Feedback during Web-Based Homework: The Role of Hints

Ravi Singh[1], Muhammad Saleem[1], Prabodha Pradhan[1], Cristina Heffernan[1], Neil T. Heffernan[1], Leena Razzaq[2], Matthew D. Dailey[1], Cristine O'Connor[3], and Courtney Mulcahy[3]

[1] Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609 USA
{ravisingh,msaleem,prpradhan,ch,nth,mdailey}@wpi.edu
[2] University of Massachusetts Amherst, 140 Governors Drive, Amherst, MA 01003 USA
leena@cs.umass.edu
[3] Oak Middle School, 45 Oak Street, Shrewsbury, MA 01545 USA
{COConnor,CMulcahy}@shrewsbury.k12.ma.us

Abstract. Prior work has shown that computer-supported homework can lead to better results over traditional paper-and-pencil homework. This study about learning from homework involved the comparison of immediate-feedback with tutoring versus a control condition where students got feedback the next day in math class. After analyzing eighth grade students who participated in both conditions, it was found that they gained significantly more (effect size 0.40) with computer-supported homework. This result has practical significance as it suggests an effective improvement over the widely used paper-and-pencil homework. The main result is followed with a second set of studies to better understand this result: is it due to the timeliness of feedback or quality tutoring?

Keywords: evaluation of CAL systems; intelligent tutoring systems; interactive learning environments; secondary education; teaching/learning strategies.

1 Introduction

The increasing popularity of computer assisted learning (CAL) applications in schools and colleges has led to the development of various web-based CAL tools that aim towards improving the quality of student learning. The scope of CAL systems has expanded from tools used in classrooms to web-based applications that are capable of supporting and guiding students through homework as well. Many preparatory tools for mathematics tutoring have been developed and tested. WebWork (www.webwork.rochester.edu), WebAssign (www.webassign.com) and Blackboard (www.blackboard.com) are web applications that are already popular across colleges in the US. The introduction of educational technology has also been increasing recently in K-12 grades. For instance, since 2002, the state of Maine has introduced a 1:1 laptop program for 7th and 8th grade students and their teachers. A study [1] on the impact of 1:1 computing programs has shown increased motivation and engagement in classrooms and better retention of content material based on reports from teachers.

G. Biswas et al. (Eds.): AIED 2011, LNAI 6738, pp. 328–336, 2011.

The use of web-based homework by teachers is more feasible through programs such as Maine's program and through the increasing development in educational technology. Teachers may use web-based homework to supplement or replace conventional teaching methods such as paper-and-pencil homework. However, the increasing popularity of such technology brings into question the advantages in terms of effectiveness of web-based homework for students. Concerns are often raised about the cost-effectiveness of technology towards improving the standard of education.

Feedback on homework has been shown to have a large positive effect on student learning [2]. Most CAL systems used for homework attempt to improve the quality and timeliness of the feedback for homework. But, the effectiveness of computer-supported homework is still largely debated. Thus, in order to determine how a computer-supported tutoring system may affect student performance on their homework, a study by Mendicino, Razzaq & Heffernan [3] analyzed the effectiveness of computer-supported homework over traditional paper-and-pencil homework for K-12 students. Mendicino et al. reported an effect size of 0.61 in favor of computer-supported homework over paper-and-pencil homework. However, a study conducted using WebAssign [4] to deliver computer-supported homework for college level students showed no significant difference in the student performances.

In this study, we aim to improve the experimental design of Mendicino et al. [3] to further understand the effectiveness of web-based tutoring systems for delivering homework and improving student learning for K-12 students. We further analyze the characteristics of this tutoring mechanism to determine what factors contribute towards its effectiveness. The ASSISTment System was used for this study.

The ASSISTment System (www.assistments.org) is a web-based tutoring system, capable of offering instructions to students while providing detailed evaluations of their performance to teachers [5]. The system integrates assistance and assessment to efficiently tutor students in mathematics and is being used by middle and high school teachers throughout Massachusetts. Teachers may use the system as part of their coursework to assist students in learning while also obtaining detailed reports on individual students. Teachers may then identify difficulties students may be facing to tailor their instruction to be more effective. The system is free to use and supported by grants from the U.S. Department of Education and the National Science Foundation.

