

STUDENT LEARNING IN A RAPID FEEDBACK ENVIRONMENT

Nicholas Ball, Utah Valley University, Nicholas.ball@uvu.edu
Gove Allen, Brigham Young University, gove@byu.edu

ABSTRACT

The research presented in the paper examines how instructor policies related to the use of feedback in a university course can improve student learning. Using MS Excel assignments and a rapid feedback computerized grading system, we examine two specific course policies. The first of these is whether to allow students more than one attempt at the same assessment, using feedback from the first attempt to inform student work on subsequent attempts. The second involves whether the highest or the average score across multiple attempts is recorded in the grade book. We find that the student learning, measured as performance on a comprehensive final exam, improves when students are allowed multiple attempts on other course assignments. Additionally we find that student learning improves when the instructor records the average score instead of the highest score in course grade book for assignments with multiple attempts.

Keywords: Student Learning, Rapid Feedback, Assessment, MS Excel

INTRODUCTION

Determining how to best design a course to maximize student learning can be a difficult undertaking. In addition to selecting the correct set of course objectives and materials and activities for meeting those objectives, instructors must effectively assess student learning and provide meaningful feedback to students on their performance. Often large class sizes or inadequate instructor resources compound these difficulties. While there is not a silver bullet for all classroom situations, computer-aided assessment holds promise for improve assessment while reducing the workload for instructors.

One hallmark of computer-aided assessment is the benefit of rapid feedback for students. In many cases, feedback is delivered to students within a matter of moments after they complete an assessment. With computer-generated rapid feedback, instructors are provided with course policy options regarding how students are allowed to use that feedback. These options are not all equal with respect to student learning. In this paper, we examine how course policy regarding student feedback can be set to enhance student learning.

LITERATURE REVIEW

Two important bodies of literature are relevant to this research. The first of these is the well-established body of literature related to feedback and learning. The second of these is the growing body of literature related to computer-aided assessment.

While a comprehensive review of the literature related to feedback and learning is beyond the scope of this work, several studies inform the hypotheses developed in this research. Most of this research highlights the beneficial effects of feedback. For example, Ilgen, et al., [3] find that individuals prefer specific and timely feedback. Specific feedback can be particularly useful because it provides more meaningful information about performance than vague feedback. Additionally, it allows individuals to develop specific goals and strategies for improving their performance [3]. Liden and Mitchell [5] provide compelling evidence that specific and timely feedback lead to improved individual performance. In a classroom setting, timely and specific feedback is important to student learning.

The prior literature also highlights different types of feedback. Feedback can be classified as summative, formative, and diagnostic [1]. Summative feedback includes assessments of individual performance. Formative feedback involves assessments of the processes individuals follow to complete a task, and diagnostic feedback is used for tutors or individuals to gain a snapshot of the level of mastery an individual has achieved at a given point in time. Of

the three, formative feedback is generally the most helpful for improving individual performance [4], because it provides the opportunity for individuals to examine not just the outcome of their work, but how they are completing their work [3]. To the extent that students can incorporate feedback as part of their learning process (rather than just a summative assessment of classroom performance), student learning will increase.

The research related to computer-aided assessment is a growing area. The research in this area that is most relevant to this work examines how computer aided assessment overcomes disadvantages of traditional assessment techniques [6]. As we discuss in the Atlas and Prometheus section of this paper below, the computer-aided assessment technology described in this paper was developed to overcome problems with traditional assessment in Excel. Computer-aided assessment can provide detailed, timely feedback to students and create forms of assessment that are not possible without technology, such as adaptive assessments [6]. Additionally, computer aided assessment eases resource constraints of instructors, facilitates consistent assessment, and enhances the defensibility of grading [2]. Gibbs and Simpson [2] also suggest that computer-aided assessment will computer-aided assessment can give rise to innovative assessment strategies.

In this study, we use a computer-aided assessment engine to provide students with detailed immediate feedback on their work. This feedback is used to help test two hypotheses. First, learning is most enhanced by formative (process-centric) feedback. Feedback becomes formative when it can be used to improve the process followed to complete a task. Our grading technology provides students with detailed feedback in a short time window. This will facilitate, if course policy allows, students to complete an assessment more than once using feedback from a prior attempt to improve how they complete subsequent attempts – making the feedback formative in nature.

