# Exams and Student Feedback: An experiment in Marking Efficiencies

Denard Lynch, Andrew Kostiuk University of Saskatchewan denard.lynch@usask.ca, andrew.kostiuk@usask.ca

**Abstract:** Providing summative feedback to students in a timely fashion, and managing the associated marking in larger classes has been a perpetual challenge in an education environment, and is even more so in a resource challenged environment. This paper discusses the results of an experiment in evaluation in an engineering course by implementing a modified evaluation and grading approach. The objectives were to i) provide timely feedback to students, ii) improve engagement and reduce overall course loading for students, and iii) reduce marking effort for instructors, all without negatively affecting student grade performance. The results show that improvements over traditional methods can be made in two of the three areas.

The course in question, (redacted), covers basic electrical concepts and devices for non-electrical engineering students. The course had been offered in four previous years using a traditional evaluative approach: weekly assignments (submitted, marked and returned), laboratory exercises (comprehensive reporting or exercises submitted, marked and returned), midterm(s) (graded and returned), and a final examination. The modified approach was implemented over the past two vears and included the same learning strategies, but with a potentially lower resource commitment for students and instructors. Modifications to the strategy were implemented the first and second years. The experiment introduced procedural and administrative modifications in assignments, laboratories and examinations, and the addition of short weekly quizzes to improve engagement in an active learning environment.

Approximately ten assignments were offered to help students test and improve their understanding and knowledge. In the first year, assignments and solutions were posted simultaneously; no submission was required and there was no grade contribution offered. The rationale for this strategy was that students would receive virtually instant feedback by having solutions immediately available, and the freedom to judge the quantity and level of completion required to meet their individual learning needs. In the second year, assignments were administered through an online assignment system for CEEA18; Paper 117 mark credit. This was intended to reinstate the incentive of mark credit to improve student engagement while still providing instantaneous feedback on correctness.

The course has always included a critical "hands-on" laboratory component which was traditionally time intensive for both students and instructors. While the laboratory submissions were still required for mark credit, the reporting requirement was reduced to a minimal, specified sampling of results to provide evidence that the practical work was addressed. Expected outcomes were again provided for students to provide relevant and timely feedback. In the second year, a 3-bin grading system was adopted to improve the granularity of the marks while still requiring considerably less marking effort.

Examinations were also modified to improve timeliness of feedback and reduce marking effort. In the first year, three "midterm" examinations were distributed through the term to monitor student learning and verify student participation in the self-directed parts of the course. Each of these exams consisted of 12 questions and were simply graded on a correct response (no "partial marks) to reduce marking effort. In the second year, two midterm examinations were deemed sufficient, but were graded using a 3-bin approach, thus allowing for "partial marks". Exams were returned to students in the next lecture period in both cases. The Final examination in the first year was designed using a 3-bin scheme to allow for partial marks while still reducing marking effort. In the second year, this was increased to a 4-bin scheme to improved granularity. The increase in granularity had very little effect on marking effort for both the midterm and final.

One additional modification was made in the second year with the addition of brief weekly quizzes, for mark credit, to encourage students to complete assigned prereading exercises and keep up with course work. The quizzes consisted of two brief questions: one on assigned reading for the coming week and one on the previous week's material. These quizzes were administered and graded using a classroom response system and

University of British Columbia; June 3 - 6, 2018

automatically integrated with the learning management system.

Analysis consisted of comparison of grades with previous years, anecdotal evidence and observations on student effort, course evaluation data and survey results. Preliminary results indicate student load and instructor marking effort were significantly reduced. While grade results were approximately the same. A direct objective comparison with previous years is not significant due to variations in course content and cohort.

*Keywords:* Student assessment, Assessment tools, summative, grading systems

## **1. INTRODUCTION**

Improving efficiency or "doing more with less" is, arguably, the purview of the engineer. This is certainly applicable in engineering education which, like most educational environments, is challenged by increasing expectations and decreasing resources. "Binning" to simplify marking is certainly not a new technique, and providing timely feedback to students on understanding and performance are believed to be beneficial to learning.

It is these postulations that formed the basis of modifications in a fundamental course in basic electrical and electronics principles in an attempt to improve the learning environment for students and also reduce the marking resource requirement in a relatively large class.

### 2. THE RESEARCH QUESTION

This paper discusses the results of implementing a number of pedagogical changes in the subject course and attempts to answer the question: "Can we reduce resource requirements for timely feedback and marking while improving, or at least maintaining, student learning, and student engagement?". Changes were made in the first year of the experiment and modified, based on feedback and experience, in the second year.