2 Experiments

Three experiments for investigating the effectiveness of web-based homework were conducted. In the first experiment (Experiment-1), we attempt to strengthen the claims made by Mendicino et al. [3] comparing computer-based homework with paper-and-pencil homework. This study compares learning gains for students in two conditions; *Immediate Feedback with Tutoring* (IFT: where students received homework with tutoring and immediate feedback on each problem) and *Business as Usual* (BAU: where student received no feedback). However, this study confounded the effects of Immediate Feedback and Tutoring. Experiment-2A & 2B were designed to understand the independent effect of Tutoring controlling for immediate feedback.

2.1 Experiment-1

Mendicino, et al. [3] compared web-based homework with paper-and-pencil homework. The students in the web-based homework condition used ASSISTments to complete their homework and so received immediate feedback through hints and scaffolding. The paper-and-pencil homework group completed their homework on paper and received feedback the next day in class. Hence, this study analyzed whether homework could be improved with immediate feedback with tutoring and found positive results in its favor.

However, as stated earlier Mendicino et al. [3] chose student classes as units of assignment for their conditions, but analyzed data at the student level. According to the What Works Clearinghouse [6] the unit of assignment should be the unit of analysis. Having classes as the unit of assignment and analyzing at the student level may lead to overestimation of the observed effects. Also a large Randomized Controlled Trial (RCT) number is recommended to evaluate educational software in reading and math. Mendicino, et al. had an RCT number of 4, as there were 4 different classes. Also, the use of the same test for pre- and post-tests may have created a test-retest effect and contributed to overestimation of learning rates.

In Experiment-1 we attempt to replicate this experiment by expanding the sample size and making some critical changes to the design and procedure. One major change in the experiment is the use of ASSISTments by both treatment and control groups to complete their homework. Instead of providing paper-and-pencil homework to the control group the students received Test Mode problems, and the treatment group received Tutoring Mode problems. Problems in Test Mode provide no feedback to the students and so were used as a replacement for paper-and-pencil homework. The unit of assignment was at the student level and not the class level and counter-balanced pre- and post-tests were used rather than using the same tests for both cases.

Experimental Design. The students in eight classes were randomly assigned based on their last names to the Immediate Feedback with Tutoring (IFT) condition or the Business as Usual (BAU) condition. After treatment, the conditions for the student groups were then switched. This provided a repeated measure for each participant. All students received two computer-based homework assignments as per their conditions. Before the homework, every student was given a pre-test and after completion of the homework a post-test was administered. In order to account for any test-retest effect two different test forms (Form-A1 & Form-B1) were randomly distributed to students for the first pretest. The students who received Form-A1 were provided with Form-B1 for the first post-test and vice versa. The same was done with test forms (Form-A2 & Form-B2) for the second round of pre- and post-tests when the second homework was assigned. These tests were paper-and-pencil based.

The pre- and post-test and homework assignments consisted of 10 problems each that were intended to be a Geometry and Number-Sense Review for the students. The assignment tested understanding of supplementary angles, properties of triangles, properties of quadrilaterals and parallel lines, transversals and the Pythagorean Theorem. The pre- and post-test consisted of problems that were very similar to problems from the homework assignment.

Procedure. Eight 8th grade classrooms with computers and the students' home computers were the settings used for this study. The students were familiar with the ASSISTment system and had used it for math homework before. Two teachers instructed four classes each and the total number of students was 172.

On the first day of the experiment all students completed a pre-test in class. The students then were assigned homework to be done with ASSISTments. The IFT group received homework with immediate feedback in terms of correctness with tutoring, which consisted of 3-4 hints for solving the problem, and the BAU group received no feedback at all. The following day, the teachers reviewed selected problems from the homework in class. After the review, students completed a post-test. The next day, students completed another pre-test and were assigned a second homework where the treatment and control groups were switched to provide a repeated measure for each student. The teachers reviewed the homework on the following day. This review session was videotaped to analyze the quality of their feedback. The students completed the second post-test after the review. The data from the first and second round of pre- and post-tests were then analyzed as paired samples**.**

Results. The eight classes included in this study had a total of 172 students. For the first homework assignment, 22 students in the *BAU* condition and 15 students in the *IFT* condition did not complete the homework. For the second homework, 14 students in the *BAU* conditions and 23 students in the *IFT* condition did not complete the homework. After the first homework assignment was assigned, some students in the *BAU* condition might have received feedback after completing the assignment by visiting a report page on the ASSISTments website. Due to this fact, we excluded 30 students who received this form of immediate feedback when in the *BAU* group. For the second homework assignment, the report page was disabled so that students in the *BAU* condition could not receive immediate feedback or tutoring from the system. Excluding these students and those who were not present for all or part of the experiment, 68 students participated in the study.

