

# IMPROVING STUDENT LEARNING IN MATH THROUGH WEB-BASED HOMEWORK

Kim Kelly<sup>1</sup>, Neil Heffernan<sup>1</sup>, Cristina Heffernan<sup>1</sup>,  
Susan Goldman<sup>2</sup>, James Pellegrino<sup>2</sup>, Deena Soffer-Goldstein<sup>2</sup>

<sup>1</sup>Worcester Polytechnic Institute, <sup>2</sup>University of Illinois Chicago

*Much debate surrounds the effectiveness of the common educational practice of homework (Cooper et al., 2006). A randomized-controlled trial has shown that using a web-based homework system that provides immediate feedback to students, while they are doing their mathematics homework, and detailed item reports to teachers significantly improves student learning. The use of that data also changed the homework review process, leading to a more comprehensive and meaningful review of student errors and misconceptions.*

## INTRODUCTION

Like much of the research in education that has focused on improving student learning, the present study examines the role technology can play in increasing student performance through homework. The common educational practice of homework has been criticized as Cooper et al. (2006) highlight the point that poorly conceived homework does not help learning. However, if we leverage technology to provide immediate feedback while students complete their mathematics homework, can we improve student learning?

Several studies have shown the effectiveness of intelligent tutoring systems (ITS) when used in the classroom (Singh et al. 2011). However, very few studies have explored the effectiveness of ITS when used as homework. Therefore it was very encouraging when Van Lehn et al. (2005) presented favorable results when ANDES, an ITS, was used in this fashion. Yet, most systems are not currently designed to be used for nightly homework. Computer aided instruction (CAI), which gives all students the same questions with immediate end-of-question feedback, is more applicable than complex ITS for nightly homework as teachers can easily build the content from textbook questions or worksheets. Kulik and Kulik's (1991) meta-analysis reviewed CAI and reported a low effect size for simple computer based immediate feedback systems. However, these studies were not in the context of homework use and did not focus on how teachers use the data to respond to student performance. Web-based homework systems (WBH) like WebAssign ([www.webassign.com](http://www.webassign.com)) are commonly used in higher education. These systems are similar to web based computer aided instruction (CAI), providing students immediate feedback and reports to teachers. While VanLehn et al. (2011) reported on three such systems used at the higher education level for physics, there are no known studies at the K12 level that allow this contrast.

In this study we look to measure the effect on learning by comparing simple WBH to a traditional homework (TH) condition representing the type of practice that millions of students perform every night in America and probably around the world. Additionally, we explore how the teacher can use the data to modify and improve mathematics instruction.

The current study employed ASSISTments.org, a web-based intelligent tutoring system to provide “end-of- problem-correctness-only” feedback during homework in the WBH condition. The ASSISTments system was also used for the TH condition by further removing the correctness feedback thus emulating traditional paper and pencil homework assignments. ASSISTments is currently used by thousands of middle and high school students for nightly homework. Students can receive immediate feedback on the homework and the teachers can then access item reports detailing student performance. In the current study we were interested in examining the effects of teacher review of homework performance based on information derived from the ASSISTments system under each of the two different homework conditions. The goal was to estimate the additional effects of teacher-mediated homework review and feedback following each of the two homework practice conditions – TH and WBH – and also study differences in how teachers might approach homework review given variation in student performance following each type of homework practice.

## **EXPERIMENTAL DESIGN**

Participants were 63 seventh grade students, who were currently enrolled in an eighth grade math class, in a suburban middle school in Massachusetts. They completed the activities included in the study as part of their regular math class and homework. Students were assigned to conditions by blocking on prior performance in math class. This was done by ranking students based on their overall performance in ASSISTments prior to the start of the study. Matched pairs of students were randomly assigned to either the TH (n=33) or WBH (n=30) condition.

The study began with a pre-test that was administered at the start of class (see Kelly, 2012 for all study materials and data). This test consisted of five questions, each referring to a specific concept relating to negative exponents. Students were then given instruction on the current topic. That night, all students completed their homework using ASSISTments. The assignment was designed with three similar questions in a row or triplets. There were five triplets and five additional challenge questions that were added to maintain ecological validity for a total of twenty questions. Each triplet was morphologically similar to the questions on the pre-test.

Students in the WBH condition were given correctness-only feedback at the end of the problem. Specifically, they were told if their answer was correct or incorrect. If a student answered a question incorrectly, he/she was given unlimited opportunities to self-correct, or he/she could press the “show me the last hint” button to be given the answer. It is important to emphasize that this button did not provide a hint; instead it provided the correct response, which was required to proceed to the next question.

