### SUCCESSES WITH TWO-STAGE EXAMS IN MECHANICAL ENGINEERING

### Markus Fengler and Peter M. Ostafichuk University of British Columbia ostafichuk@mech.ubc.ca

**Abstract** Two-stage exams consist of a traditional pencil-and-paper examination written in class by individual students, followed immediately by a second sitting in which the students retake the same exam in teams (i.e. a collaborative test). The team test provides an immediate opportunity for students to discuss, debate, teach, and receive feedback on the subject matter. It draws on principles of goal-directed practice, timely targeted feedback, and collaborative learning.

The practice of two-stage testing is a defining feature of the Team-Based Learning approach, and is used for introductory reading quizzes that begin each course module. These have been part of the instructional approach in Mechanical Engineering at the University of British Columbia for over a decade. In 2014, we have extended two-stage testing to include midterm and final examinations. To accommodate the team portion, exams were shortened by approximately one third and questions were reformatted to be easier to complete in teams.

Students report a strong preference this approach (72% in favour) and report a resulting improvement in their understanding of the course material (75%). Exam performance gains have also been observed. In almost all cases, teams outperform their strongest member, and it is not uncommon that the weakest team outperforms the strongest individual in the class. As an added benefit, the revised question structure that makes it easier for students to collaborate on exam writing has also simplified and expedited the marking process.

*Keywords:* Assessment, two-stage exams, team exams, Team-Based Learning, collaborative learning

### **1. INTRODUCTION**

Student emotional and cognitive engagement is greatly heightened during examinations, yet, traditional examination formats, which are focused solely on individual assessment, do not leverage this engagement to improve learning and retention. In the traditional format, students leave the exam without feedback. In contrast, two-stage exams begin with a traditional, individuallyperformed examination, followed immediately by a teamanswered version of the same exam. The team exam provides all students with a chance to receive feedback and improve their understanding while they are intensely engaged. This is true for both high- and low-performing students. This process draws on principles of goaldirected practice, timely targeted feedback, and collaborative learning.

The practice of two-stage testing is a defining feature of Team-Based Learning [1], and is used for in-class quizzes on preparatory readings that begin each course module. This quiz format has been used in courses in Mechanical Engineering at the University of British Columbia since 2004. Beginning in 2014, we extended two-stage testing to midterm and final examinations.

Epstein et al.[2] showed students displayed improved learning and increased retention when given immediate feedback on their answers to test questions. Gilley and Clarkston showed that the use of two-stage tests provided learning gains almost three-times better than those achieved with an equal amount of time spent in self-study [3]. This finding has also been reported by others [4],[5]. Others have argued that two-stage exams increase student enjoyment [5], reduce student anxiety [6], and more closely link the process of testing with the process of learning [7].

In this work, the implementation and outcomes of two-stage testing in a second year mechanical engineering design course are presented. Details about the course are provided in Section 2, results and discussion in Section 3, and conclusions in Section 4.

### 2. COURSE CONTEXT

This study was conducted at the University of British Columbia (UBC) in a second year mechanical design course (MECH 223). The course is part of the integrated Mech 2 Program [8]. The typical course enrollment is 115-125 students split into 20 teams of 5 to 7. The course is delivered using the Team-Based Learning (TBL) approach [1],[9]. Course-specific details regarding this TBL implementation are extensively documented [10],[11],[12]. This is a seven credit course split into two parts (four weeks in January and three weeks in April). The course runs all day Monday to Friday and each part includes a major design project culminating in a competition.

As part of the TBL teaching approach, each of the 6 modules covering the course topics begin with a quiz on readings assigned from the textbook [13]. The quiz typically consists of 20 multiple choice questions. Students first complete and submit the quiz individually using Scantron cards. These cards give no feedback to students on the correct answer. After time expires, students submit their cards and then repeat the same quiz with their team. The team quiz responses are recorded with instant feedback using an Immediate Feedback Assessment Technique (IF-AT) scratch card (www.epsteineducation.com) that indicates whether or not the selected answer is correct.

The teams in this study consisted of six to seven students (20 teams each year) and were instructor-formed [14] in order to maximize diversity [15],[16] and to minimize previously established subgroups [1],[16]. For the TBL preparatory reading quizzes, the entire team worked together on the team portion of the quizzes; for the two-stage midterm and final exams, each team was divided into two sub-teams of at least 3 students. The teams and sub-teams rosters were maintained for the course duration.

### 3. RESULTS AND DISCUSSION

Results and discussion is provided below in four sections: TBL reading quiz performance, midterm and final exam performance, student perceptions of two-stage exams reported on surveys, and lessons learned in writing effective questions for two-stage exams.

### 3.1. Comparing Individual to Team Reading Quiz Performance

Individual and team reading quiz results for a six year period of the course (750 students, 120 teams), were analyzed and showed the following:

- The number of students who outperformed their team: 1 (0.13%)
- Average margin by which teams outperformed their *strongest* member: 13%
- Proportion of students outperformed by the *weakest* team in the class: 95%

These results clearly show that even a mediocre team outperforms the great majority of individuals in the class.

