

# Does Immediate Feedback While Doing Homework Improve Learning?

Paul Kehrer, Kim Kelly & Neil Heffernan

Worcester Polytechnic Institute  
nth@wpi.edu

## Abstract

Much of the literature surrounding the effectiveness of intelligent tutoring systems has focused on the type of feedback students receive. Current research suggests that the timing of feedback also plays a role in improved learning. Some researchers have shown that delaying feedback might lead to a “desirable difficulty”, where students’ performance while practicing is lower, but they in fact learn more. Others using Cognitive Tutors have suggested delaying feedback is bad, but those students were using a system that gave detailed assistance. Many web-based homework systems give only correctness feedback (e.g. web-assign). Should such systems give immediate feedback or might it be better for that feedback to be delayed? It is hypothesized that *immediate* feedback will lead to better learning than *delayed* feedback. In a randomized controlled crossover-“within-subjects” design, 61 seventh grade math students participated. In one condition students received correctness feedback immediately, while doing their homework, while in the other condition, the exact same feedback was delayed, to when they checked their homework the next day in class. The results show that when given feedback immediately students learned more than when receiving the same feedback delayed.

## Introduction

The field of Intelligent Tutoring Systems (ITS) has had a long history (Anderson et al. 1995, Koedinger et al. 1997, Corbett et al. 1997). Recently, Kurt VanLehn (2011) claims that ITS can be nearly as effective as human tutors. VanLehn also concludes that Computer Aided Instruction (CAI) is not as effective as ITS. The distinction is the type and granularity of feedback provided. ITS provide fine-grained, detailed and specific feedback and tutoring often at the step or sub-step level. In contrast, CAI provides immediate feedback on the answer only. The focus of this past research has predominately been the use of these systems in the classroom not as homework support.

Some studies have shown the effectiveness of ITS used in the context of a Web-Based Homework Support

(WBHS) (Mendicino et al. 2009, Singh, 2011, Bonham et al. 2003). Similarly, VanLehn et al. (2005) have shown significant learning gains in students using the Andes Physics tutoring system in place of traditional homework. However, these learning gains are the result of sophisticated feedback rather than correctness-only feedback. Kelly et al. (submitted) shows that correctness-only feedback and unlimited attempts to self-correct result in significant learning gains compared to no feedback at all. However, this feedback was provided while students were completing their homework. Does immediacy of this type of feedback matter?

## Timing of Feedback

In addition to the type of feedback affecting efficacy, timing of feedback has also been studied. Shute (2007) summarizes the inconsistencies in the research on immediate versus delayed feedback and concludes that both types of feedback have pros and cons. Much of the research cited in her analysis was conducted in laboratory settings or within the context of a classroom. However, the reality is that students in America are given homework every night and traditionally receive feedback the following day. ITS, as WBHS, provide an opportunity for students to receive feedback immediately, while doing their homework instead of waiting. But does this immediacy of feedback impact learning in the unique case of homework?

The current research question is, do students learn more when they are getting correctness feedback as they work on their homework than when they get the same feedback the next day. Given that the quality of the feedback is lacking compared to previous studies, one might wonder is it critical for students to receive feedback immediately. We seek to determine if there is a difference in learning gains, but also how large an effect does the immediacy of feedback have when used in a real educational setting?

## Current Study

The present study used, ASSISTments, an intelligent tutoring system, which is capable of providing scaffolding and tutoring. Because this study focuses on the effectiveness of *correctness only* feedback, tutoring features were turned off.

### Experimental Design

A total of 65 seventh grade students in a suburban middle school in MA participated in this study as part of their regular math homework and Pre-Algebra math class. The topics covered during this study included surface area and volume of 3-dimensional figures.

A pre-test was administered for each topic. The pre-test consisted of one question for each sub-topic included in the lesson. For example, the lesson on surface area of 3-dimensional figures actually had four sub-topics that were being taught: surface area of a pyramid, surface area of a cone given the slant height, surface area of a cone given the height, and surface area of a sphere. The lesson on volume of 3-dimensional figures had five sub-topics, which included: volume of a pyramid, volume of a cone, volume of a sphere, volume of a compound figure, and given the volume of a figure find the missing value of a side. All of the study materials including the data can be found in Kelly (2012).

