THE IMPACTS OF PROVIDING NOVICE COMPUTER SCIENCE STUDENTS WITH A SECOND CHANCE ON THEIR MIDTERM EXAMS*

Ben Stephenson
Department of Computer Science
University of Calgary
2500 University Drive NW
Calgary, Alberta T2N 1N4
Canada
ben.stephenson@ucalgary.ca

ABSTRACT

Over the years, we have been approached by numerous students seeking a "second chance" after performing poorly on a midterm exam in an introductory computer science course. While many students have provided creative justifications for why they should be treated differently from everyone else, some students have articulated reasons why all students in the class should receive a second chance. These reasons have included lack of experience writing university level exams, the cumulative nature of the final exam, and lack of prior experience writing exams in this subject area.

After considering the arguments, the instructor for the course adjusted the grading scheme in a subsequent year so that students would receive a second chance on the midterm exam. Specifically, students who achieved a higher grade on the final exam than on the midterm exam would have their midterm exam grade replaced with their final exam grade. Making this change resulted in improved student perception of the fairness of the grading scheme while having little impact on the class average grade for the course. While these positive outcomes were observed, some negative impacts also occurred including lower midterm exam participation rates and a higher failure rate for the course.

1.0 INTRODUCTION

Course outlines often describe learning outcomes that students are expected to achieve by the end of the course. Yet instructors often determine a significant portion of students' grades at the midpoint of the academic term using a test or exam. Locking in a significant portion of students' grades at this point of the term fails to reward students who master this material later in the course. During our teaching career, we have also had several students express that they believe their mark on the midterm exam didn't accurately reflect their knowledge because of lack of experience writing either university level and/or computer science exams.

Seeing some validity in these arguments, we decided to provide students with a second chance on the midterm exam. This second chance was implemented by replacing students' midterm exam grades with their final exam grades when the final exams showed improved performance. Our hope was also that this would change the midterm exam from being purely an assessment tool into a combination of an assessment tool and a learning experience.

The remainder of this paper is organized in the following manner. Section 2 provides a brief overview of the students taking the course, the topics covered in the course, and the students' expected level of understanding. Section 3 describes the process we followed to assess the impact of our change in grading practice. The differences we observed in students' level of participation in various assessments and students' grades are described in Section 4. It also outlines changes observed in the course evaluations completed by the students. Section 5 presents our conclusions.

2.0 COURSE DESCRIPTION

At our institution the Faculty of Science offers a Natural Sciences Program. This program is targeted at students "who have a broad interest in science and wish to emphasize the fundamental importance of a multidisciplinary approach" [5]. As part of this program, students are required to take courses from at least four departments within the faculty of science, including at least one half-course from the department of computer science with significant programming content. As students advance in their program, they select two concentration areas where they take more advanced courses.

Our introductory computer science course serves multidisciplinary students, many of whom are enrolled in the Natural Sciences Program. As such we have students who are scientifically and mathematically literate, but that are not necessarily inherently interested in computer science. Separate courses are offered for students majoring in computer science programs, and a separate course with little programming content is offered to support programs outside of the faculty of science.

The overall goal of our course is to provide students with the skills necessary to write small computer programs that will assist them in their pursuit of knowledge in other scientific disciplines. Upon completing the course, students are expected to be able to take a description of a problem, decompose it using top down design, and then synthesize a solution to the problem as a Python program. During the course students gain experience using common programming constructs such as if statements, loops and functions. They are introduced to array and dictionary data structures which they must

use to complete the final two assignments in the course. By the end of the course they are also expected to be able to load data from files and perform basic exception handling. Throughout the course students make use of a graphical package [4] which allows them to perform simple visualizations such as graphing a dataset.

3.0 EXPERIMENTAL DESIGN

We have taught introductory computer science to multidisciplinary students on several occasions. In fall 2008 we taught the course for the first time using materials inherited from the previous instructor. We made some adjustments to the course content and delivery before teaching the course for a second time in spring of 2009. A few additional adjustments were made between the spring 2009 and fall 2009 terms. Since fall 2009 the course has remained stable, both in terms of its content and the techniques used to deliver the content.