H1: Student learning will be greater when course policy allows students to complete an assessment more than once, using feedback from prior submissions rather than when course policy allows only one submission.

Of course, providing students with more than one attempt to complete an assignment can be problematic. Common concerns with allowing students to complete an assessment more than once include: the potential for grade inflation, worries that this will reduce the rigor of the course, and timing constraints that may limit the amount of content that can be covered during the course. Again computer-aided assessment can ameliorate timing concerns. Moreover, course policy can address other concerns raised by instructors. For example, an instructor may choose to record the average score rather than the highest score in the course grade book when a student completes an assessment multiple times. This could at least partially address grade inflation and rigor concerns. It may also be the most advantageous for students.

H2: Student learning will be greater when course policy prescribes that the average score be recorded in the course grade book than if the highest score is recorded for multiple attempts.

Atlas and Prometheus

Underlying the ability to assess hypotheses suggested in this paper is the ability to provide detailed and immediate feedback to students. To accomplish this, we have developed a novel approach to assessing student work on Microsoft (MS) Excel assignments. Excel is the most commonly taught electronic spreadsheet package in the world. Courses that with significant Excel content are often offered at many universities as a general education or business core requirement.

Some topics that are taught broadly at universities lend themselves to “textbook, lecture, quiz, and exam” model that makes teaching large sections possible. However, Excel, is much more hands-on in nature and is not effectively taught using this model. Schools have typically taken two approaches to delivering this content. The traditional avenue to teach this material involves small sections with assignments requiring students to become familiar with the software through direct use. Teaching small sections with tenure-track faculty is expensive and using adjunct and teaching assistants leads to variations in quality that are undesirable in what may be the student’s first exposure to a particular department. Perhaps more importantly, the hand grading of student work leads to long feedback cycles and imprecise application of rubrics.

In recent years, another option has become popular. The advent of simulation software from text publishers such as Pearson, McGraw-Hill, and Cengage allows students to get rapid feedback on performance but they require that students work not in the real application they are learning, but in a facsimile that reproduces only a portion of the main product's functionality. Our experience indicates that students find both of these approaches dissatisfying.

We wanted an approach that: (1) provides rich, meaningful assignments for students to solve by working directly in Excel, (2) gives detailed feedback instantly, (3) allows students learn from the feedback to improve their skills and submit again, (4) alerts us when students are working together on assignments, (5) presents detailed analytics on student engagement and performance, (6) delivers content in both textual and video presentations, (7) allows students multi-year access to the instructional content, (8) lets us explore how students solve problems, not just see their final solution, (9) did not cost students an arm and a leg, and (10) supports Mac users as well as PC users. We realized that the only way we could get this system is if we built it ourselves—so we did. We call it Prometheus.

The Prometheus Approach™ leverages our understanding of feedback cycles to increase the effectiveness of student learning. It accomplishes this by providing step-by-step demonstration (in both video and text); hands-on, meaningful practice; rapid, detailed feedback on performance; and an instructor configurable assessment environment.

The Prometheus Approach™ is powered by the ATLAS Feedback Engine™, a patent-pending technology that delivers very detailed feedback to students in a matter of a few seconds. ATLAS (Automated Technology Learning Assessment System) requires no installation and works on both Microsoft and Apple operating systems. Because of its distributed architecture, we are able to handle vast numbers of simultaneous submissions (some classes have more than 2,500 students) without experiencing server-load bottlenecks – ensuring that students receive feedback within a few seconds of completing an assessment. Both of the studies described in this paper utilized this system to facilitate the assessment of student learning and deliver feedback to students on their performance.

CONTRIBUTIONS

This research makes at least two significant contributions to teaching. First, it highlights the effectiveness of a computer-aided assessment engine for Excel assignments, ATLAS. While not the focus of the research, the ATLAS grading engine was used to provide detailed and immediate feedback to the students studied in the paper. The fact that we demonstrate meaningful differences in student performance when instructors configure the ATLAS grading engine in different ways highlights the effectiveness of the assessment approaches facilitated by ATLAS.

