. In this case, with authentic environmental science material, focusing on narratives about tracking and monitoring the status of endangered and threatened species (hence the AnimalWatch system's name). By connecting mathematics problem solving with science, the student encounters many examples of mathematics in real-world problem contexts, at a point in the curriculum when many students begin to complain that mathematics is disconnected from their lives. The topic of environmental science is engaging to many pre-teens, both boys and girls, and aligns with the science frameworks for middle school (Grade 6 focuses on Earth Science, and Grade 7 on Life Science) as well.

Word problems. AnimalWatch includes approximately 1400 word problems organized into 20 *learning objectives*. Fourteen of these learning objectives were considered "core" for the study. Each learning objective centers on one topic in sixth grade mathematics such as operations with integers, fractions, or rates. Within each learning objects are narratives about endangered and threatened species, including the Giant Panda, Przewalski Wild Horse (Takhi), Right Whale, California Condor, Snow Leopard, White Shark, Great Cats, and Snakes of Asia. Each word problem includes an introduction with authentic background information, a graphic (image, figure or table) and a question derived from the introduction. Scientific terms in the word problems that may not be familiar to students are linked to an integrated glossary. Students enter their answer, including units, into an answer box and receive immediate feedback. For example, in one problem, a white shark is described as swimming from the ocean surface down to a depth of 30 meters below the surface, and then swimming up to a depth of 16 meters. The numbers in the problem are consistent with current scientific knowledge about shark behavior. The student is asked to compute the total distance traveled. Students must understand the concept of the number line, the analogy of the ocean's surface as the zero point, and the concept of elevation (defined in the integrated glossary) to solve the problem. The problem is linked to the state's mathematics standards for Number Sense (working with integers) and Mathematical Reasoning (deciding how to approach the problem, evaluating the answer for reasonableness in the context of the original situation). If the student makes one or two problem-solving errors, text feedback about accuracy is presented, followed by an operations hint (e.g., "No, that's not quite right." "Are you sure you're subtracting?"). A third error elicits a recommendation to view the associated help resources. When the student clicks on the "help" icon, a menu window appears, showing the options available for that topic, including text explanations (e.g., how to find the least common denominator), worked examples, interactive solutions, and video lessons. Students' requests to view help resources are automatically tracked, and these process data will be extracted for later analyses. A fourth error elicits the correct answer, which the student is required to enter before moving on to a new problem. Students also have the option of deciding a question is "too hard" and the answer is shown, but the question marked as incorrect.

Basic skills practice. Royer et al.'s (1999) work indicates that improved fluency with simple computation and mathematics facts is linked to improved problem solving performance on achievement test items. Thus, AnimalWatch includes *skill builders* that provide students with opportunities to build fluency through practice with multiplication facts, recognizing decimal-fraction-percent equivalents (e.g., identifying that 1/8 = 0.125 = 12.5%), rounding and estimation, one-variable equations (e.g., finding the value of x when 2x = 10), finding the least common denominator of two unlike fractions, and operating with negative integers. Each skill builder set presents the student with a series of 10 mathematical statements, one at a time (e.g., 35 + (-15) = 20) which the student identifies as either true or false; the immediate feedback is a green check if correct and red x if not. Skill builders can be repeated multiple times with new items. The skill builders are motivating for students because they can achieve near-perfect scores with repetition, and repetition in turn helps to builds fluency with the concepts and operations, indicated by decreases in time required to solve the items.

Instructional resources. Teachers can view and download reports on students' efforts – these grade books include the option of clicking to view the audit trail details of the student's work on each word problem. The teacher side of AnimalWatch also includes online resources:

- A professional development manual and curriculum guide that can be viewed online or downloaded in PDF format; and
- A users' wiki, with documentation, troubleshooting tips, discussion forum, and frequently asked questions.

Data sources

Data sources about students' included mathematics and English/language arts scores on the end-of-year state achievement test (not a matched pre/post designed test), pre- and post-intervention test scores on the Mathematics Diagnostic Testing Project (MDTP) test (a matched, pre/post design), and pre- and post-intervention scores on a study-specific test of word problem solving. Complete data were not available for all students. Sample size for each of these is summarized in Table 1.

Table 1. Numbers of student scores on study measures.