Based on the gain scores for the students, overall learning was observed in both conditions. The mean gain for the students in the *IFT* condition was 2.4 (SD=1.81), whereas it was 1.63 (SD=1.93) for the *BAU* condition. Both gain scores were reliably different than zero, $t(67)=10.9$, $p < 0.001$ for *IFT* and $t(67)=6.97$, $p < 0.001$ for *BAU*. Furthermore, comparing the gain scores of the two different conditions showed a reliable difference, $t(67) = 2.322$, $p = .023$, with higher gain scores in the *IFT* condition than in the *BAU* condition. The effect size observed in the direction of *IFT* was 0.40 with the 95% confidence interval of (-0.03 – 0.86).

The results suggest that students learned more from homework in the *IFT* condition as opposed to the *BAU* condition. Figures 1a and 1b show the distribution of gain scores for the students in the two conditions. From the graphs, it can be seen that the students in the *IFT* condition earned higher gain scores than those in the *BAU* condition and that our analysis is not sensitive to a few students.

However, certain factors such as the difference in the pre- and post-test forms and the quality of delayed feedback provided by the two different teachers may have had a significant impact on the results of student learning. We decided to dig deeper into the effect of these factors by doing additional analysis. We chose to focus on the second day of the experiment since the teacher reviews were videotaped then.

To test the effect of the different pre- and post-tests, a one-way ANOVA was performed with respect to Form as the independent variable. The test Forms that were assigned did not reliably predict post-test gains, $F(1, 119) = 0.78$, $p = 0.38$. We conclude that the test forms were balanced and excluded Forms from the analysis.

The next factor considered was the difference in homework reviews provided by the two teachers. Based on the video recordings of the homework reviews, there was a significant difference in the way they reviewed homework. Teacher B spent significantly more time reviewing more problems (mean = 19.8 minutes, 4.5 problems) than Teacher A (mean = 8 minutes, 3.5 problems). Teacher B also spent time reading the problems and answers to the students while Teacher A did not read the problems. We had no reason to believe that the quality of feedback provided by the two teachers was significantly different, but given the differences in the review methods we decided to test for the effect of teacher.

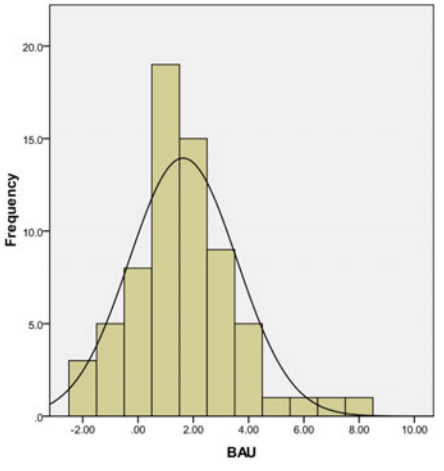

Fig. 1a. Distribution of gain scores for BAU

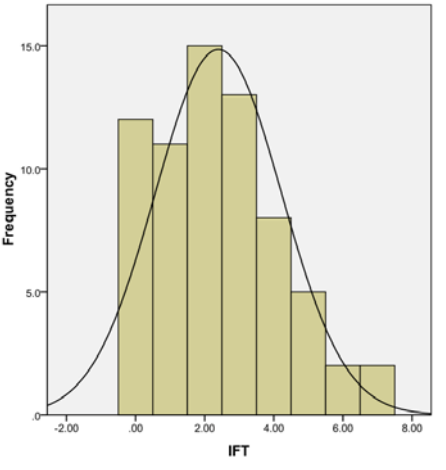

Fig. 1b. Distribution of gain scores for IFT

Comparing gain scores of students of Teacher A and Teacher B showed that the average gain for students who had Teacher B (M = 2.44, SD = 1.76) was higher than the average gain for students of Teacher A (M = 1.92, SD = 2.03). This could be due to the amount of time spent by Teacher B on reviewing homework problems the next day in class. However, a one-way ANOVA with Teacher as a factor showed that the difference was not significantly reliable, $F(1, 118) = 2.150$, $p = 0.145$, based on Teacher. But it did seem reasonable to keep in the model as we know from the review sessions that the two teachers spent different amounts of time going over problems. We continue our analysis considering Teacher as a potential factor in the model.

The results of a two-way ANOVA with Teacher and Condition as factors showed Condition to be a reliable factor, $F(1, 118)=8.27$, $p=0.005$, and the Teacher by Condition was also significantly reliable, $F(1, 118)=8.38$, $p=0.005$, in predicting post-test gains. Also, while Teacher A's students in the *IFT* condition had higher mean scores they are not reliably higher than Teacher B's students in the *IFT* condition.