The second, and main, contribution of this research highlights approaches to improving student learning that are achieved by manipulating how students are allowed to utilize the feedback provided to students and how students scores are recorded in the class grade book. Specifically, we find that student learning improves under two specific circumstances:

- (1) Students are presented with immediate and detailed feedback on their performance and then allowed to utilize that feedback to complete the assessed task a second time instead of presenting students with feedback on their performance without the opportunity to apply the feedback in a subsequent attempt at the same task.
- (2) When students are allowed to complete an assessment more than once, if the instructor records the average score across attempts rather than the highest score.

Together these findings support a powerful approach to the computer-aided assessment of Excel assignments that has a significant impact on student learning. We recommend that instructors provide students with two attempts at assignments and then record the average score across those attempts for that assessment in the grade book.

METHODOLOGY AND FINDINGS

To test our hypotheses we conducted two separate studies. Both studies involved college students taking an introductory Excel class. Each class required students to complete a set of five assignments in Excel to assess their mastery of the topics covered in the class. Each of the classes utilized the ATLAS grading system that provides

In both studies, policy changes by the instructors of the course facilitated the treatments we needed to test our hypotheses. In this way, both studies provided us with a natural experiment on the relationship between aspects of feedback and student learning. However, care was taken for both studies to account for any biases in the data that might result from the fact that students were not randomly assigned to the courses. The first study examined hypothesis 1 and the second study examined hypothesis 2.

Study 1: The Effect of Applying Feedback on Student Learning

The first study involved students in eight different sections of an introductory management information systems (MIS) class at the same university over the course of two consecutive semesters, spring and fall. Four sections of the courses were offered each semester. All sections of the course required that students complete each of the five untimed Excel assignments and a timed comprehensive final exam to assess their mastery of the topics covered in the class. Each student received detailed and immediate feedback about his or her performance for each assignment and the final exam.

During the first semester, students were allowed to submit each of the five assignments twice. The average score for each assignment was recorded in the grade book. The pedagogical reasoning behind the course policy was that students would learn from the feedback generated in the first attempt on an assignment. This learning could then be directly applied to fixing any mistakes made during the first attempt on the students' second attempt at completing each assignment. The instructors felt that allowing students to use the feedback achieved on the first attempt to complete the second attempt would reinforce any learning achieved from the feedback. Students were not allowed a second attempt of the comprehensive exam.

The instructors were concerned that allowing multiple attempts on the same assignment was artificially inflating their scores for the course. Between semesters they instituted a policy change. Students in the second semester were allowed only one attempt at each assignment, though they did still receive detailed and immediate feedback on their work.

The change in course policy provides us with a good setting to examine hypothesis 1. Does learning improve if students are allowed to apply feedback from completing a task to completing the same task again? To directly test this hypothesis, we examine student performance on the comprehensive final exam. We expect that students who took the course when they were allowed to complete each assignment a second time after receiving feedback from their first attempt to outperform students who were received only feedback on their performance for each assignment. The results of this study are reported in the **Study 1 Findings** section of this report.

Because students were not randomly assigned to the courses, we were concerned that there may be systematic differences between the students who took the course in each semester. These differences, if present, may heavily bias or confound our findings. To examine if there were systematic differences between students across the semesters, we analyzed student performance on each of the five assignments in the course.

We expected that students in the first semester, who were allowed to complete each assignment twice, would have higher overall scores on each assignment. This is the case. However, there is no reason to believe that these students would exhibit higher performance on their first attempt on these assignments than their counterparts in the other semester. We examined student scores on their first attempts at each assignment across the two semesters. If there were differences in student performance on the first attempts, we would conclude that systematic differences exist between the two student groups. Conversely, if there are not differences in student performance on the first attempts across the semesters, this would provide at least some evidence that no such systematic differences exist.