Instrument	*Control*	*Treatment*	*Total*
State Tests	476 (75%)	795 (79%)	1271 (77%)
MDTP (paired)	371 (58%)	676 (67%)	1047 (64%)
Topic quizzes (paired)	377 (59%)	632 (63%)	1009 (61%)

Treatment group students' problem solving in software. Students' actions while using the software were automatically recorded, including time logged in, problem solving performance, and completion of Learning Objectives.

Teacher logs. Teachers completed weekly online logs to document their usage and report concerns or comments about implementation. Near the end of the school year, classroom observations were also conducted with a subset of teachers who had exhibited high, medium or low fidelity to the implementation guidelines based on the audit trail of their class' use of the system and their self-reports in the logs.

Results

Table 3 summarizes the score analysis and significance of student achievement gains on the various assessments. For example, in the "Mean Gain in MDTP scores" column the Control group mean gain on MDTP was 18.6% and the top two green boxes indicate Treatment group mean gain on MDTP of more than 27% was statistically significantly higher *when Treatment group students used AnimalWatch enough that they completed 5 or more Learning Objectives*. For Treatment group students completing 4 or fewer Learning Objectives, their MDTP gain (25.4%) was not statistically significantly different from that for Control. That is, if students used AnimalWatch enough (5 or more lesson objectives completed), it was reflected in a statistically significantly greater achievement gain on the MDTP. A companion result, when students completed 8 or more learning objectives, is that CST and Topic Quiz mean gains or effects were significant. The effect size, Cohen's *d*, is included.

Table 3. Scores and gain in scores between Treatment subgroups and control.

Treatment by AnimalWatch usage		CST 2011 Mathematics	Mean Gain in MDTP scores	Mean Gain in Topic Quiz Scores	Mean Gain in Topic Quiz - Word Problems Only
Learning	1 to 4	357.56	25.4% (*d*=0.1)	44.4% (*d*=-0.1)	42.4% (*d*=-0.1)
Objectives	5 to 7	379.13	27.5% (d=0.3)	52.8% (*d*=0.2)	52.1%% (*d*=0.2)
Completed	8 to 21	412.55	22.9% (*d*=0.3)	49.9% (*d*=0.2)	51.4 %(*d*=0.3)
Control		*381.80*	*18.6%*	*44.8%*	*42.2%*

Key: Treatment scores shaded in green have a significantly higher score than the control group, $p < .05$. Cells shaded in lighter green indicate a small effect size.

MDTP. Difference scores representing the change from pre- to post-test were created for each student, and then compared by condition, controlling for differences in the two groups. The results are that students in the Treatment group improved more than those in the Control group (14 percentage points versus 10 points score gain).

State achievement test. The analytic strategy focused on using the previous year's mathematics test score as a predictor of mathematics test scores at the end of the study year, along with condition (Treatment, Control). We first eliminated 114 Treatment students who completed 5 or fewer of the core Learning Objectives over the course of the school year, leaving a sample of 909 students in the Treatment group. A regression analysis was conducted with prior-year test scores and condition (Treatment, Control) as predictors of current-year test scores. Not surprisingly, the results showed a very strong effect of prior year on current year test scores. In addition, there was a small but significant effect of condition, with Treatment students showing more improvement than Control students.

Study-specific test of word problem solving. Treatment group students who completed a high number of learning objectives had significantly greater gains than Control group students on project-designed tests.

Treatment group students' software use. Students in the Treatment group completed an average of 261 word problems over the school year, roughly corresponding to 14 learning objectives. Average performance across problems is shown in Table 2. Students solved about 75% of the problems correctly within three attempts, got about 20% wrong, and indicated that just over 5% were "too hard." Although some researchers have reported concerns about students "gaming" software, meaning that they attempt to avoid effort, in the present study students only used the "too hard" option on about 5% of the problems over the course of the school year.

Table 2. Average performance types across AnimalWatch word problems.