The difference in mean scores between the *BAU* and *IFT* conditions for Teacher B suggest that delayed but quality feedback from teachers can make *BAU* homework as effective as the *IFT* homework if they spend a lot of time going over the questions.

Discussion. The positive results obtained from this experiment certainly reinforced the observation that computer-supported homework can produce superior results to more traditional approaches. The observed effect was smaller than the effect size reported by Mendicino et al., which was 0.61. The analysis based on Teacher by Condition showed that teachers may be able to make delayed feedback as effective as immediate feedback with tutoring. However, to do this Teacher B spent significantly more time than Teacher A reviewing the homework in class. The marginal gains for the two teachers were not significantly different. This seems to suggest that an effective strategy would be to give computer-based homework with tutoring and utilize the homework review time more effectively, perhaps going over new material.

Given the positive gains for students in the IFT condition, it seemed reasonable to examine the effects of immediate feedback and tutoring separately. Thus Experiments 2A and 2B look at student gains with immediate feedback with and without tutoring.

2.2 Experiment-2A

This experiment was conducted to analyze the effects of tutoring over immediate feedback. The two conditions for this experiment were *Tutoring* and *No Tutoring*. In the *No Tutoring* condition no tutoring is provided and the students are only given feedback on the correctness of their answers and provided with the right answer if they answered incorrectly using Correctness Mode problems. In the *Tutoring* condition students could ask for up to 3-4 hints on solving the problem before being presented with the final answer. With this experiment we hoped to understand the size of the effect with respect to Tutoring while controlling for the timeliness of feedback.

Experimental Design. The students were assigned four homework assignments. The homework assignments were completed by the students using the ASSISTment system at home. The students were randomly placed into either the *Tutoring* condition or the *No Tutoring* condition by the system. The homework assignments were designed such that the first half of each of the four assignments could be treated as the pre-test and the second half would act as the post-test for the experiment.

The content for this experiment consisted of problems that required the use of the Pythagorean Theorem to find the lengths of sides of triangles or deduce the area of geometric figures. The problems were similar to problems from the Connected Mathematics Project – "Looking for Pythagoras" unit.

Procedure. The setting and participants were the same as Experiment-1. The students were familiar with Correctness and Tutoring Mode in ASSISTments. The homework was assigned as a review after students were done with their regular bookwork. The students were not told that the assignment included pre- and post-tests.

Results. There were 72 students who finished at least one out of the four homework assignments. Out of the 72 students, 32 students were placed in each condition at least once over the period of the four homework assignments. The gain score in a condition

for each student was calculated to be the average gain score for homework assignments completed by the student in that condition. The average gain score for the Tutoring group was 1.0 (N = 32, SD = 1.16) and the average gain score for the No Tutoring group was 0.4 (N = 32, SD = 1.16). The analysis for comparing the gain scores in the two test conditions showed a reliably significant difference between the two conditions, $t(31) = -2.178$, $p = 0.037$, in favor of *Tutoring* with an effect size of 0.54. The 95% confidence interval of the observed effect size was (0.14 – 0.95).

Discussion. The effect observed in the direction of *Tutoring* indicates that providing tutoring for students does significantly improve their learning. These results suggest that tutoring homework is more effective than immediate feedback alone. However, the effect size (0.54) is higher than that observed in Experiment-1 in favor of *IFT* over *BAU*. We expected the effect to be smaller when controlling for immediate feedback.

2.3 Experiment-2B

After observing a surprisingly large effect size in Experiment-2A, the purpose of Experiment 2B was to see if the result could be replicated.

Experimental Design. The students were randomly placed in the two experiment groups based on their last names. The structure of the homework assignment was similar to the ones used in Experiment-2A but contained more problems.

The problems assigned for this experiment dealt with exponential and linear growth rates. The problems were similar to problems from the Connected Mathematics - "Growing Growing Growing" unit and were used as a review.

Procedure. The procedure was the same as Experiment-2A.

Results. Out of the 172 students, 20 students did not start the homework assignment. Three students, two placed in *Tutoring* and one placed in *No Tutoring* started but did not complete the assignment. The remaining 149 students completed the assignment. We excluded students who received perfect scores on the pretest, which left us with 107 students for our analysis. Overall, the average gain score was 0.80 (SD = 1.48) and the overall scores were reliably different than zero, $t(106) = 5.61$, $p < 0.001$.