Table 1. Student Performance on First Attempts Across Semesters

| Assignment | Semester 1 Mean | Semester 2 Mean | Mean Difference | P-value (t-test) |
|------------------------|------------------------|------------------------|------------------------|-------------------------|
| Excel Basics | 95.07 | 93.58 | 1.49 | 0.0808 (NS) |
| Formulas and Functions | 95.91 | 93.76 | 2.15 | 0.0794 (NS) |
| Logic and Reference | 91.18 | 89.89 | 1.28 | 0.4919 (NS) |
| Time and Text | 97.03 | 95.79 | 1.24 | 0.1043 (NS) |

Charts 96.44 96.76 -0.32 0.7696 (NS)

Table 1 outlines the results of this validity check. The average scores for the first attempts in the first semester are compared to the average scores in the second semester. All assignments are scored out of 100 points. Five separate t-tests were conducted to examine if the calculated mean differences are statistically significant. We first tested for equality of variance to know which t-test to use. We found the variance to be unequal for each assignment across the two semesters. After we determined to use the t-test for two samples with unequal variance, we found no statistical difference between the student scores for any of the assignments across either of semesters.

Study 1 Findings

After concluding that there was not any evidence of systematic differences between the students across the semesters, we examined the possible effect of allowing students the opportunity to apply feedback on an assignment to complete second attempt on that assignment on student learning. Our measure of student learning for the semester is student performance on the final exam for the course. **Table 2** outlines this analysis.

Table 2. Student Performance on the Comprehensive Final Exam

| | Semester 1 | Semester 2 |
|--------------------------|------------------|------------|
| Number of Students | 123 | 109 |
| Average Final Exam Score | 279.37 | 267.66 |
| Variance | 1043.11 | 2328.70 |
| Mean Difference | 11.71 (3.9%) | |
| P-value (F-test) | 0.4663 (NS) | |
| P-value (t-test) | 0.0145* (2.1984) | |

In the first semester, 123 students completed the final exam. The average score for these students was 279.37. One hundred nine students complete the exam in the second semester with an average score of 267.66. There were 300 total points possible on the exam. Students in the first semester scored an average of 11.71 points (3.9%) higher on the final exam than did students in the second semester. In other words, students who were allowed to apply the feedback achieved on the first attempt scored 11.71 points higher on the final exam, on average, than students who received only the feedback.

We performed a t-test to determine if this raw difference in scores is statistically significant. We first calculated an F-statistic to determine if the exam scores across the semesters had equal variance. We found that the two samples had unequal variance. We determined to complete the t-test for two samples with unequal variance. The P-value for this t-test was calculated at 0.0145, which is statistically significant. We therefore conclude with students who are allowed to apply feedback exhibit higher levels of learning than those who receive only feedback, supporting hypothesis 1.

Study 2: Examining How Score Resolution Impacts Student Performance

Much like study 1, study 2 examines the effect of a course policy change between semesters. This study involves students at a different university. At this university, students used the same teaching materials used for study 1, including the same five homework assignments and the technology that provides the detailed immediate feedback to students on their performance. Students were not required to complete the comprehensive final exam for either of the courses at this university.

During the first semester of an introduction to Excel course, students were allowed two attempts at each assignment. The second attempt would occur after students received feedback from their first attempt. In the first semester course, the score recorded in the grade book for each assignment was the highest score achieved across the two attempts at the assignment.

In the second semester, students were also allowed to complete each assignment twice. The second attempt was also informed by the feedback achieved at the end of the first attempt. In this semester, the instructor was concerned that the course policy of the first semester was too generous and lead to grade inflation. Moreover, the instructor was

concerned the some students were not taking the first attempt seriously. Rather, some students would complete the first attempt, without engaging with the other learning materials for that unit of the course, to see what they could learn about the assignment rubric from the feedback received on the first attempt. They would then use that information on the second attempt. To combat both of these problems, the instructor implemented a change between semesters. This change involved recording the average score in the grade book across the two attempts for each assignment.

Just as in the first study, there is concern that since students were not randomly assigned to the courses, there may be systematic differences in students' prior level of Excel capability. Fortunately, the instructor at this university administered a pre-test to students during the first several days of each semester to determine where he needed to focus his teaching efforts during the semester. We used the pre-test scores from each semester to examine possible differences in students' Excel capabilities before coming into the class. **Table 3** highlights this analysis.