## **3.2.** Comparing Individual to Team Midterm and Final Exam Performance

For the 2014 academic year course offering, the performance of each individual on midterm and final exams was compared to the performance of their subteam. This was quantified by the realized performance gain: the proportion of actual mark improvement from the team normalized by the maximum possible mark improvement for the individual,

$$\frac{S_{ieam} - S_{indiv}}{100\% - S_{indiv}}$$
(1)

where  $S_{team}$  is the team's score on the test and  $S_{indiv}$  is the individual's score. This was considered for all four quartiles of individual score performance, as shown in Fig. 1 for the two midterm tests and the final exam.



Fig. 1. Realized performance gains

Some important observations from this chart are, first, that there is a positive average gain for all quartiles (the team does better than all individuals, on average), and second, the top quartile sees improvements of 20% to 30% (even the top students benefit in the process).

### 3.3. Student Perceptions

An optional and anonymous survey of the 2013-14 class included, among other things, questions about the two-stage exam format. The section on the survey regarding the two-stage exams was preceded by the following text,

"The midterm format this year was new, with a team test completed after the individual test. In past, the midterms were roughly twice as long and were only done individually. Please indicate whether you agree or disagree with the following statements."

and the three prompts were

- Working through the test as a team helped me to solidify and/or improve my understanding of at least some aspects of the course material.
- The length of the midterm allowed me to sufficiently demonstrate my understanding of the course material.
- The new midterm format (individual followed by team) is better than a longer individual midterm without a team component.

Responses were given on a 5-point Likert scale (strongly agree, mildly agree, neutral, mildly disagree, strongly disagree). The survey had a response rate of 35% and the results are shown in Fig. 2. The number in parentheses after each prompt is the percentage of students who mildly or strongly agreed with the prompt.



Fig. 2. Survey responses to two-stage exams

The result show positive student impressions of twostage testing. Importantly, students reported a preference for the two-stage format (71%). Even more compelling, 76% of student reported that they felt it improved their understanding of the course material.

# 3.4. Question Structure Supporting Two-Stage Exams

Based on multiple iterations of setting two-stage exam questions, our typical question structure has evolved over time. Accounting for the reduction in time available for each exam (since the new individual and team exams must both be completed in the same time as the traditional assessment), we worked to reduce the amount of timeconsuming writing required to answer each question. Similarly, while long, written questions can be effective for individual assessment, the need to have a team member write out a long answer in a team setting encumbers continued engagement of the other team members as they wait for the writer to finish. This reduces the time available for team discussion. Where possible, we restructured questions so that deeper understanding could be assessed without a significant amount of writing.

An example of a question structured for two-stage testing is shown in Fig. 3 on the subsequent page. The question asks the students to identify the most significant errors in a sample solution and to briefly justify their reasoning. (The specific scenario shown involves the translation of needs into design requirements.) Errors the students are to identify range from obvious to subtle, and identifying them tests deep understanding rather than memorization and mechanics. Rather than having a student busied with writing while other members wait, all team members can be simultaneously engaged. In similar questions to Fig. 3 in earlier years, prior to our adoption of two-stage testing, students were asked to develop and write out their own needs statements and associated requirements based on supplied information. While this earlier style of question was relatively easy to construct, it was time-consuming for students to answer. There were also some issues with dependency (an incorrect needs statement leads to an incorrect requirement, thereby making assessment more difficult).

The first implementation of our new style of question required only circling of errors. This was too simplistic as it presented difficulties when marking as the students' reasoning was not evident. The subsequent versions reintroduced limited writing, where students were asked to state a brief reason for selecting each error.

This style of question permits effective individual assessment and facilitates the team exercise by focusing on higher-level thinking, reducing writing and permitting division of work. While division of work may in theory allow some students to avoid answering questions in the team setting, our observations in the classroom are that the great majority of students are actively participating and disengagement is very rare.

The construction of questions of the style described above requires careful thought and additional time for construction, but this is more than offset in a reduction in marking time. In addition, while answering the question shifts students to higher-level thinking (judgment, evaluation, etc.), marking becomes much simpler and relies on less expertise and judgment on the part of the grader. Through this change we have found only modest differences to the evaluating power of the questions. In the last five years of data, grading dispersion (standard deviation) on questions similar to those in Fig. 3 shifted from roughly 26% in the original format to 18% in the More importantly, the biserial evolved format. coefficient was essentially unaffected, correlation shifting from approximately 0.14 to 0.13.

### 4. CONCLUSIONS

The evidence that teams of students generally outperform the strongest individual student and that immediacy of feedback improves learning supports the use of two-stage exams to capitalize on an opportunity for enhanced learning. In our experience, students have generally responded positively to the practice. Development of assessments that facilitate traditional individual assessment as well as a team-based exercise that contributes to learning and assessment is possible with careful question construction. The benefit of reduced marking effort required to complete the assessment bolsters the case for adoption of this practice. The following is a list of some of the needs for a search and rescue and fast oil spill response vessel for English Bay. Need statements and associated requirements have been developed and some of these contain some errors. Circle and number the 7 most significant errors in the table. For each error you identify, at the space provided at the bottom, provide a short description of why you selected it. 18.