The accompanying homework assignments were completed using ASSISTments, a web-based tutoring system. Students were accustomed to using the program for nightly homework. The homework was designed using triplets, or groups of the 3 questions that were morphologically similar to the questions on the pre-test. There were three questions in a row for each of the primary topics. Additional challenge questions relating to the topic were also included in the homework to maintain ecological validity.

Post-tests for each topic were also administered. There was one question for each sub-topic and they were morphologically similar to the questions on the pre-test and homework assignments. Therefore, the tests on surface area had four questions while the tests on volume had five.

### Procedure

Students were blocked based on prior knowledge into two conditions, *immediate feedback* and *delayed feedback*. To do this, overall performance in ASSISTments was used to rank students. Pairs of students were taken and each was randomly assigned to either of the conditions. Students in the *immediate feedback* condition were given correctness feedback immediately on each question as they completed their homework. Students in the *delayed feedback* condition completed their homework on a worksheet but were given the same feedback the next day.

At the start of the study, all students were pre-tested in class, which was part of the typical routine in this classroom. They were then formally instructed on surface area of 3-dimensional figures. That night, they completed a related homework assignment. Students in the *delayed feedback* condition completed their assignment on a worksheet, receiving NO feedback. Students in the *immediate feedback* condition completed their homework using ASSISTments, which immediately told if their response was correct. In the case of an incorrect response, students were given unlimited attempts to correct their answer. A correct response was required to move on to the next question. Therefore, students could ask for the correct response if needed by pressing the "Show hint 1 of 1" button. It is important to note that when tutoring features are active, this button would provide a hint. However, to explore correctness-only feedback, this button provided the correct response.

The following day, all students reviewed their homework. Students in the *delayed feedback* condition used ASSISTments to enter their answers from their worksheet, providing them the same correctness-only feedback and unlimited attempts to self-correct that was given in the experimental condition. Students in the *immediate feedback* condition reviewed their responses using the item report in ASSISTments. The item report shows students which questions they answered incorrectly and what response they initially gave. They were encouraged to look back over their responses and work. To end class, all students were then given a post-test on surface area of 3-dimensional figures.

The study was replicated the following week with students switching conditions and with a new topic. Again, students were pre-tested during class and formally instructed on volume of 3-dimensional figures. That night, students completed their homework in the opposite condition. Specifically, students who had received *immediate feedback* now completed the homework on a worksheet, without feedback and those who had received *delayed feedback* now used ASSISTments to receive feedback immediately. The next day, in class, students reviewed their homework. Students in the *delayed feedback* condition used ASSISTments to receive correctness feedback and those in the *immediate feedback* condition used the ASSISTments item report to review incorrect responses. A post-test was then given.

## Results

Data from 61 students were included in the data analysis. Students were excluded from the analysis if they were absent for any part of the study ( $n=4$ ). A two-tailed t-test analysis of the pre-tests showed that students were evenly distributed for both assignments. (Surface Area: *Immediate Feedback*  $M=14$ ,  $SD=16$ , *Delayed Feedback*  $M=13$ ,

$SD=13$ ,  $p=0.82$ . Volume: *Immediate Feedback*  $M=3$ ,  $SD=0.7$ , *Delayed Feedback*  $M=5$ ,  $SD=0.9$ ,  $p=0.25$ ).

While between subject analysis are common, this study was conducted to provide a within subject analysis. Results showed that when students received immediate feedback ( $M=60$ ,  $SD=27$ ) they performed better than when receiving the same feedback delayed ( $M=51$ ,  $SD=30$ ), however this difference is only marginally significant ( $t(60)=2.1$ ,  $p=0.057$ ).