Table 3. Student Pre-test Scores for Each Semester

| Course | Semester 1 | Semester 2 |
|-------------------------|-------------|------------|
| Number of Students | 124 | 934 |
| Average Pre-test Score | 38.67 | 37.88 |
| Variance Pre-test Score | 230.22 | 238.54 |
| P-Value F-test | 0.8215 (NS) | |
| P-Value t-test | 0.5812 (NS) | |

During the first semester, 124 students took the course. In the second semester, 934 students completed the course. The average score on the pre-test the first semester was 38.67 while the average score for the second semester was 37.88. The pre-test was scored out of 100. After using the F-Test to determine that the variances for the two semester samples is unequal, we conducted a t-test to examine statistical differences in the mean scores for the two semesters. The P-Value for the t-test is 0.5812. We therefore conclude that there is not a statistical difference between the students with regard to prior Excel skills between the two semesters.

With this assurance that there is not a systematic difference in the students' Excel skills across the semesters, we examined the effect of the course policy change on student performance in the class. As we postulate in hypothesis 2, we expect student learning to be higher when the average score is recorded in the grade book for an assignment with multiple attempts than when the highest score is recorded. Unfortunately, the instructor did not require that students complete the final exam in this course. The only measures of student learning and performance we have for this study is the scores for each of the assignments.

Study 2 Findings

When conducting the analysis for this study we were concerned that the results would be biased by fact that recording the highest score in the grade book would likely lead to higher overall scores for each assignment even if student learning on those assignments was not as great in the higher score semester when compared to the average score semester. We thought that we may need to collect additional secondary data from students, asking them to report on how much they felt they had learned in the class. What we found was surprising.

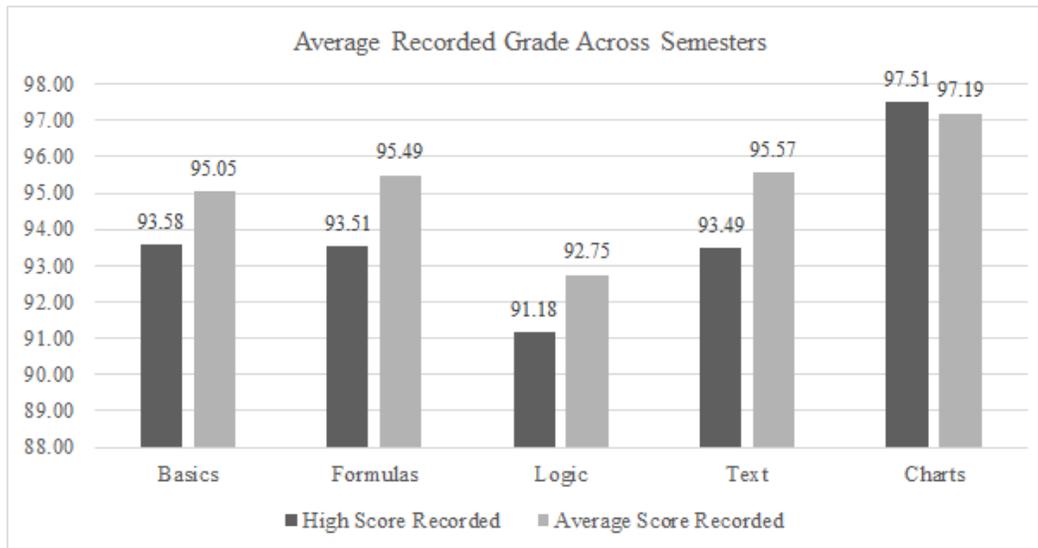


Figure 1. Average Assignment Scores

Figure 1 compares the mean recorded scores for each semester. Surprisingly, the mean recorded score was higher for four of the five assignments in the course where the average score was recorded in the grade book than for the course where the highest score was recorded in the grade book. This was counter to our expectation. The single exception to this pattern was the charting assignment, where the mean scores are very close. We assumed that the highest score would be higher for most students in the course where the high score was recorded than the average score for most students in the course where the average score was recorded – even if students were actually learning more in the average score course. We expected this because lower first-attempt scores would not factor into the grade that is recorded in the grade book. Given that students scored higher in the average score class despite this possible bias, we determined to see if the differences in raw scores were statistically significant. This would provide compelling evidence that student learning was higher in the score where the average score was recorded in the grade book compared to the course where the highest score was recorded.