Students in the TH condition completed their homework using ASSISTments but were simply told that their answer was recorded but were not told if it was correct or not (it said “Answer recorded”). It is important to note that students in both conditions saw the exact same questions and both groups had to access a computer outside of school hours. The difference was the feedback received and the ability for students in the WBH condition to try multiple times before requesting the answer.

The following day all students took post-test1. This test consisted of five questions that were morphologically similar to the pre-test. The purpose of this post-test was to determine the benefit of feedback while doing their homework. At that point, students in the WBH condition left the room and completed an unrelated assignment. To mimic a common homework review practice, students in the TH condition were given the answers to the homework, time to check their work and the opportunity to ask questions. This process was videotaped and can be seen in Kelly (2012). After all of the questions were answered (approximately seven minutes) students in the TH condition left the room to complete the unrelated assignment and students in the WBH condition returned to class. The teacher used the item report, generated by ASSISTments to review the homework. Common wrong answers and misconceptions guided the discussion. This process was videoed and can be seen at Kelly (2012). The next day, all students took post-test2. This test was very similar to the other pre and post-test assessments as it consisted of five morphologically similar questions. The purpose of this test was to measure the value-added by the different in-class review methods.

## **RESULTS**

Several scores were derived from the data collected by the ASSISTments system. Student’s homework average was calculated based on the number of questions answered correctly on the first attempt divided by the total number of questions on the assignment (20 questions). A partial credit homework score accounted for the multiple attempts allowed in the WBH condition. Students were given full credit for answers, provided they did not ask the system for the response. The score was calculated by dividing the number of questions answered without being given the answer by the number of total questions on the homework assignment (20 questions). Time spent on homework was calculated using the problem log data generated in ASSISTments and is reported in minutes. Times per action are truncated at five minutes. Recall that the homework assignment was constructed using triplets. Learning gains within the triplets were computed by adding the points earned on the third question in each triplet and subtracting the sum of the points earned on the first question in each triplet.

### **Learning Gains from Homework**

One student, who was absent for the lesson, was excluded from the analysis (n=63). A t-test comparing the pre-test scores revealed that students were balanced at the start of the study ( $t(61)=0.29$ ,  $p=0.78$ ). However, an ANCOVA showed that students in the WBH condition reliably outperformed those in the TH condition on both post-test1

( $F(1,60)=4.14$ ,  $p=0.046$ ) and post-test2 ( $F(1,60)=5.92$ ,  $p=0.018$ ) when controlling for pre-test score. See Table 1 for means and standard deviations. If the difference was reliable a Hedge corrected effect size was computed using CEM (2013). The effect sizes do not take into account pretest. The key result for post-test2 of 0.56 effect size had a confidence interval of between 0.07 and 1.08.

	TH	WBH	<i>p</i> -value	Effect Size
Pre-Test	9% (17)	7% (14)	0.78	NA
Post-Test1	58% (27)	69% (21)	0.046*	0.52
Post-Test2	68% (26)	81% (22)	0.018*	0.56
HW Average	61% (20)	60% (15)	0.95	NA
Partial Credit HW Score	61% (20)	81% (18)	0.0001*	1.04
Time Spent (mins)	22.7 (9.6)	23.2(6.2)	0.96	NA
Learning Gains	0.03 (0.9)	1.73(1.1)	0.0001*	2.21

Table 1: Means, standard deviations (in parenthesis), and effect size for each measure by condition. \*Notes a reliable difference.

A comparison of homework average shows that students scored similarly ( $F(1,60)=0.004$ ,  $p=0.95$ ). An ANCOVA revealed that when calculating homework performance using the partial credit homework score, students in the WBH condition performed reliably better than those in the TH condition ( $F(1,60)=17.58$ ,  $p<0.0001$ ). This suggests that with unlimited attempts, students are able to self-correct, allowing them to outperform their counterparts. Similarly, comparing learning gains revealed that students with correctness feedback and unlimited attempts to self-correct learned reliably more while doing their homework ( $F(1,60)=45.72$ ,  $p<0.0001$ ).

A review of the item report further describes this difference in learning gains. As expected, students in the TH condition continued to repeat the same mistake each time the question was encountered resulting in three consecutive wrong responses. Conversely, students in the WBH condition may have repeated the mistake once or twice but rarely three times in a row, accounting for the learning.

The first thing that we want to point out is that students in the WBH condition had a significantly lower percentage correct on the first item. Presumably students in the WBH condition would use the hint button when they were not sure of the answer or were willing to guess as they had unlimited attempts. However, in the TH condition, there was no such button, therefore perhaps students were more likely to take other steps such as looking at class notes, asking a parent or calling a friend for help before responding.