#### **3. COURSE MODIFICATIONS**

In the first four years, the course was assessed using traditional methods. Assignments were posted online, with a due-date typically one week following (assignments were submitted on paper except for the last year of this format where an online submission was implemented). Target time for returning marked assignments was one week following submission. Midterm exams, which included a mixture of short and comprehensive questions, were marked giving partial credit and returned to students one to two weeks after they were written. The final exam followed the same general format as the midterm, except for length. The modifications to the course assessment structure consisted of changing the quiz, exam and assignment formats.

# 3.1 Modifications in the First Year

The initial modifications to the assessment structure of this course focused on providing timely feedback to students on their understanding of class material as reflected in their performance on assignments and exams. The objective of providing results to students in as short a time as practical follows mainly from the observation that, historically, a large number of returned assignments were never retrieved or reviewed by students. While they would eventually see the overall result in terms of a grade, this clearly would not provide sufficiently specific feedback on where any misunderstandings might lie, especially not in time to correct it before moving on to subsequent, dependent material.

The simplest strategy to adopt to improve assignment timeliness was to post full, worked solutions with assignments so that students could immediately verify their own work and allow them to address any misunderstandings or lack or knowledge. Because the solutions were immediately available, no mark credit was given and assignment completion by students was not verified. The risk, of course, was that students would i) not do the assignments at all, or ii) just read through the worked solutions without first attempting a solution on their own. However, if students did attempt the assignments as recommended, feedback would be essentially immediate. It was generally believed that the timely feedback from exams also provided an opportunity for formative feedback.

The initial exam marking strategy was similarly focused on reducing marking time in order to allow a quicker return to students. The two-bin methodology ("all or nothing"; 0 or 1 out of 1 mark) followed from the common observation among markers that either a correct answer or a blank space require very little time to assess. The midterms were 12 questions each, marked using a two-bin scheme (0 or 1). The final was 24 questions marked using a three-bin scheme (0, 1 or 2 out of 2 marks). The risks were that i) very minor mistakes could mask any knowledge the student *did* have, and ii) the instructors may not gain feedback on weak or difficult areas in the class. In addition, students expressed concern that relatively minor oversights on their part could result in an undeserved low mark.

Mitigating these risks and addressing these concerns led to subsequent modification in the second year.

# 3.2 Modifications in the Second Year

In the second year of the experiment, further modifications were made based on student feedback, instructor observations and availability of an online assignment system. These further adjustments were implemented to address some of the risks and improve the student experience and learning opportunities.

First, effort was spent to create an assignment set within an available online assignment system. This system allowed students to do the assignments online, for credit, and allowed a number of attempts at each progressive "milestone" within a question. Feedback on correctness was immediate, although online help was limited. Each student received a unique set of values for each assignment problem, which encouraged every student to work through the calculations individually even if they were working in a group. The overhead of preparing assignments for this system was significant, but can be exploited in future years by using the same question base. This system also re-instated the incentive of mark credit while retaining, and even improving, the advantages of immediate feedback and encouraging engagement.

The only modifications made to exams were to increase the number of bins by one for both the midterm and final exams, and to reduce the number of midterms over the term to two (from three). These changes were implemented mostly in response to student feedback. From an instructor perspective, increasing the granularity by one bin only increased marking time by an estimated 20% and still allowed return of exams by the next class period. Also, the format of having a number of shorter questions focused on the individual concepts covered and recording question-related statistics provided valuable feedback to instructors on strong and weak points within the students' understanding. The reduction from three to two midterms was due in part to the addition of short, weekly quizzes to the assessment regime.

One of the secondary objectives of this experiment, and part of the vision for this course, was to continue to shift toward an active learning environment. The addition of weekly quizzes was intended to support this effort as well as provide another assessment vehicle that would provide frequent and immediate feedback to students. A weekly quiz was implemented using the Top Hat classroom response system integrated with the Blackboard learning management system, which met the goal of adding no marking effort and minimal overhead to the course load for instructors. Creating weekly quiz questions was balanced partially by removing one midterm exam. The quizzes were given at the beginning of the first class each week and consisted of two brief questions: one on the reading material assigned for that week, and one based on material covered the previous week. Because of the limitations of Top Hat, questions were generally graded either correct or incorrect (twobin). To overcome this limitation, some questions were composed of more than one part.

Finally, the time-saving advantages of binning were extended to the laboratory assignments in order to reduce the require marking resources required for this part of the course.

The laboratory component for this course is selfdirected, and to assign mark credit, it was assumed sufficient to verify that the assignment had been completed by examining a requested sampling of results (measurement or screen shots; submitted online). In the second year, these were evaluated using a three-bin scale: full marks if the expected result was submitted, zero if completely invalid or no result was submitted, and fixed, partial mark otherwise.

The result of these implementations was to reinstate mark credit incentive for all work, provide timely feedback, and reduce required marking resources for the course. Table 1 summarizes the initial and second-phase modifications and mark distribution implemented for this experiment.