Table 4. Comparing Student Performance for Difference Grade Resolution Policies

| | Excel Basics | Functions | Logic | Text and Time | Charts |
|-----------------------------|--------------|-----------|-------------|---------------|-------------|
| Count Semester 1 | 124 | 121 | 120 | 122 | 121 |
| Count Semester 2 | 933 | 930 | 926 | 927 | 922 |
| Mean Semester 1 | 93.58 | 93.51 | 91.18 | 93.49 | 97.51 |
| Mean Semester 2 | 95.05 | 95.49 | 92.75 | 95.57 | 97.19 |
| Mean Difference | 1.47 | 1.98 | 1.57 | 2.07 | -0.33 |
| Variance Semester 1 | 53.38 | 66.21 | 107.90 | 36.57 | 51.05 |
| Variance Semester 2 | 44.44 | 48.64 | 89.52 | 30.36 | 63.12 |
| P-Value F-test [†] | 0.0783 | 0.0089 | 0.0774 | 0.0767 | 0.0710 |
| P-Value t-test | 0.0175* | 0.0020* | 0.0592 (NS) | 0.0002* | 0.3206 (NS) |

[†]**Note:** Non significant p-values for the F-test mean variances are unequal.

Table 4 presents the results of the significance tests we conducted. As already noted the mean recorded score was higher in the average score class than for the higher score class for all assignments with the exception of the charts assignment. In order to conduct the correct t-test to examining if the calculated mean differences are statistically significant, we first conducted F-tests to see if the sample variances are equal. We determined that the sample variances were unequal for all assignments except the logic assignment. We used the appropriate t-test to examine if the calculated mean differences for all assignments are statistically significant. We found that the mean differences for three of the five assignments (Basics, Functions, and Text and Time) are statistically significant. The mean difference for the Logic assignment is not statistically significant at the alpha = .05 level, but it nearly significant. The mean difference for the Charting assignment is also not statistically significant.

Taken together these tests provide substantial support for hypothesis 2. Students do appear to learn more when the average score is recorded in the grade book than they do when the highest score is recorded. The learning advantage for students in the average score course appears to be strong enough that it shows up in the overall scores students earn on the assessments in the class, despite the fact that the scores for the highest score class would likely be inflated by the fact that lower first-attempt scores do not factor in grade recorded in the grade book.

Results Summary

In summary, we conducted two studies to examine the two hypotheses reported in this paper. Hypothesis 1 suggests that students will exhibit higher levels of learning if they are able to apply feedback about their performance than they will if they are just presented with the feedback. Hypothesis 2 suggests that when students are allowed to completed an assessment multiple times with feedback from initial attempts, they will exhibit higher levels of learning if the average score for their attempts is recorded in the grade book than if the highest score is recorded.

Table 5. Summary of Results

| Hypothesis | Result |
|---|--|
| Hypothesis 1: student learning improves when students apply feedback rather than just receive feedback. | Study 1 confirms this hypothesis |
| Hypothesis 2: student learning improves when average scores are recorded for multiple attempts rather than highest scores. | Study 2 provides partial confirmation of this hypothesis |

The table highlights a summary of the results of our tests of these hypotheses. Study 1 confirms hypothesis 1. Students exhibit higher levels of learning when they apply feedback from an initial attempt at a task opposed to just receiving feedback on their performance. Study 2 provides substantial support for hypothesis 2. Students demonstrate higher levels of learning when they are allowed to complete more than one attempt at a task if the average score of the attempts is recorded in the grade book than if the highest score is recorded.

Discussion

Our research provides substantial support for the idea that the way that students are allowed to use feedback can have a dramatic impact on learning. First we found that when students are allowed to apply feedback to a second attempt at completing a task, their learning increases over just being presented with the feedback. Providing immediate and detailed feedback to students on their task performance is critical to this approach. Computer-aided assessment techniques provide a good way to ensure that feedback is delivered in a timely fashion.

Providing the opportunity for students to apply feedback directly to a second attempt at an assignment works because students use feedback differently when they can apply it than instead of using it to explain their performance. When students can apply feedback on a second attempt at an assessment the feedback becomes both summative (performance-based) and normative (process-based). As mentioned earlier prior research [4] has highlighted the benefits of process-based feedback in learning.