The average gain score for the Tutoring group was 1.16 (N=64, SD=1.26) and the average gain score for the No Tutoring group was 0.28 (N=43, SD=1.6). When comparing the gain scores in the two conditions a reliable difference in favor of *Tutoring* was observed, $t(74.34)=2.97$, $p=0.004$. The observed effect size was 0.54 with a 95% confidence interval of (0.22–1.01).

Discussion. Upon replication of Experiment-2A, we found that the size of the effect was comparable to that observed in Experiment-2A. This indicates that tutoring had a large impact on student gains and that it is more beneficial to have tutoring in addition to immediate feedback. Most math textbooks provide answers to selected problems that can serve as immediate feedback for homework, but students often do not get immediate tutoring. This suggests that learning can be significantly improved from computer-based homework by providing immediate feedback with tutoring.

3 Conclusions and Contributions

The experiments presented in this paper help strengthen the claims made by Mendicino et al. [3], while improving the experiment design. The results suggest that spending proportionately more time in class going over homework can make learning equivalent to giving computer-based homework. Furthermore, the results show an advantage of tutoring when controlling for immediate feedback.

Strictly speaking, we did not look at the amount of learning in the *IFT* condition if the Teacher *did not* give a review the next day. It would be interesting to see if the *IFT* condition results stay as high as they do if the teacher did not go over the homework at all the next day. But, it seems that there is some value in going over the homework in class as seen by the strong gain scores for students in the *BAU* condition for Teacher B. We propose a future study that tests the value of homework review after receiving *IFT*. If the value of reviewing homework is small then a cost-benefit analysis should be considered to see if time is better used to introduce new content.

We did a survey of six curriculum supervisors from different towns and asked them "What is the appropriate amount of time for teachers to spend going over homework?" We got the following answers: 8-10, 10, 10, 10, 5-10 and 10-15 minutes. Based on these responses, we assume that 10 minutes is the right amount of time teachers should be spending on average going over homework. If we assume that students are spending 20 minutes doing their homework and their class period is effectively 40 minutes long, they have a total "math" time of 60 minutes combining homework and class. If students spend 20 minutes doing homework and 10 minutes reviewing homework in class, they are spending half of their "math" time on homework. This amount of time for review is probably significant for helping students learn, but also means that a better method can have an impact of practical significance. If we can improve homework by even half a standard deviation, it would be reasonable to see if we can improve student performance on state tests.

Our results reinforced the observation [3] that computer-supported homework is better for students compared to traditional homework approaches. It can be claimed that detailed scaffolding and hints can improve homework performance significantly. Systems such as ASSISTments can provide the necessary tools for improving homework by providing quality tutoring and immediate feedback and allowing teachers to identify areas in which students are struggling at an individual and class level, advantages that are much harder to achieve with traditional homework.

Acknowledgments. This material is based upon work supported by the National Science Foundation (NSF), under Grant DGE-0742503 and under Grant #0937060 to the Computing Research Association (CRA) for the CIFellows Project. For the full list of over a dozen funders of ASSISTments please see http://www.webcitation.org/5xp605MwY.

References

1. Bebell, D.: Technology promoting student excellence: An investigation of the 1st year of 1:1 computing in New Hampshire middle schools. In: Technology and Assessment Study Collaborative, Boston College (2005)

2. Walberg, H.J., Paschal, R.A., Weinstein, T.: Homework's powerful effects on learning. Educational Leadership 1995, 76–79 (1985)
3. Mendicino, M., Razzaq, L., Heffernan, N.T.: Comparison of Traditional Homework with Computer Supported Homework. Journal of Research on Technology in Education 41(3), 331–359 (2009)
4. Bonham, S.W., Deardorff, D.L., Beichner, R.J.: Comparison of student performance using Web- and paper-based homework in college-level physics. Journal of Research in Science Teaching 40(10), 1050–1071 (2003)
5. Razzaq, L., Heffernan, N., Koedinger, K., Feng, M., Nuzzo-Jones, G., Junker, B., Macasek, M., Rasmussen, K., Turner, T., Walonoski, J.: Blending Assessment and Instructional Assistance. In: Nedjah, N., de Macedo Mourelle, L., Neto Borges, M., Nunesde Almeida, N. (eds.). Intelligent Educational Machines within the Intelligent Systems Engineering Book Series, pp. 23–49. Springer, Heidelberg (2007)
6. What Works Clearinghouse Standard of Evidence for Reviewing Studies, Institute of Education Sciences, `http://ies.ed.gov/ncee/wwc/pdf/wwc_version1_standards.pdf` (retrieved May 2010)