			Ē	aluation Criteria		
Needs	Need Statement	Metric		Accepta	ble Threshold	
		Entity to be measured	Units	Minimum	Maximum	Go / no-go
The vessel has to be able carry and deploy a containment boom and these can be un to 600 m long	The vessel carries and deploys containment booms	Length of boom carried and	metres	N/A	600	N/A
Coast Guard estimates of required top speed indicate 30 knots and it	The vessel top speed should be at least 30 knots	Top Speed	knots	30	N/A	N/A
should be able to reach this speed in 10 seconds	The vessel reaches top speed in 10 seconds or less	Time to Top Speed	seconds	10	N/A	N/A
The vessel carries a crew of at least 4 people and has space for at least 4 rescued people.	The vessel carries 4 to 8 people	Occupant Capacity	# of people	4	œ	N/A
The vessel is capable of reorienting its heading within its hull length (can turn on a vertical axis passing through the middle of the boat)	The vessel has steerable jet- drives	Presence of jet drives	N/A	N/A	N/A	Go
The vessel will be subject to Transport Canada regulations	The vessel complies with Transport Canada regulation	Regulatory Compliance	N/A	Compliance with TC Regulations	N/A	N/A
		5				

9

2

 $\mathfrak{c}$ 

4

Fig. 3. Sample Team Assessment Question

#### References

- Michaelsen, L.K., M. Sweet, M., and D.X. Parmelee, Team-Based Learning: Small Group Learning's Next Big Step. *New Directions for Teaching and Learning*, Jossey-Bass, San Francisco (2008).
- [2] Epstein, M.L., Lazarus, A.D., Calvano, T.B., Matthews, K.A., Hendel, R.A., Epstein, B.B, and G.M. Brosvic, "Immediate Feedback Assessment Technique Promotes Learning and Corrects Inaccurate First Responses," The Psychological Record, 2002, 52, 187-201
- [3] Gilley, B.H., and B. Clarkston, "Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students." *Journal of College Science Teaching* 43.3 (2014): 83-91.
- [4] Cortright, R. N., H.L. Collins, D.W. Rodenbaugh, and S.E. DiCarlo, "Student retention of course content is improved by collaborative-group testing," *Advanced Physiological Education*, 27(3), 2003, 102–108.
- [5] Drouin, M. A., "Group–based formative summative assessment relates to improved student performance and satisfaction," *Teaching of Psychology*, *37*(2), 2010, 114–118.
- [6] Zimbardo, P. G., Butler, L. D., and V.A. Wolfe, "Cooperative college examinations: More gain, less pain when students share information and grades," *The Journal of Experimental Education*, 71(2), 2003, 101– 125.
- [7] Zipp, J.F., "Learning by exams: The impact of twostage cooperative tests," *Teaching Sociology* 35.1, 2007, 62-76.
- [8] Ostafichuk, P.M., E.A. Croft, S.I. Green, G.S. Schajer and S.N. Rogak, "Analysis of Mech 2: An Award-

Winning Second Year Mechanical Engineering Curriculum," Proc. of EE2008, Loughborough, UK, July 2008.

- [9] Sibley, J. and P.M. Ostafichuk, *Getting Started with Team-Based Learning*, Stylus, Sterling, VA (2014).
- [10] Ostafichuk, P.M., Hodgson, A.J., Bartek, S., and Naylor, C., "Teaching Team Dynamics: Experiences in Second Year Mechanical Engineering Design", Proc. CDIO Conference, Montreal, QC, June, 2010.
- [11] Hodgson, A.J., and P.M. Ostafichuk, "Team-Based Learning in the Design Modules of a New, Integrated, 2nd Year Curriculum at UBC," Proc. CDEN, Kananaskis, AB, July, 2005.
- [12] Hodgson, A.J. and P.M. Ostafichuk, "Designing Extended Assignments for Team-Based Learning Modules," Proc. CDEN 2006, Toronto, ON, July, 2006.
- [13] Ostafichuk, P.M., Hodgson, A.J., and M. Fengler, *The Engineering Design Process: An Introduction for Mechanical Engineers*. Vancouver: Ostafichuk, 2013 (2<sup>nd</sup> ed.), 434 pp., {ISBN 978-0-9920587-2-2}
- [14] Brickell, J., Porter, A., Reynolds, M., and R. Cosgrove, "Assigning Students to Groups for Engineering Design Projects: A Comparison of Five Methods," *Journal of Engineering Education*, Vol. 83, Issue 3, 259-62, July (1994).
- [15] Feichtner, S., and E. Davis, "Why Some Groups Fail: A Survey of Students' Experiences with Learning Groups," *The Organizational Behavior Teaching Review*, Vol. 9 No. 4, 58-73, (1984).
- [16] Weimer, M., "Why Groups Fail: Student Answers," *The Teaching Professor*, Vol. 5, No. 9, November (1991).