	2016	2017	
Assignments	~10/0%	~10/10%	
(#/% of grade)	self-directed;	online; with	
	posted with	immediate	
	solutions	feedback	
Quizzes <sup>1</sup>	~none	~10/10%	
(#/% of grade)		best 8/10	
Laboratories <sup>2</sup>	~6/12%	~6/10%	
(#/% of grade)	formal grading	sampled/bin	
		grading	
Midterm Exams	3/36%	2/20%	
(#/% of grade)	12 Q's each	12 Q's each	
	2-bin grading	3-bin grading	
Final Exam	1/52%	1/50%	
(#/% of grade)	24 Q's	24 Q's	
	3-bin grading	4-bin grading	

 Table 1. Course Assessment Comparison 2016 - 2017

Notes:

- 1 The best 8 of 10 quizzes were used to determine grade credit
- 2 Lab assignments were substantially the same for both years. In 2017, a selected sample of the results was submitted online and marked either 0, 7 or 10/10.

The feedback gathered and relevant observations are discussed in the following section.

#### 4. RESULTS

Student feedback was solicited using a voluntary, postclass survey administered using the Top Hat classroom response system. In addition, student comments submitted via SEEQ (Student Evaluation of Education Quality) were also considered anecdotally.

The resource savings experienced from implementing these course modifications is also estimated.

Although not statistically significant or adequately controlled, a simple analysis of the grades on the final exams is included assuming the final exam performance would be a reasonable representation of student learning in this course.

# 4.1 Top Hat Post-course Survey

Student were given approximately a two week window in which to complete a brief survey via Top Hat. The survey consisted of four multiple choice questions with some opportunity to add free-form comments. Approximately 1/3 of students responded (total course enrollment: 147).

The questions offered to students were:

- 1. I would say the weekly review/readings quizzes (select all that apply):
- 2. Which of the following do you feel we should continue to do in this course?
- 3. Which of the following do you feel we should discontinue doing in this course?
- 4. How much value do you feel you would have received from this course if it met only once a week (say Tuesdays) and you were expected to do additional course-related work on your own (by referencing posted material and exercises).?

The data from the first three questions is shown graphically in Figure 1 through Figure 3 below. Question 4 was included to gather student opinion on possible future course structure or resource saving options. The results are shown in Figure 4.

The introduction of weekly quizzes was intended to encourage students to complete assigned pre-reading and to provide some feedback on how well they understood recently covered material. Only one selection was allowed for Question 1, and it is assumed the most significant reason was selected. Student comments generally corroborate the results shown in Figure 1, with the addition of the suggestions to decrease the mark weighting for the quizzes.



Figure 1: Weekly Quizzes Survey Responses

Question 2 was intended to determine the perceived value of key course elements, and students were allowed to select "all that apply".



Figure 2: Course Elements to Continue

For Question 3, students were again allowed to select "all that apply".

Note that while approximately 70% of respondents (Figure 1) found the weekly quizzes beneficial for one of the intended purposes, approximately 24% of respondents also suggested they be discontinued.



Figure 3: Course Elements to Discontinue

While not directly relevant to this current question, meeting less frequently has been considered as a strategy to encourage students to take a larger role in managing their own learning. In addition, this could reduce the amount of physical space required (e.g. classrooms) as enrollments increase. It is offered here for information.



Figure 4: Class Frequency Response

# 4.2 Estimate of resource savings

In this course, instructors have typically marked exams, and teaching assistants or markers have marked laboratories and assignments. While timely feedback was the primary consideration in reducing marking effort (and thus, time), the financial savings can also be a factor in a resource-challenged environment. Table 2 shows estimates of the hours saved by implementing bin marking and online assignments. The average enrollment over the history of this course (164 students) was used as a basis for total savings estimates.

	Traditional	Modified	Savings (hrs)	
Assignments <sup>2</sup>	197 0 <sup>1</sup>		197	
Midterms <sup>3</sup>	100	40	60	
Final <sup>3</sup>	65	32	33	
Laboratories <sup>2,4</sup>	96	48	48	
Quizzes <sup>5</sup>	0	0	0	
Totals	458	120	338 hrs	

Table 2.	. Resource	Savings	(hours)	per	Term
----------	------------	---------	---------	-----	------

Notes / assumptions:

- 1 This ignores instructor effort to create or modify assignments to be compatible with the online assignment system.
- 2 Marking resources for assignment and laboratories was traditionally budgeted as 1 hour per 10 enrolled students per week for 12 weeks of the term.
- **3** Based on an estimated observed average of 0.61 minutes per question. Finals estimate based on .91

minutes per question due to increased bin count and extra verification time.

- 4 This course included 6 laboratories though the term; half that of assignments.
- 5 This ignores instructor effort to create quiz questions and any administrative effort required.