% Correct 1st Try	% Correct 2nd Try	% Correct 3rd Try	Strike Out (Incorrect)	Too Hard
49.8	16.5	8.6	19.5	5.4

Overall, students accessed the multimedia help features in the software on about 8% of problem attempts. There was a significant relation between the number of problems on which the student viewed the help features and improvement from pre to post-test on the MDTP (with $F(1,571) = 18.434$, $p < .001$). In addition, the percentage of word problems on which students viewed the integrated help features varied significantly across teachers (from 2% to 27% of problems). In addition to the prior year state achievement test score, the teacher was also a significant predictor of the current year's achievement test score for Treatment students. Teachers whose students used more of the software resources were also more likely to have students who performed better than would have been expected based on their prior year's test score.

Teacher Fidelity of Implementation

In addition to weekly logs completed by teachers, we observed 8 teachers in their classrooms in the spring term: 3 whose AnimalWatch student usage was low (below the minimum threshold for the study), 2 for whom it was moderate (just meeting requirements) and 3 for whom usage was high. The chart in Figure 1 summarizes findings on fidelity of implementation components among the observed teachers. Highly faithful implementation would be represented by score of 4 in each of the 4 areas (or a total of 16 points). The chart is representative of the distribution of responses to log items about the four constructs across all 35 Treatment teachers.

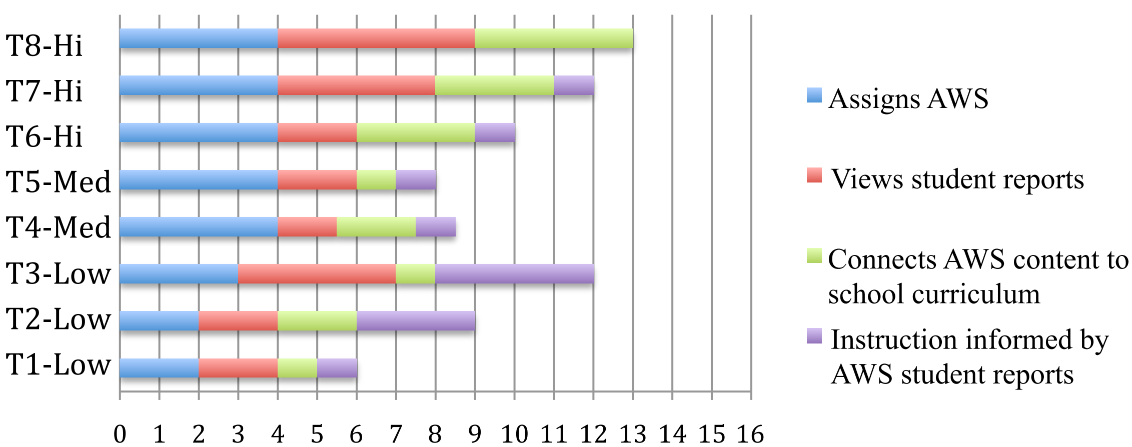

Figure 1. Observed teachers responses to fidelity of implementation questions.

The results shown in the chart and coding of observations suggest that in a Fidelity Matrix that compares Stages of Concern with Innovation Configuration factors, the teachers observed cluster around Routine-Consequence (based on Hall & Hord, 2006). That is, most teachers observed were comfortable enough with the use of AnimalWatch to treat it as a routine part of planning and instruction while also having mostly resolved concerns about management of software use and having a focus on the local instructional consequences of folding AnimalWatch use into instruction (e.g., consequences for learners, for the curriculum, for pedagogy). These two category placements (Innovation Use Level of 4A-Routine and Stage of Concern 4-Consequence) coincided with the minimal goal level for faithful implementation set by the AnimalWatch designer. One implication of this result is that different or additional supports for teachers would need to be put into place to see refinement (Level 4B) or smooth integration (Level 5) of the use of the software into classroom planning and practice as well as to scaffold teachers in further development along the concern dimension into more substantive concerns about impact (e.g., that involve collaboration (Level 5) and potential agency (Level 6) in the massaging of curriculum and instruction to incorporate the software as a valued component of instruction).

Conclusion

Results from this evaluation study were encouraging with regard to the potential of tutoring systems technology to support mathematics learning. There was a small but significant positive effect on test performance, with students who used the software at least 8 times during the year showing more improvement than those who did not. The effect was small but consistent across multiple measures of mathematics proficiency. In addition, variations in improvement for the Treatment students were related to variations in implementation across classrooms. More specifically, students whose teachers encouraged use of the integrated instructional resources in the software showed more improvement than those in Treatment classrooms where help use was low. At the classroom level, performance improved more in classrooms where the Treatment group teachers reported high fidelity of implementation. That is, the active component in the Treatment condition appeared to be the effective implementation by teachers, not simply the presence of software in the classroom. Continuing analyses seek to understand relationships among variables by examining a second year of study of the same software with a larger group of students (the same teachers as in the study reported here, plus 20 new teachers) and the use of hierarchical linear modeling.