In this study we compare data from the fall 2009 academic term with data from the fall 2010 academic term. The course was a large enrollment class each year, with 150 students in 2010 and 180 students in 2011. Two sections of the course were offered each year, all with the same instructor. All of the comparisons performed in this study consider the entire year's data. We don't attempt to perform any analysis for individual sections because we noticed that some students choose to attend the earlier lecture section even though they are actually enrolled in the later lecture section, and vice versa. The class times were the same each year.

The instructor made a conscious effort to minimize differences in the course between these terms. Students were recommended to purchase the same textbook each year, and students were provided with nearly identical course notes (a handful of typos were corrected from 2009 to 2010). The instructor did not consciously change any aspects of the course beyond the grading policy that is the focus of this study.

Students completed the same number of assignments each year. While the assignments were not identical, they were highly similar. Assignment 1 included a small programming task along with a collection of written questions about data representation. Students were asked to complete the same programming task each year. The data representation questions used slightly different numbers in 2009 and 2010.

Later assignments were adjusted in similarly small ways. In 2009, assignment 2 asked students to create a program that plotted an equation of the form $y = ax^3 + bx^2 + cx + d$. In 2010 students created a program that plotted an equation of the form $y = a(x-b)^c + d$. Assignment 3 asked students to implement four image transformations. In 2009, the operations were darken, chroma key, rotate by 180 degrees and blur. In 2010 the operations were negative image, chroma key, rotate 90 degrees counterclockwise and blur. Assignment 4 asked students to read data from a file, process it, and generate a bar graph. No significant changes were made to this assignment between 2009 and 2010.

Students were also assessed using a midterm exam and a final exam. Because the midterm exam was returned to the students it was necessary to make significant changes to it between 2009 and 2010. As such, we do not perform any comparisons on midterm exam results between the two years.

Final exams were not returned to students in either year. While students had the opportunity to review their final exam after final marks for the course were posted, no students elected to utilize this opportunity. As such, we were able to use the same final exam in both 2009 and 2010. The exams were carefully controlled between 2009 and 2010 to ensure that no students gained access to them.

4.0 RESULTS

We examined the impact our change in grading policy had on a variety of factors. Section 4.1 examines changes in the number of students that elected to participate in each assessment. The number of students that withdrew from or failed the course is explored in Section 4.2. The impact this change had on final exam grades is described in Section 4.3 and Section 4.4. Section 4.5 describes the impact this change had on the class average grade. Course evaluation data is presented in Section 4.6.

4.1 Participation Rates

We examined the proportion of students that submitted each assessment each year. Table 1 summarizes the behavior we observed.

	2009	2010	Difference	Chi Square
Number of Students (n)	150	180		
Assignment 1	90.0%	91.7%	1.7%	0.56
Assignment 2	86.0%	83.9%	-2.1%	0.67
Assignment 3	79.3%	78.9%	-0.4%	0.02
Assignment 4	71.3%	73.9%	2.6%	0.57
Midterm Exam	96.0%	90.6%	5.4%	13.89
Final Exam	80.7%	81.7%	1.0%	0.12

Table 1: Assessment Participation Rates in 2009 and 2010

Table 1 shows that most of the participation rates were similar each year. A Chi Square (?2) test of proportions was applied to each assessment to determine if the differences observed were statistically significant. The differences for the assignments and the final exam were not found to be statistically significant at a 95% confidence level (critical value: 3.84).

A statistically significant difference was observed in the Midterm Exam participation rate (critical value: 3.84, observed value: 13.89). In 2009, 144 of 150 students (96.0%) wrote the midterm exam for the course. In 2010, the midterm exam participation rate dropped to only 90.6% (163 of 180 students).

