REVELATIONS OF A COLLABORATIVE MIDTERM IN AN ENGINEERING TECHNOLOGY MANAGEMENT COURSE

Allan MacKenzie

McMaster University, W. Booth School of Engineering Practice & Technology
mackenza@mcmaster.ca

Abstract – *Collaborative testing or two-stage testing as it commonly referred, is where students work together to discuss examination questions and arrive at a consensus to test answers. Two-stage testing is an extension of collaborative learning that has been purported to decrease test anxiety and increase critical thinking, communication, and cooperation skills. Concurrently, there has been considerable research and attention focused on the development of these same transferable skills since employers highly value them. This paper explores the outcomes over two academic terms from administering a two-stage midterm in a mandatory management course for undergraduate students in an engineering technology program. The pedagogical consideration, application to transferable skills, and administration of this type of assessment method are discussed, along with the student perspective regarding the experience and impact on student test performance.*

Keywords: Engineering education, two-stage exams, team exams, test anxiety, collaborative learning, transferable skills, pedagogy, strategic management

1. INTRODUCTION

There is considerable research supporting the use of collaborative group testing across course disciplines [2, 5, 7]. Collaborative testing or two-stage testing as it commonly referred, is where students work together to discuss examination questions and arrive at a consensus to test answers, is an extension of collaborative learning that has been purported to decrease test anxiety and increase critical thinking, communication, and cooperation skills [10]. An appealing aspect of this type of testing is that it potentially transforms the necessary process of assessment into a more significant opportunity for learning [9].

Concurrently, there has been considerable research and attention focused on the development of transferable skills, such as critical thinking, communication, and cooperation since employers highly value them. These skills are so critical that employers identify them as the number one differentiator for job applications in all types of industries, which is consistent across research in Canada, the USA, Australia, and the European Union [4].

Increasing many engineering and technology schools and programs across Canada are under increasing scrutiny to provide a variety of learning experiences and initiatives which can further develop transferable skills to help enhance the employability of their graduates.

If collaborative testing can increase the learning potential of an exam experience and also further develop valuable transferable skills, such as cooperation and communication; why is it still not more of a common pedagogical practice? Mostly because faculty report being afraid that students who have not prepared for the exam may inappropriately benefit from the knowledge of students who have studied [14]. Improved test scores are one of the most commonly reported benefits of collaborative testing [6]. The improvement of test scores, however, is not a universal benefit for all students. It is well established that low-ability students benefit more from collaborative testing than do high-ability students [13]. This was reinforced when administering a two-stage midterm in an undergraduate engineering technology management course.

The objective of this paper is to share the experience of orchestrating a collaborative midterm exam over two academic terms in a mandatory strategic management course delivered to final year undergraduate students in the W. Booth School of Engineering Practice and Technology at McMaster University. The pedagogical considerations, application to transferable skills, and administration of this type of assessment method are discussed, along with the student perspective regarding the experience and impact on student test performance.

2. COURSE CONTEXT

The collaborative midterm in this study was valued at 20 percent of the student's final grade. This was one of the assessments within a final year mandatory undergraduate strategic management capstone course in an engineering technology management program. The students were from one of three areas of specialization: Automation Engineering Technology, Automotive, and Vehicle Engineering Technology or Biotechnology. The students formed into self-selected groups of four members at the beginning of the term in order to participate in a nine-week global online management simulation, along with writing a case analysis report, and completing the

collaborative midterm together, which is the focus of this paper.

This study analyzed the midterm results from cohorts in 2016 and 2017 as shown in Table 1. Students were distributed across three sections, with two sections being taught by the researcher and the other by an adjunct faculty member. The instructors have taught this course together for five years and have collaborated extensively in the development of course materials, assessment instruments and methods of delivery.

Table 1: Strategic Management Student Cohorts

PROGRAM	2016 COHORT				2017 COHORT			
	C01	C02	C03	TOTAL	C01	C02	C03	TOTAL
Automation	1	27	27	55	1		46	47
Automotive	42			42	51			51
Biotechnology		9	12	21		41		41
TOTAL	43	36	39	118	52	41	46	139

The students were given information at the beginning of the term and again during a test review session about the reasons for using a collaborative midterm. The three main reasons shared with the students included the increased learning potential of this type of testing experience, the opportunity to develop transferable skills, such as cooperation and communication, and the prospect of decreased assessment anxiety which often compromises student's ability to perform on exams.