The ability to attempt each question multiple times is unique to students in the WBH condition. We suggest that this feature may play an important role in the presented learning gains. While this specific feature was not empirically tested in this study, we can only speculate on its effect. However, it is important to note that students in the WBH condition had on average 49 attempts (standard deviation=24) to answer the 20-question homework assignment. The fewest attempts made by any student was 25

and the most was 140. The average number of times the answer was requested was 4 with a standard deviation of 3.5. This suggests that students in the WBH condition took advantage of the ability to try questions multiple times to learn the material without requesting the correct answer.

We were not expecting that correctness only feedback was going to be time efficient. In fact, students in both conditions spent the same amount of time to complete their homework ( $F(1,60)=0.002$ ,  $p=0.96$ ). However, it appears that the time spent was apportioned differently in the conditions. Specifically, the TH condition took longer to generate a first response, but the WBH condition took time making multiple attempts as well as requesting the answer. It seems that students in the TH group spend more time thinking about the problem but the WBH group can get the problem wrong, and then use their time to learn the content.

### **Learning Gains from Homework Review**

To address the second research question of the effectiveness of using the data to support homework review, a paired t-test revealed that students in both conditions did reliably better on post-test2 than on post-test1 ( $t(62)=3.87$ ,  $p<0.0001$ ). However, an ANCOVA revealed that when accounting for post-test1 scores, there is not a reliable difference by condition in the gains from post-test1 to post-test2 ( $F(1,60)=2.18$ ,  $p=0.15$ ). This suggests that both methods of reviewing the homework lead to substantially improved learning. Interestingly, the results indicate that TH feedback, while students complete homework (69% post-test1), is as effective as receiving no feedback and then having the teacher review of the homework (68% post-test2). This suggests that to save time, teachers may not even need to review the homework if students have access to web-based homework systems.

### **Observational Results**

In addition to examining the effects of immediate feedback on learning, this study explored the potential changes to the homework review process the following day in class. In the TH review, time was spent first on checking answers and then the teacher responded to students' questions. However, in the WBH review the teacher reviewed the item report in the morning to determine which questions needed to be reviewed in class. The item report shows individual student performance as well as class performance at the question level. Common wrong answers are also displayed for each question. The teacher noted that in triplet 2, students incorrectly applied a previously learned concept. Specifically, 39% of students initially got this type of question right (multiplying powers with coefficients and variables). However, learning took place as 68% got the next similar question right. It was therefore puzzling to see that on the third question in that triplet (question number 10), only 45% got the question right. Upon investigating the question, the teacher was able to identify the misconception and therefore addressed it with the class.

We designed the experiment with ecological validity in mind. That is, we wanted the teacher to naturally review the homework, giving students enough time to ask

questions. The hope was that approximately the same amount of time would be spent in each class and by each condition. We were disappointed to find that the classes and conditions varied greatly in the amount of time spent going over the homework. Half of the sections took over nine minutes to review the homework while two of the sections in the TH condition and one in the WBH condition spent substantially less time. This is a threat to the validity of drawing statistical inferences, but given the desire to maintain realistic homework review conditions, these inconsistencies highlight important differences in the homework review methods.

An observational analysis of the video recordings of the teacher reviewing the homework revealed that while the time spent in the WBH condition was often longer than the TH, it was also far more focused than in the TH. Specifically, when students were in the TH condition, on average 1 minute passed before any meaningful discussion took place. Whereas, when students were in the WBH condition, homework review began immediately with the teacher reviewing what she perceived to be the most important learning opportunities.

Other notable differences in the type of review include the number of questions answered. In the TH condition, 2 classes saw 3 questions each and one saw 7. However, in the WBH condition each class saw 4 targeted questions and 2 classes requested 1 additional question. The variation in question types also is important to note. The teacher was able to ensure that a variety of question types and mistakes were addressed whereas in the TH condition students tended to ask the same types of questions or even the same exact question that was already reviewed. Additionally, students in the TH condition also asked more general questions like “I think I may have gotten some of the multiplying ones wrong.” In one TH condition only multiplication questions were addressed when clearly division was also a weakness and similarly, another TH condition only asked questions about division. This accounts for much of the variability in overall review time.