## 4.3 Final Mark Analysis

One of the implied constraints of this experiment was that student learning should, at the least, be maintained. Uncontrolled variables such as cohort, exam question equivalence and prerequisite efficacy present challenges in validating grade comparisons. Although not experimentally verifiable, mark performance on the final (comprehensive) exam is discussed as an anecdotal indication of student learning.

Table 3 shows a comparison of the average, minimum, maximum, standard deviation and number of students enrolled in the course from its inception in 2012 until the most recent offering in 2017.

	2012	2013	2014	2015	2016	2017
Avg	51.5%	48.7%	58.0%	58.9%	59.3%	57.2%
Min	10.0%	3.0%	30.0%	30.8%	18.1%	18.1%
Max	93.0%	89.0%	97.0%	87.5%	97.2%	87.5%
StD	15.8%	15.8%	13.5%	11.1%	15.6%	15.6%
#	127	177	203	163	168	147

Table 3: Final Exam Marks Comparison (2012-2017)

Figure 5 shows the distribution of marks on the final exams over the history of this course. With the exception of a lower average in the first two years, the data reasonably approaches a normal distribution. While not statistically significant, the slightly lower performance in 2017 is of concern. However, it is speculated that this is a result of implementation of a new classroom model ("flipped" and "active") rather than a result of modifications introduced to improve feedback timeliness and conserve marking resources.



Figure 5: Distribution of Final Exam Marks (2012-2017)

## **5. CONCLUSIONS**

From direct observation, even with some potential error in estimates, it is clear from Table 2 that marking based on more granular assessment of performance leads to resource savings, and the reduce time required for marking results in more timely feedback to students.

Instructor observation and some student comments suggest some value in timely and more frequent feedback, and that simplifying (binning) marking schemes to achieve resource efficiency in the assessment process had no negative effect on instructor feedback. However, there is no defendable evidence (based on Figure 5 or Table 3) that improving the timeliness of feedback or increasing engagement through the modifications involved in this experiment had any positive effect on student learning.

#### 6. DISCUSSION/FUTURE WORK

The opportunity to reduce resource requirements through the implementation of appropriate strategies tested in this experiment is clearly available. However, determining the possible effect on student learning requires more specific control and focus. If a clearly separate control group is note feasible (and in this case it would be exceedingly challenging), perhaps utilizing an appropriate vehicle to categorize individual student's learning style (as defined by Kolb: diverging, assimilating, converging, and accommodating. [1]) and map and compare responses and results. The effect of applying a flipped" classroom style and its efficacy with different learning styles may be masking both the benefits and drawbacks of these strategies.

An ideal, albeit illusive, goal would be to determine a strategy that would achieve resource efficiency for both students and instructors and also improve student learning. The results of this experiment to date would appear to have missed the latter part of this goal. A similar observation was implied by Miller-Young describing the results of implementing a similar "pre-class reading quizzes" [2]. Her observations that "*The high-achieving students gave exclusively positive feedback about the format of the course.*" and "*It is not surprising that students who were the highest achievers in the course also did well on the quizzes. It is likely that the strong students already had successful reading and learning strategies before starting the course.*" also apply in this case, and we share her concern that it is a different part of the student cohort that should perhaps be our main target [2]. Also shared is the observation that there was no significant evidence that even the high achieving students improved their deep learning with a supposedly "better" classroom strategy [2].

Having specific and different strategies available in the same classroom may improve learning for all types of learners, should that even be possible. Being able to identify or develop styles that could be shown to improve learning for any or all student learning styles would be valuable for moving beyond resource efficiency and into clear teaching effectiveness.

#### Acknowledgements

The authors wish to acknowledge the support of the University of Saskatchewan for providing the means and opportunity for its instructors to explore and enhance their pedagogical skill and knowledge and contribute to the body of knowledge in this area. The authors would also like to thank Dr. Sean Maw and the other members of the Innovative Teaching and Research in Engineering Education Group (InTREEg) at the University of Saskatchewan for their support and suggestions. Finally, the authors thanks the students in the subject course for their comments, feedback, patience and participation while we attempt to improve their learning experience.

# **Works Cited**

- [1] D. A. Kolb, R. E. Boyatzis and C. Mainemelis, "~kgilbert/educ5165-731," August 1999. [Online]. Available: https://www.d.umn.edu/~kgilbert/educ5165-731/Readings/experiential-learning-theory.pdf. [Accessed March 2018].
- [2] J. Miller-Young, "Using Peer Instruction Pedagogy for Teaching Dynamics: Lessons Learned from Pre-Class Reading Quizzes," in *Proc. 2013 Canadian Engineering Education Association (CEEA13) Conf.*, Montreal, 2013.