The grading scheme had the individual component of the midterm weighted at 80% and the collaborative group component at 20% for a total test score. Students were assured that their overall midterm score would not be negatively impacted due to the results of the collaborative component. The course instructors agreed that if a student's score on the individual component were better than the combined individual/collaborative test score (i.e., total test score), the individual score achieved would be weighted at 100% for the absolute midterm grade in order to address concerns about fairness with the assessment procedure.

The 2016 cohort Midterm-Part 1 individual component had 44 multiple-choice questions and two short written answer type questions from which the student's selected only one short answer question for a possible six marks. Midterm-Part 1 was a total of 50 marks to be completed within 70 minutes. The 2017 cohort had a similar setup for Midterm-Part 1, with 54 multiple-choice questions and the same option for the short answer section, resulting in an overall total of 60 marks within the same 70-minute time-limit.

After the 70 minutes, the students had a 10-minute break, returning to the same room afterward with their pre-selected group members to complete the collaborative component of the midterm. Once the invigilators established all groups were ready, one copy of the Midterm-Part 2 was distributed to each group to begin the collaborative phase of the test. The groups had 45 minutes

to discuss and answer 30 multiple-choice questions. The multiple-choice selections in Midterm-Part 2 were all different questions from the ones in Midterm-Part 1; however, both midterm parts assessed comparable course topic areas at either the cognitive levels of 'application' or 'analysis' based on Bloom's taxonomy.

3. RESULTS

Four responses were considered and included the score for individual multiple-choice (MC) section, a score for individual short answer question, a total score of the individual assessment (Midterm-Part 1) and the score of the collaborative multiple-choice section (Midterm-Part 2). Test scores are given as a percentage correct for the multiple-choice and short answer questions parts of the midterm. For comparison, a difference score was determined for each student by subtracting the mean percentage correct earned on the tests taken individually from the mean percentage correct earned on the tests taken collaboratively. A difference score greater than zero indicates that a student performed better when working collaboratively than individually, whereas a negative score indicates the reverse. The magnitude of the difference score indicates the strength of that benefit or loss.

3.1 Student Cohort 2016

The 118 student cohort had a mean of 69.49% on the multiple-choice and a mean of 68.83% correct for the short-answer written question. The Midterm-Part 1 had an overall mean score of 69.41% (SD = 8.43) earned under the individual testing condition. A mean score of 72.66% correct (SD = 11.30) was earned under the collaborative condition (Midterm-Part 2). Figure 1 graphically depicts the midterm results of the 2016 cohort.

Figure 1: Boxplot Cohort 2016 Midterm Scores

A paired t-test was used to determine whether there was a statistically significant difference in student performance between collaborative testing and individual testing. The test between the collaborative multiple choice section (Midterm – Part 2) and the total individual score (Midterm – Part 1) resulted in a mean difference of 3.25% (SD = 11.77). As the P-Value of 0.003 is less than 0.05, it can be concluded that the difference in performance is statistically significant in favour of collaborative testing given that the mean difference is greater than 0. These results are shown graphically in Fig. 2.

Figure 2: Collaborative MC vs Individual Total Score 2016

Again, a paired t-test was used to determine whether there was a statistically significant difference in student performance but this time between multiple choice sections only (no short answer). The test resulted in a mean difference of 3.16% (SD = 11.16). As the P-Value of 0.003 is less than 0.05, it can be concluded that the difference in performance is statistically significant in favour of collaborative testing given that the mean difference is greater than 0. These results are shown graphically in Fig. 3.

Figure 3: Collaborative vs Individual MC Score 2016

3.2 Student Cohort 2017

The 139 student cohort had a mean of 70.07% on the multiple-choice and a mean of 67.12% correct for the short-answer written question. The Midterm-Part 1 had an overall mean score of 69.77% (SD = 10.78) earned under the individual testing condition. A mean score of 71.77% correct (SD = 9.10) was earned under the collaborative condition (Midterm-Part 2). Figure 4 graphically depicts the midterm results of the 2017 cohort.