This difference in midterm exam participation rate is worrisome. When students fail to write the midterm exam 70% of the students' final grades are determined from a single 2-hour assessment. This is worrisome, both because students gave up the

opportunity to "lock-in" a portion of their grade by writing the midterm exam, and because, on average, students do more poorly on the final exam. We believe that this difference resulted from some students taking the midterm exam less seriously because they knew that they would have a second chance to earn those marks. Walker has previously noted that "Student's typically view points as identifying what matters" [7]. The same conclusion is also reached by Gibbs and Simpson [3]. In this case, our grading policy may have inadvertently sent the message that the midterm didn't matter since choosing not to write it did not result in the direct deduction of any points.

4.2 Course Failures vs. Course Withdrawals

At our institution, students are able to withdraw from a course until the last day of classes. When a student withdraws from a course a grade of W is recorded in the student's academic record. However this W does not impact the student's grade point average. When a student fails a course a grade of F is recorded in the student's academic record. A grade of F is included in a student's grade point average, contributing 0 grade points. As such, students benefit from withdrawing from courses instead of failing them.

As part of this study, we examined the pattern of student failures and withdrawals. Table 2 shows the distribution of withdrawals and failures each year.

	2009	2010	Difference
Number of Students (n)	150	180	
Passed	72.7%	72.8%	0.1%
Withdrew	18.0%	13.3%	-4.7%
Failed	9.3%	13.9%	4.6%

Table 2: Student Failures and Withdrawals

While the proportion of students who didn't receive credit for the course was nearly identical in each year (27.3% in 2009, 27.2% in 2010), more students failed the course in 2010, with the resulting negative impact on their grade point averages. Using a Chi Square test of proportions revealed this difference was statistically significant (critical value: 5.99, observed value: 6.18).

We believe that this difference resulted from students having unrealistic optimism about their ability to achieve a substantially better grade on the final exam than on the midterm exam, and that this optimism lead them to write (and fail) the final exam when withdrawing from the course would have been the more prudent choice. This phenomenon is known as optimism bias [1], and can occur both inside and outside the academic arena [2].

4.3 Final Exam Grade Improvement

Of the 147 students that wrote the final exam in 2010, 28 earned a higher grade on the final exam than on the midterm exam. As such, approximately 19.0% of students improved their grade as a result of this change. For comparison purposes, 102 students

(69.4%) achieved a lower grade on the final exam than on the midterm exam (and as such, retained their midterm exam grade), and approximately 11.6% of students achieved the same grade.

The amount of improvement varied greatly among students that achieved a higher grade on the final exam than on the midterm exam. While many students only improved by a third of a letter grade (for example, moving from a C to a C+), one student improved from a D+ on the midterm exam to an A on the final exam. On average, students that improved increased their grade by approximately 0.73 grade points. The following figure shows the distribution of the improvement.