Our informal discussions with students indicate those students who receive just feedback see the feedback as instructor initiated justification for their score. When students are allowed to apply the feedback to complete an assessment again, the feedback is viewed as suggestive of how the students can improve their work. Put another way, when feedback used on a second attempt at an assessment, students view the feedback more in terms of highlighting *why* their work was incorrect, not just *what* was incorrect. We believe that this subtle and powerful shifting from what to why explains why student performance not only on the individual assessments, but also on the comprehensive final exam for the class.

This finding has some important implications for instructors, because designing a course that allows students multiple attempts at the same assessment are not cost-free. First, course designs should allow enough time for students to complete multiple attempts for each assessment, or provide multiple activities that assess similar tasks. Using this approach will likely limit the number of topics covered in a class on increase the amount of work that students complete outside of the classroom. Additionally, as student performance increases from a learning

perspective, student scores will also increase. Many instructors are feel pressures to reduce or eliminate grade inflation. Instructors will need to weigh these costs against the increases in student learning that occur when students are provided with the opportunity to complete an assessment more than once.

Our second main finding – that student performance increases when the average score is recorded in the grade book instead of the highest score, also has implications for course designs. While this policy will be easy for instructors to implement, it can seem counter-intuitive to both instructors and students. Many of the instructors we work with that use our automated grading technology are tempted to record the highest score students achieve across attempts rather than the average score. Recording the highest score feels more generous to students. At the very least students perceive it to be a more generous approach, because it removes any risk the student may incur in the first attempt at an assessment.

We believe that it is the risk associated with the first assessment that makes this approach more effective. Again we asked students to informally comment on their thoughts with both approaches. We found an interesting recurring theme with students in the class where the highest score was recorded. Students in this class felt that the first attempt was cost-free. Many chose to take this first attempt before engaging in any of the learning activities of the class (watching instructional videos, reading the textbook, or completing practice problems). This anecdotal evidence is confirmed by examining server logs of student interactions with the course materials. Students would use the first attempt to diagnose those areas they needed to study. Some even suggested that they used the first attempt to see what the grading engine was “looking for” in order to manipulate the assignment rather than internalize the content in assessed by the assignment. While this is more efficient for students from a time point of view, it is likely not effective for long-term learning or for applying the skill being taught to another context. Many students even suggested that they were able to successfully complete the class without engaging with the learning materials.

We found the behavior patterns and attitudes of students toward the first attempt to be dramatically different in the average score class than in the highest score class. Because the first attempt would ultimately factor into the score recorded in the grade book, students were motivated by the riskiness of the first attempt. While there were still some students who attempted the assessment before viewing the course learning materials (videos, textbook, and practice problems), a higher percentage of students engaged with the learning materials before completing the assessment. Moreover, students did not report using the first attempt to see what the grading engine was “looking for”, because the risk of earning a very low score on the first attempt was too high. Students invested the time to learn before completing the first attempt, because the first attempt ultimately mattered for their grade in the class.

We believe the attitude and behavioral differences students expressed about how the first attempt was viewed under the different ways the assessment score was handled in the grade book explains differences in how well students learned the topics covered in the class. Students took the first attempt more seriously because of the risk associated with that attempt for their grade. This lead to greater engagement in the learning materials for the course before the assessment was completed.

However, we were still surprised that the recorded student scores for the individual assessments were actually higher when the average score is recorded in the grade book than when the highest score was recorded. We believe that this may have been due to another difference in how students viewed the assessments in the different classes. Anecdotally, students suggested to us that there was a threshold score (somewhere below 100%) that was acceptable to them in the class. While this threshold may be different for each student, it served as a benchmark to determine if they should invest the time to complete the second attempt at an assignment. While this was the same for students in both courses, we believe that students in the average score class ended up with higher recorded scores on the assessments because they had already incurred much of the cost of learning before the first attempt. This meant that they were both more likely to get a higher score on the first attempt and less likely view the cost of the second attempt as too costly (since students in the highest score class likely had not engaged the course materials as extensively before the first attempt).