In listening to the comments made by students it appears that the discussion in the TH condition was not as structured as the WBH condition. Not all students had their work and therefore couldn't participate in the review. One student said, “I forgot to write it down.” Another said, “I left my work at home.” Because students were asking questions and the teacher was answering them, we suspect that only the student who asked the question was truly engaged. In fact, one student said, “I was still checking and couldn't hear” which led to the teacher reviewing the same question twice. In the WBH condition, the teacher used the information in the report, such as percent correct and common wrong answers to engage the entire class in a discussion around misconceptions and the essential concepts from the previous question.

Other notable differences include the completeness of the review. In the TH condition, the review was dominated by student directed questions. This means that each class experienced a different review and the quality of that review was directly dependent on the engagement of the students. Conversely, in the WBH condition, all 3 classes were presented with the same 4 troublesome questions and common mistakes. Additional

questions were reviewed when asked (as in two sections) but the essential questions as determined by the data in the item report were covered in all three sections.

### **Student Survey Results**

Following participation in this study, students were questioned about their opinions. We want to acknowledge that students might have been telling the teacher what she wanted to hear: the whole classroom of students had been using ASSISTments for months and the teacher had told them on multiple occasions why it's good for them to get immediate feedback. So with that caveat, we share the following results. 86% of students answered ASSISTments to the question "Do you prefer to do your homework on ASSISTments or a worksheet?". 66% mistakenly think that it takes longer to complete their homework when using ASSISTments (we showed in this study that that was not the case) and 44% feel that they get frustrated when using ASSISTments to complete their homework. However 73% say that their time is better spent using ASSISTments for their homework than a worksheet. When asked what students like best about ASSISTments, student responses included:

"That if you get stuck on a problem that it will give you the answer."

"You can redo your answer if you get it wrong and learn from your mistakes."

"How it tells you immediately that you are right or wrong."

"I like how I know if I'm right or wrong. This helps because often times when I get things wrong I just go back to my work and I see what I'm doing wrong which helps me when doing other problems."

While the learning benefits are profound and students prefer a web-based system, there is a sense of frustration that must still be addressed. Specifically, student feedback suggests that students appreciate the features of intelligent tutoring systems, including hints, worked examples and scaffolding. Therefore, future studies should explore adding additional feedback to determine if added AIED features improve learning or if maybe learning requires some levels of frustration. All of the survey results are made available without names, including students' comments at <http://www.webcitation.org/6DzciCGXm>.

### **DISCUSSION**

This paper's contribution to the literature is exploring the potential use of ITS for mathematics homework support. Used as designed, ITS are somewhat cumbersome for teachers to use for homework as the content is not customizable. However, if ITS were simplified they could be used like web-based homework systems, providing correctness feedback to students and reports to teachers. This begs the question, is correctness only feedback enough to improve the efficacy of homework and what effect does teacher access to reports have on homework review? This randomized controlled study suggests that simple correctness-only feedback for homework substantially improves learning from homework. The benefit of teachers having the data to do a more effective homework review was in the expected direction (but not

reliable). But taken together (immediate feedback at night and an arguably smarter homework review driven by the data) the effect size of 0.56 seems much closer to the effect of complex ITS. Of course the large 95% confidence interval of [0.07 to 1.08] tells us we need more studies.

Future studies can explore features of other web-based homework systems like Kahn Academy to determine which aspects of the systems are particularly effective. Incrementally adding tutoring features to determine the effectiveness of each feature would also be valuable. Finally, the role of data in formative assessment should be further explored. In what way can teachers use the data to improve homework and review and instruction?

In this fast-paced educational world, it is important to ensure that time spent in class and on mathematics homework is as beneficial as possible. This study provides some strong evidence that web-based homework systems that provides correctness-only feedback are useful tools to improve mathematics learning without additional time.

### **Acknowledgement**

Funding for this work was provided by US Dept of Ed: IES grant # R305C100024.

### **References**

- CEM. (2013). Effect size calculator [software]. Retrieved from <http://www.cemcentre.org/evidence-based-education/effect-size-calculator>
- Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987-2003. *Review of Educational Research*, 76(1), 1-62.
- Kelly, K. (2012). *Study materials*. Retrieved from <http://www.webcitation.org/6E03PhjrP>
- Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior*, 7, 75-94.
- Singh, R., Saleem, M., Pradhan, P., Heffernan, C., Heffernan, N., Razzaq, L., ... Mulchay, C. (2011). Feedback during web-based homework: the role of hints. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of the Artificial Intelligence in Education Conference 2011* (Vol. 6738, pp. 328-336). Berlin, DE: Springer.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- VanLehn, K., Lynch, C., Schulze, K. Shapiro, J. A., Shelby, R., Taylor, L., ... Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*, 15(3), 1-47.