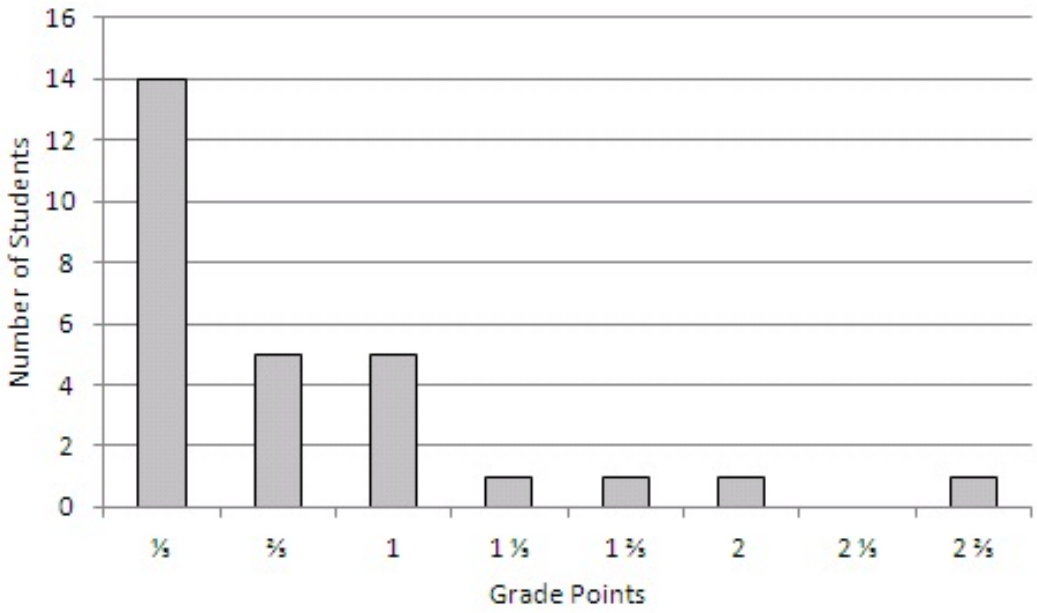

Figure 1: Amount of Improvement among Students that Achieved a Higher Grade on the Final Exam

4.4 Which Students Showed Improvement?

As noted previously, approximately 19.0% of students improved their midterm exam grade by earning a higher grade on the final exam. In this section, we examine which students benefited from this opportunity. Figure 2 shows the proportion of students that improved on the final exam, divided by midterm exam grade.

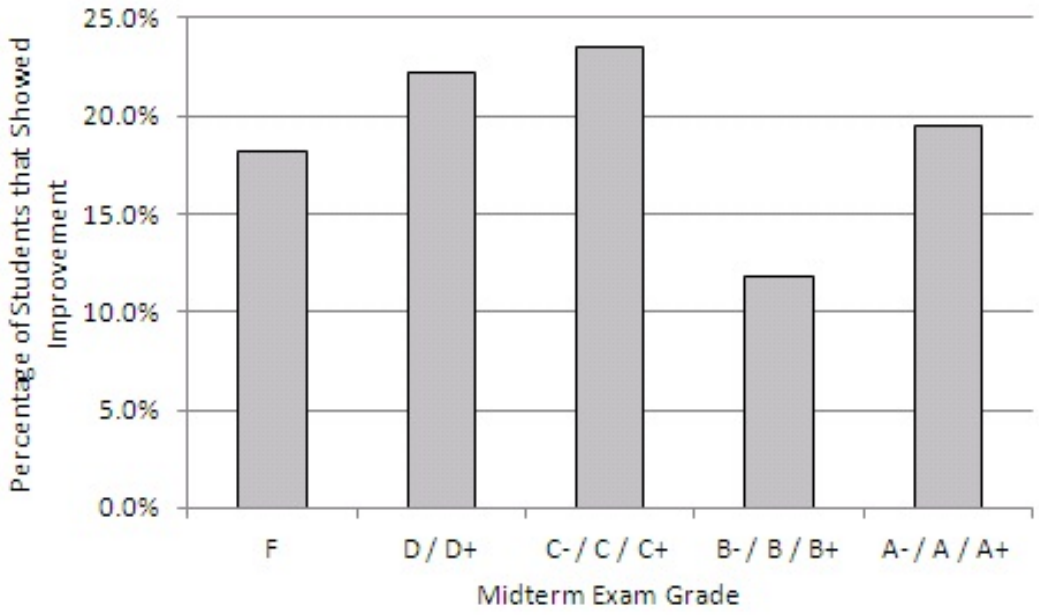

Figure 2: Distribution of Students that Improved

As Figure 2 shows, students that achieved a C or D on the midterm exam were the most likely to improve. On one hand, this is intuitive - students with these grades have among the most to gain by expending extra effort before the final exam. However, we had assumed that improvement in this category of students was less likely because their lackluster midterm exam performance indicated that they did not have a solid foundation to support learning the more advanced topics later in the course. This result suggests that the students' arguments that their midterm exam performance was not an accurate reflection of their knowledge may have some merit. It could also indicate that the midterm exam was a learning experience in addition to an assessment.

Previous work has explored the factors that make exams a learning experience [8]. In that study students wrote the same exam twice, once individually and once in a small group. This was found to improve student learning considerably. While our students didn't write the same exam twice, there are some parallels. Our students were assessed on the same material twice, in a similar way each time, and they did have ample opportunity to learn from each other, their teaching assistants or the course instructor between the two assessments.

