In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), Artificial intelligence in education: Building technology rich learning contexts that work, pp. 536-538. Amsterdam: IOS Press.

Intelligent tutoring and human tutoring in small groups: An empirical comparison

Carole R. BEAL, Erin SHAW, and Mike BIRCH
Information Sciences Institute - USC
4676 Admiralty Way, Marina del Rey CA 90292 USA
{ cbeal, shaw, mbirch@isi.edu }

Abstract. The efficacy of a tutoring system for pre-algebra instruction plus human tutoring was compared to instruction provided to small groups of middle school students by experienced human math tutors, with instructional time held constant. Students completed pre- and post-tests of computation, fractions, algebra and rational numbers skills. Results indicated that students showed significant improvement from pre- to post-test, but there was no difference as a function of type of tutoring. The findings help to establish the efficacy of ITS instruction relative to skilled human tutoring of students in small groups.

Introduction

Individualized instruction is generally recognized as the "gold standard" for instruction. Students learn most rapidly and deeply when they can work one-on-one with an experienced human tutor. Research indicates that the instruction provided by intelligent tutoring systems is also effective [1, 2]. However, to date tutoring systems have typically been compared to whole-class instruction, i.e., "business-as-usual" instruction. We do not yet know how tutoring systems compare to human instruction provided in a more individualized manner. Ideally, this would involve a comparison of an ITS to one-on-one instruction with a human tutor, yet this situation is difficult to arrange in the context of in-vivo classroom research. However, as an initial investigation, in the present study we compared the efficacy of an ITS to instruction provided by experienced human tutors to small groups of students.

1. Method

1.1 Data sources

Participants included 32 middle school students (incoming Grade 7) attending an academic summer program held at the university campus. Students attended the program every Saturday for four hours, with instructional time divided between language arts (2 hours) and mathematics (2 hours). The program was structured so that students worked in small groups, with 6-8 students per adult tutor. Tutors for the math sessions were experienced mathematics teachers from local schools.

1.2 Design

For the present study, two groups of students for each of two tutors were assigned either to receive one hour of computer tutoring and one hour of human tutoring for each math session,

or to receive two hours of human tutoring (small-group instruction) per session, for 6 sessions. Human tutors provided blackboard lessons and worksheet practice.

1.3 Computer tutoring

Students worked with AnimalWatch, a web-based tutoring system in which instruction in pre-algebra mathematics was integrated with environmental science content. The program presented word problems about endangered species. Problem selection and math topic were determined by the pedagogical model, based on the student's ongoing problem solving performance. Students also practiced basic math skills through AnimalWatch SkillBuilder units on math facts, decimal equivalents, rounding and estimating, and other pre-algebra material.

AnimalWatch provided immediate feedback when an answer was submitted. Incorrect answers elicited a text message, e.g., "Your work is improving but that's not correct". A second incorrect answer elicited an operations hint, e.g., "Are you sure you are adding 312 and 434?". Students who still needed help after the text hints could click a "Hints" icon to view complete worked examples, as well as interactive hints. Students' answers and use of hints were logged automatically into the database.

1.4 Assessments and scoring

All students completed a pre-test before the tutoring sessions, followed by a post-test at the conclusion of the program. The tests included 30 items: ten computation items; 6 fractions items; 6 algebra items; and 8 items assessing rational numbers skills such as ratios, proportions, and unit conversions. Tests were constructed so that items had multiple clones that were equivalent in difficulty. Students completed the tests in 45 minute sessions under the supervision of the program instructors. Students received scores for the proportion of computation, fractions, algebra and rational number items completed correctly on the pre- and post-tests.

2. Results

2.1 Pre-test

An ANOVA with Tutoring Condition (Small Group; ITS + Small Group) as the grouping factor and pre-test proportion correct scores (Computation, Fractions. Algebra, Rational Numbers) as dependent measures. Results indicated no differences in performance as a function of Tutoring Condition, indicating that students in the two tutoring conditions had similar skills before the program activities began. There was a significant effect of Math Topic: Students scored highest on Computation (65% correct) and Fractions (66% correct) items; then Algebra (52%), with the lowest scores for Rational Numbers (13% correct). Mean scores are shown in Table 1.

2.2 Pre- to post-test comparisons

To evaluate the impact of the program, an ANOVA was conducted with Tutoring Condition as the grouping factor, Test (Pre-, Post-) as a within-subjects factor, and total proportion correct on each test as outcome measures. There was a significant effect of Test, $F(1,23) = 27.82$, $p < .001$. As may be seen in Table 1, students showed significant improvement from pre- to post-test. However, there was no effect of Tutoring Condition, indicating that both groups improved to the same extent.

Table 1. Mean proportion correct for Pre- and Post-tests (Standard deviations in parentheses)

Group:	Computation	Fractions	Algebra	Rational Nos.
Only human tutoring Pre test	0.62 (0.19)	0.62 (0.36)	0.47 (0.45)	0.09 (0.17)
ITS + Human tutoring Pre test	0.66 (0.19)	0.63 (0.37)	0.49 (0.40)	0.13 (0.28)
Only human tutoring Post test	0.74 (0.13)	0.74 (0.20)	0.80 (0.19)	0.49 (0.27)
ITS + Human tutoring Post test	0.72 (0.19)	0.73 (0.24)	0.84 (0.20)	0.45 (0.40)

To investigate the math topic where tutoring had most impact, we conducted a repeated measure ANOVA comparing Pre and Post test scores of each of the four math topics, with Group as a between subjects factor. There were no effects associated with the type of tutoring (only human tutoring, or half-human and half-ITS tutoring). For Computation and Fractions problems, the increase from pre- to post-test, although consistent in direction and across participants, was not significant. For Algebra items, there was a significant improvement from pre- to post-test, $F(1,26) = 63.77$, $p < .001$. There was also a significant improvement for items involving rational numbers concepts, $F(1,25) = 25.82$, $p < .001$. This suggests that tutoring had most impact on the most challenging material.

3. Discussion

The results of the study suggest that ITS instruction can be a valid supplement for instruction. When instructional time was held constant, students who received half of their classroom learning opportunity at the computer performed as well on the assessments as students who received all instruction from an experienced human tutor working with small groups. Both program formats were effective: students showed significant increases in math skill from pre- to post- test, particularly on the more difficult topics such as fractions, ratios, proportions, and equivalents.

The next step is to drill more deeply in to the patterns of ITS usage exhibited by individual students who received ITS instruction for half of their instructional time. For example, we would expect that a student who showed significant improvement on fractions problems on the assessments should have received ITS problems and scaffolding in that topic. Failing to find such a relation would suggest an alternative interpretation focusing more on the motivational benefits of computer-based practice.

Acknowledgments

The study was supported by a grant from the Institute of Education Sciences Mathematics and Science Education program. We would like to thank the program tutors and program director Kim Thomas-Barrios for their assistance with the study, as well as project members Lei Qu, Jean-Philippe Steinmetz, Hyokyeong Lee, Sharan Volin, and Ivon Arroyo.

References

[1] Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P. (2007). On-line tutoring for math achievement testing: A controlled evaluation. Journal of Interactive Online Learning, 6, 43-55.

[2] Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. J. (2000). Carnegie Learning's Cognitive Tutor: Summary of research results. White Paper. Pittsburgh PA: Carnegie Learning.