Figure 4: Boxplot Cohort 2017 Midterm Scores

A paired t-test between the collaborative multiple choice section (Midterm – Part 2) and the total individual score (Midterm – Part 1) resulted in a mean difference of 2.01% (SD = 10.03). As the P-Value of 0.02 is less than 0.05, it can be concluded that the difference in performance is statistically significant in favour of collaborative testing given that the mean difference is greater than 0. These results are shown in Fig. 5.

Figure 5: Collaborative MC vs Individual Total Score 2017

A paired t-test between multiple choice sections only (no short answer) resulted in a mean difference of 1.702% (SD = 9.937). As the P-Value of 0.045 is less than 0.05, it can be concluded that the difference in performance is statistically significant in favour of collaborative testing given that the mean difference is greater than 0. These results are shown in Fig. 6.

Figure 6: Collaborative vs individual MC Score 2017

The results from the collaborative midterm used in the strategic management course across both cohorts show improved test score for the majority of the students. Mirroring the pedagogical research about the direct effects of collaborative testing, which are improved test scores for most students [6].

4. STUDENT FEEDBACK

The 2017 student cohort in the strategic management course was asked to participate in a post-test focus group to gather feedback on the experience of completing a collaborative assessment. Participation in the focus group was voluntary and was conducted by researchers at McMaster's MacPherson Institute for Leadership, Innovation, and Excellence in Teaching. The data collected during the focus group was anonymized and returned to course instructors after final grades for the course were submitted.

The participation rate was 59% with a total of 82 students in the strategic management course participating in the focus groups. The researchers visited all three-course sections and facilitated participant discussions about the collaborative midterm. The session was 75 minutes in length of which the students spent the first 15 minutes responding to a survey, followed by 60 minutes of open discussions.

The survey asked for the student's level of agreement with 12 statements using a 5-point Likert scale ranging from 'strongly agree' (5) to 'strongly disagree' (1). The summary survey responses are given in Table 2. Student responses to the questions on the survey were generally

positive towards the collaborative-testing experience. All the affirmative survey question averages are above three which indicates that in most cases the students agreed to the statements given and the benefits described.

Table 2: Cohort 2017 Evaluation of Collaborative Testing

Question	Mean	SD
Writing my midterm test using the two-stage test format increased my understanding of the course material.	3.256	1.016
The two-stage format lessened the amount of stress (or nervousness) I typically feel when writing tests or exams.	3.354	1.023
I would like to write two-stage midterms in other B.Tech. courses as well.	3.366	1.222
Answering test questions as a group helped me understand what was being asked more clearly.	3.61	0.966
Answering test questions as a group increased my confidence in the answers we submitted.	3.268	1.207
I came away with a better understanding of the course concepts having written the midterm in a group.	3.159	0.987
The two-stage test format helped me understand why each answer was correct.	3.256	1.016
The group test writing format helped me appreciate other perspectives (understanding) on our course content.	3.5	1.033
The group test writing format helped me remember information that I had (personally) forgotten during my individual midterm component.	3.61	1.051
The group test writing format helped develop good working relationships with my classmates.	3.427	0.956
The two-stage midterm format is better than having one independently written midterm (without a team component).	3.244	1.117
Answering test questions as a group hurt my ability to submit a correct answer due to the need for a majority consensus of the group members.	3.256	1.12

When asked during the focus group discussions about the benefits of writing a midterm in two stages, students noted several positive aspects, primarily in terms of stress. In each section, the majority felt less stressed going into the midterm knowing that they would have a chance to collaborate on the group component. Many students felt that they learned more because of this format. Given a chance to go over the answers with their peers, who may have different areas of knowledge, allowed them to learn the correct answers to questions they had been unsure about. Students felt that the collaborative component was more enjoyable because it could only help their mark, although some students felt the potential boost was not worth the extra effort required during the collaborative part of the midterm.