4.5 Impact on Class Average

Before implementing this change in grading policy we had some concerns about grade inflation. Having now completed our study, we have seen that grade inflation is not a significant concern. The grade point average determined from the four assignments, original midterm exam score, and final exam score, excluding students who withdrew from the course was approximately 2.26 grade points. When the midterm exam grades were replaced, this increased to approximately 2.30 grade points, a difference of just less than 0.04 grade points. While this is an increase, it is of

sufficiently small magnitude that we do not consider it to be a problem. Furthermore, we believe that this increase is a more accurate reflection of student's knowledge at the end of the course, and as such, is fully justified.

4.6 Course Evaluations

Near the end of the course our institution conducts a survey of the students. Students are asked to rate 12 aspects of the course [6]. Overall instruction is rated on a scale from unacceptable to excellent. Students are then asked about their level of agreement with 11 additional statements such as "The course content was communicated with enthusiasm", "Students were treated respectfully" and "I learned a lot in this course". Students are also asked about their level of agreement with the statement "The evaluation methods used for determining the course grade were fair".

A Chi Square (?2) test of proportions was performed on each question to identify any differences between 2009 and 2010. The number of students that somewhat agreed, agreed or strongly agreed with positive statements were compared to the number of students that were neutral or disagreed with such statements. No statistically significant differences were observed in 11 of the 12 categories. A statistically significant decrease in the number of neutral or negative responses was observed for the statement "The evaluation methods used for determining the course grade were fair" (critical value: 3.84, observed value: 4.43). In 2009, 9.5% of students had a neutral of negative response to this question. This decreased to 2.5% of students in 2010.

5 CONCLUSION

We evaluated the impact of replacing student midterm exam grades with their final exam grades when their final exam grades exceeded their midterm exam grades. Some positive impacts were observed, including an increase in midterm exam grade for approximately 19.0% of students, and a statistically significant decrease in the number of students that were neutral or negative toward the statement "The evaluation methods used for determining the course grade were fair" on the course evaluations. However, some negative impacts were also observed. The participation rate on the midterm exam showed a statistically significant decrease from approximately 96.0% to 90.6%, and while the combined total for withdrawals and failures remained essentially unchanged, the proportion of failures increased considerably. Based on these results we believe that it would be worthwhile to continue our policy of replacing midterm exam grades with improved final exam grades. However, we also believe that additional steps should be taken to try and combat students' optimism bias so that the number of students failing the course can be reduced to 2009 levels.

REFERENCES

[1] Armor, D. A., Taylor, S. E., When Predictions Fail: The Dilemma of Unrealistic Optimism. In Gilovich, T., Griffin, D., Kahneman, D. (Eds.) Heuristics and biases: *The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press, 334 - 347, 2002.

[2] Dunning, D., Heath, C., Suls, J. M., Flawed Self-Assessment: Implications for Health, Education, and the Workplace, *Psychological Science in the Public Interest*, 5 (3), 69-106, 2004.

[3] Gibbs, G., Simpson, C., Conditions under which assessment supports student learning, *Learning and Teaching in Higher Education*, 1 (1), 3-31, 2004.

[4] Stephenson, B., Taube-Schock, C. QuickDraw: Bringing Graphics into First Year. *Proceedings of the 40th ACM Technical Symposium on Computer Science Education (SIGCSE '09)*, pages 211-215, 2009

[5] University of Calgary, Natural sciences program, http://www.ucalgary.ca/natsci/, 2010.

[6] University of Calgary, Universal Student Ratings of Instruction (USRI) Sample Form and Questions, http://www.ucalgary.ca/usri/ref/forms, Accessed October 12, 2011.

[7] Walker, H. M., Classroom Issues: Grading and the Allocation of Points, *SIGCSE Bulletin*, 41 (4), 14-16, 2009.

[8] Yu, B., Tsiknis, G. Allen, M., Turning exams into a learning experience, *Proceedings of the 41st ACM technical symposium on Computer science technical symposium on Computer science education (SIGCSE '10)*, 336-340, 2010.