Taken together, these findings suggest a powerful combination for instructors. Instructors should allow their students to complete assessments more than once using feedback generated from the first attempt to inform subsequent attempts. Instructors should also record the average score across submissions as the final grade for an

assessment. We believe the combination of approaches will motivate the students to engage with the learning materials in the course and use feedback most effectively.

LIMITATIONS

In considering the limitations of the studies reported in this paper, it is important to remember several aspects of the context of the studies that significantly influence the results. First, the studies examined student performance on relatively open-ended Excel assignments. Students were presented with a set of problems that required them to use a correct function or process in Excel to solve the problem. While these problems are relatively structured, the number of possible student responses for each question is much larger than for an multiple choice assessment where the granting of multiple attempts at solving the problem may be less appropriate.

Moreover, students were provided rapid feedback immediately after completing each assessment. Therefore, the tasks associated with the assessment are likely still very easily recalled by students. Therefore less cognitive effort would likely be required to interpret the feedback and apply it to a second attempt at the assessment than if the feedback cycle were delayed. In short, the findings of this report may not be generalized to other learning settings if allowing multiple attempts at an assignment is not appropriate or if the feedback cycle is long enough that the feedback is difficult to apply to a subsequent attempt at an assignment.

In addition to the generalizability limits of the study, it may well be that the study designs might suggest caution in interpreting the results. Neither study represents a strong experimental test of the hypotheses. Instead, secondary data captured before and after significant course policy changes were used to test the hypotheses. While we systematically tested for differences in students' prior skills in Excel that might explain differences in our test variables, other unmeasured differences related to collecting data from courses that span semesters might influence the results.

Finally, the hypotheses suggested in the paper were tested independently. It may well be that the two treatments are not independent. For example, we find that student learning improves when students are allowed multiple attempts at a task. We also find that recording the average score for multiple attempts is superior to recording the highest score from the student learning perspective. However, we do not know if student learning will also improve if they are allowed multiple attempts at a task and the highest score is recorded in the grade book. In other words, we do not know if both multiple attempts and average score reconciliation are needed for learning improvement or if allowing multiple attempts will improve student learning under both the average and highest score conditions. To fully understand how these two conditions interact, a follow-up study would be necessary. Unfortunately, the ethics of such a study would be questionable. It would require a researcher to assign students to treatments thought to be inferior (would any professor want to randomly assign students to a treatment where they were not allowed to submit an assignment more knowing that those students would likely not learn as much as they would in another treatment). Are would need to be taken with this future research to ensure both the ethics and the validity of the study.

CONCLUSIONS

The studies reported in the paper suggest a powerful combination of pedagogical strategies instructors can employ to improve the learning of their students. Prior research has highlighted the importance of feedback in the learning experience. This study suggests that students will benefit from course designs that allow them to apply feedback received on their performance. Students exhibit higher levels of learning when they are allowed to apply feedback from a prior attempt at an assessment than if they just receive that feedback on their performance.

Second, we suggest that how the instructor reconciles these multiple attempts at an assignment into a final recorded grade for that assignment is also important. When an instructor records the average grade for an assignment in the grade book rather than the highest score, student performance improves. Taken together the findings in the paper suggest a powerful pedagogical combination for students: allow students the opportunity to apply feedback in subsequent attempts at an assignment and then record the average score for these assessments.

REFERENCES

1. Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *Research in Learning Technology*, 13(1), 17-31.
2. Gibbs, G., & Simpson, C. (2004). Conditions under which Assessment supports Student Learning. *Learning and Teaching in Higher Education*, 1(1), 3-31.
3. Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of Individual Feedback on Behavior in Organizations. *Journal of Applied Psychology*, 64(4), 349-371.
4. Johnson, D. S., Richard, P., & Kalen, P. F. (1993). Differences in Task Performance as a Function of Type of Feedback: Learning-Oriented Versus Performance-Oriented Feedback. *Journal of Applied Psychology*, 23(4), 303-320.
5. Liden, R. C., & Mitchell, T. R. (1985). REACTIONS TO FEEDBACK: THE ROLE OF ATTRIBUTIONS. *Academy of Management Journal*, 28(2), 291-308.
6. Sim, G., Holifield, P., & Brown, M. (2004). Implementation of computer assisted assessment. *Research in Learning Technology*, 12(3).