The students indicated there were a few drawbacks too with the two-stage midterm. Some students felt that there was more bickering than learning, and students noted that because they did not get the answers to the midterm, they were still not sure of the correct answers to some questions. Most of the drawbacks mentioned were about the administrative aspects of the midterm. Noise in the testing room was a common complaint: many groups found it distracting and would have preferred having smaller, quieter rooms. There was also a consensus that multiple-choice was not the best format for a collaborative group test because they could choose one

answer, leading to the process becoming more argumentative than collaborative. Students stated they would have preferred short written answer type questions in the collaborative section so they could have strived to combine multiple opinions rather than choosing one over another. They felt that in some cases the test came down to who could argue the best, rather than who had the best answer, which meant the less assertive group members were not always heard or valued.

When asked about the experience with writing a portion of the midterm as a group, most students felt that their experience was impacted by test fatigue. After finishing the individual component, many stated they wanted to leave but had to stay for the collaborative group component. This fatigue diminished the quality of discussion amongst some student groups. Many students also mentioned group dynamics and trust as a significant factor in how their group performed on the collaborative portion of the test, and there was general agreement that a collaborative type testing would not be as effective in a first or second-year course because the group dynamics could be worse.

Asked about whether they prefer the two-stage midterm format when learning about business strategy and tactics versus courses where the content focuses more on quantitative or technical knowledge there was some disagreement among the students in the focus groups. Proponents of two-stage midterms in quantitative courses noted that it is easier to agree on an answer when there is a way to prove the answer is correct, while others noted there is less opportunity for collaboration on those types of questions. On the other hand, some students stated that business strategy and tactics would allow for better discussion and collaboration; however indicated still has the potential to degrade into an argument between two competing answers. The one thing that almost all students agreed on, as noted earlier, was that written answers would have been more conducive to a positive group experience during the collaborative portion of the midterm. Answering multiple choice questions meant the group discussion was more about trying to convince a majority of the group members to agree rather than genuine discussion and participatory solution finding.

5. DISCUSSION

5.1 Administrative Landscape

There are particular administrative challenges identified with collaborative testing. The overwhelming issue raised by the students was the noise level in the testing room during the collaborative component of the midterm. Many students indicated they found this particularly distracting and wished there were fewer groups in the room to help diminish the noise level. A recent literature review summarizes that noise exposure

does have deleterious effects on cognitive performance but that the magnitude of these effects is dependent on various factors such as noise intensity, noise duration, and task performance [12].

There is considerable sound environmental variation in two-stage assessments as students start 'heads-down' in a quiet task focus atmosphere during individual testing then transition to a 'heads-up' mode during the elevated discussions within the collaborative phase. Does this type of collaborative testing environment transition leave some students disadvantaged? Alternatively, does this moderate level of ambient noise become conducive to providing an atmosphere of creative cognition for the student groups [8]? Would it be advantageous for faculty using collaborative testing to adopt the current workplace strategy trend of 'activity-based working' where employees choose between varieties of different workspaces, each designed for a specific activity? However, given the facility constraints in most educational institutions is it even realistic to have a variety of rooms available to administer different phases of collaborative tests?

The other related issue is some students reported experiencing test fatigue with having to complete the midterm in two-stages. However, according to a study by Ackerman and Kanfer, spending hours taking a high-pressure aptitude test may make people feel mentally fatigued, but that fatigue does not necessarily lead to lower test scores. The authors of the study proposed that cognitive fatigue appears to be more about the individual's expectations of and prior experience with testing than about the length of the test [1]. These conclusions suggest that collaborative testing may be beneficial to some students and not for others with test fatigue based more on individual trait complexes, than by task requirements and format of the test.

5.2 Test Scores and Who Benefits?

In this study, there is statistical evidence of a difference in performance with a collaborative midterm. Practically speaking, 59% of the students in 2016 cohort and 55% in 2017 cohort performed better by working collaboratively based on the total midterm score. The research supports this result that low-ability students benefit more from collaborative testing than do high-ability students [13]. On the flip side, however, it means that 41% of the students in 2016 and 45% in 2017 had to have their midterm score adjusted so they would not be negatively impacted because of the collaborative component.

Begs the question – is this enough? What performance improvement threshold would be acceptable for collaborative testing to be a beneficial learning and assessment mechanism? Should the performance improvement threshold vary depending on course level or

type of material being assessed? Alternatively, perhaps it is the particular type of questions being used in the assessment, such as multiple-choice, short-answer or essay questions? These are real pedagogical quandaries that instructors must take into consideration when deciding whether to employ collaborative testing within their courses or programs.