Though we did seek information through weekly logs and surveys of enacted practice, one limitation of the study is that we do not know why some teachers did not use the software as much or as consistently as intended. The challenges teachers reported in logs ranged from confusion or frustration in gaining access to computers/labs to remembering to spend time before class assigning tasks (rather than spending their time in front of computer making assignments on-the-fly during class), to firmly expressed doubts about the effectiveness of computer tutoring. Logistical barriers to technology use in classrooms are still an issue in many districts. In the present study, school networks were often unreliable, and technical support was nearly non-existent in most of the participating schools. Future work will need to address the ways in which teachers can most effectively be supported in using technology-based instructional tools to support student learning.

Acknowledgements

Thanks to the teachers, district personnel, students, and their caregivers for their participation and support. Additionally, the study benefited from the diligent efforts of Katie D' Silva, Bob Allen, Cathy Ringstaff, and others at WestEd as well as Jane Strohm, Bill Mitchell, and Tom Hicks at the University of Arizona. The contents of this report were developed under grant number R305A090197 from the Department of Education. However, the contents do not necessarily represent the policy of the Department of Education, and the reader should not assume endorsement by the Federal Government. The findings and conclusions in this report do not necessarily represent the official positions or policies of the funders.

References

Arroyo, I., Woolf, B. P., Royer, J. M., Tai, M., & English, S. (2010). Improving mathematics learning through intelligent tutoring and basic skills training. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Lecture notes in computer science*, 6094/2010 (pp. 423-432).

Beal, C. R., & Arroyo, I. (2002). The AnimalWatch project: Creating an intelligent computer math tutor. In S. Calvert, A. Jordan, and R. Cocking (Eds.) *Children in the digital age* (pp. 183-198). Westport, CT: Greenwood.

Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy, 1(1),* 52–74.

Beal, C. R., Arroyo, I. M., Cohen, P. R., & Woolf, B. P. (2010). Evaluation of AnimalWatch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning, 9*, 64-77.

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,* 4-16.

Brown, A. L., Ellery, S., & Campione, J. (1994). Creating Zones of Proximal Development electronically. In J. Greeno & S. Goldman (Eds.), *Thinking practices: A symposium on mathematics and science education.* Hillsdale, NJ: Erlbaum.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts* (NCEE 2009-4041). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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.

CCSS - National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common Core State Standards*. Washington DC: Author.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth and eighth grade students in an international context* (NCES 2009-2011 Revised). National Center for Education Statistics, IES, U.S. Dept. of Ed., Washington DC.

Graesser, A. C., Conley, M. W., & Olney, A. (2011). Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *American Psychological Association Handbook.* Washington, DC: American Psychological Association.

Hall, G.E. & Hord, S.M. (2006). *Implementing change: Patterns, principles and potholes* (Second Edition). Boston, MA: Allyn and Bacon.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences, 13,* 129-164.

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluations studies say* (SRI Project No. P10446.001). Arlington, VA: SRI International.

Lepper, M. R., Woolverton, M., Mumme, D., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 75-105). Hillsdale, NJ: Erlbaum.

Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *Journal of the Learning Sciences, 2*, 277-305.

OCED. (2010). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science*. DOI: 10.1787/19963777

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston VA: Author.

Ritter, S., Kulikowich, J., Lei, P., McGuire, C. L., & Morgan, P. (2007). What evidence matters? A randomized field trial of Cognitive Tutor Algebra I. *Supporting Learning Flow through Integrative Technologies, 162*(1), 13-20.

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Merchant, H. (1999). Math fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24*, 181-266.

Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research, 81*, 453-492.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P.C. & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research, 81*, 4-28.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

Wimberly, G. L., & Noeth, R. J. (2005). *College readiness begins in middle school.* Iowa City, IA: ACT.

Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning.* Burlington MA: Morgan Kaufman Publishers.