However, a paired t-test analysis of the pre-test scores shows that students had significantly more background knowledge of Surface Area ( $M=14$ ,  $SD=14$ ) than Volume ( $M=4$ ,  $SD=8$ )  $t(60)=3.9$ ,  $p<.0001$ ) Therefore, relative gain scores were calculated and analyzed to determine if there was in fact increased learning as a result of immediate feedback when the potential for growth was accounted for. To calculate the relative gain score, for each student, we took his/her gain score and divided by the possible number of points they could have gained (total number of questions – pretest score). For example, if a student scored 1 correct on the pre-test out of 5 questions, and later scored 3 on the post-test, the relative gain score was  $((3-1)/4)=50\%$ . We had one student with a negative gain score, (she had one correct on the pretest, but then zero correct on the post-test, and the resulting negative score was included).

A paired t-test of relative gains shows that students learned 12% more when given immediate feedback ( $M=67\%$ ,  $SD=26$ ), than delayed feedback ( $M=55\%$ ,  $SD=32$ ), ( $t(60)=2.501$ ,  $p=0.015$ ). The effect size is 0.37 with a 95% confidence interval of 0.05 to 0.77.

We were curious to know if the effect of condition were experienced similarly across the two math topics, so we compared both the post-tests alone, their absolute gain scores, and their relative gain scores, and found similar patterns of Immediate Feedback being more effective than delayed, but with an expected, lower level of significance that was no longer reliable. See Table 2.

## Contributions and Discussion

This study adds to the delayed versus immediate feedback debate by exploring a critical context that has been ignored in previous research. Specifically, immediate feedback, while students complete homework leads to better learning than waiting until the next day to receive that same feedback. This is an extremely important situation to consider as it applies to almost every student in America. While further research is needed comparing different types of feedback, assessment, and control conditions, this study moves the debate in a new direction with respect to delay time.

### Discussion and Future Research:

There has been some controversy about whether and when immediate feedback is good especially surrounding

performance on immediate tasks versus delayed assessments. However, in the context of homework support, the goal is immediate learning gains that prepare the student for the next lesson. In this ecologically valid setting, it's very difficult to measure retention or to deliver a valid delayed assessment because other learning occurs after the intervention.

Table 2: Mean and standard deviation (in parenthesis) post-test scores, absolute gain scores and relative gain scores for both topics. Effect sizes and significance levels included.

	Immediate Feedback	Delayed Feedback	Effect Size & p-value
Surface Area:			
Post Test	77% (22)	66% (29)	0.37, p=0.11
Absolute Gains	64% (26)	53% (31)	0.36, p=0.14
Relative Gains	61% (26)	50% (29)	0.38, p=0.13
Volume:			
Post Test	63% (24)	54% (27)	0.37, p=0.14
Absolute Gains	61% (23)	48% (29)	0.42, p=0.08
Relative Gains	61% (26)	50% (29)	0.39, p=0.13

One area that should be explored further with this “overnight delay” is task transfer. For instance, according to Lee (1992) immediate feedback did worse than delayed on far transfer tasks. The lack of self-correction and error analysis was attributed with these findings. Similarly, Mathan (2003) argued, “feedback could prevent important secondary skills from being exercised.” These secondary skills include error detection, error correction and metacognitive skills. The author discusses the need to “check your work”. However, middle school students are still learning how to check their work and what it means to find errors. Quite often strong students aren't even aware they made a mistake unless it's pointed out. Similarly, many students don't know what it means to check their work. We would argue that providing correctness only feedback actually promotes these skills because it requires students to self-correct in order to move on. They are responsible for detecting their error and correcting it. Additionally, students begin to recognize the types of errors they make repeatedly and learn to check specifically for those.

Merrill et al. (1995) argues that a benefit of human tutors is that they do not intervene when learning might occur through the mistake. In the present study, the timing of the feedback allows students to make that mistake and like a human tutor simply tell the students that the answer is not quite right. Students must then detect their error and correct it, much like in the experience with a human tutor.

The results of the current study largely support our hypothesis that *immediate feedback* does improve learning compared to *delayed feedback*. As expected, students who were told if their answers were correct and were able to fix them as they completed their homework, learned more than students who completed the homework on a worksheet and