5.3 Transferable Skill Development

Considerable attention is focused on the development of transferable skills, such as critical thinking, communication, and cooperation, which is driven by the high value placed on these skills by employers. Report of 500 Canadian business leaders stated they would hire and train employees with some gaps in technical expertise if they had the soft skills for the position. The report also indicated it was difficult for these same firms to find candidates with the necessary soft skills. Despite this apparent transferable skills gap, over 40 percent of the leader's suggested that prospective candidates need to develop their soft skills before entering the competitive labor market [3]. Can collaborative testing be used as a practice in education to address this gap in employability preparedness for future graduates?

Have to start with and consider the context of students in engineering technology courses and programs. Do the students have the necessary foundational soft skills to exploit when working with others on a collaborative test? Group dynamics are a critical component in the performance and outcomes of collaborative testing. Assumptions are most likely being made that students have the prerequisite transferable skills and attributes, as well as deploying them appropriately to enhance the group's learning experience. It leads to asking whether students are collaborating on collaborative assessments. Alternatively, are the most vocal and assertive students in the group thereby convincing the other members about their particular answer selection rather than striving to create a group dynamic that enables a participatory process for collective solution finding among all the members. As identified in this study, some students felt the test came down to who could argue the best, rather than who had the best answer. As well, less assertive group members were not always heard or valued. This observation does not seem like collaboration in action.

Back to the earlier question - can collaborative testing be used as a practice in education to address employability preparedness for future graduates? The researcher advocates that a collaborative test should be partnered with collaborative learning to leverage this pedagogical opportunity. Collaborative learning has been shown to improve the acquisition of new skills and knowledge, enhance critical thinking, maintain persistence in problem-solving, enhance cooperative work, and increase student involvement [11].

Consideration must be given to the limitation of collaborative testing as a bolt-on type practice to improve the acquisition or deployment of transferable skills. Think it is best to consider it as a generative instrument of collaborative learning to assess and enable its benefits.

6. CONCLUSION

The specifics were shared from orchestrating a collaborative midterm over two academic terms in an undergraduate strategic management course. The administration of this type of assessment method was discussed, along with pedagogical considerations and the student perspective regarding the experience and impact on student test performance. The application to transferable skills was explored, leading to concerns of collaborative testing being deployed as a stand-alone mechanism to enable the attainment of transferable skills in engineering technology undergraduate students.

Acknowledgments

The author would like to thank the researchers at McMaster's MacPherson Institute for their work in facilitating focus group mentioned in this paper. Also, I wish to acknowledge my colleague K. Lawrence for the review of and feedback on this paper.

References

- [1] Phillip L. Ackerman and Ruth Kanfer, "Test Length and Cognitive Fatigue: An Empirical Examination of Effects on Performance and Test-Taker Reactions." *Journal of Experimental Psychology*, vol. 15, no. 2, pp. 163–181, 2009.
- [2] John Berry and Melvin A. Nyman, "Small-group assessment methods in mathematics," *International Journal of Mathematical Education in Science and Technology*, vol. 3, no. 5, pp. 641–649, 2002.
- [3] "Career Development in the Canadian Workplace: National Business Survey". Available as of January 21, 2014, from www.ceric.ca/perspectives
- [4] Wendy Cukier, Jaigris Hodson and Aisha Omar, "Soft Skills are Hard: a Review of the Literature". Available as of November 30, 2015 from https://www.ryerson.ca/content/dam/diversity/reports/KSG_2015_SoftSkills_FullReport.pdf
- [5] Jennifer D. Ewald, "Language-related episodes in an assessment context quiz," *The Canadian Modern Language Review*, vol. 61, no 4, pp. 565-586, 2005.

- [6] Brett H. Gilley and Bridgette Clarkston, "Collaborative Testing: Evidence of Learning in a Controlled In-Class Study of Undergraduate Students," *Journal of College Science Teaching*, vol. 43, no. 3, pp. 83-91, 2014.
- [7] Mary M. Hoke and Leslie K. Robbins, "The Impact of Active Learning on Nursing Students, Clinical Success," *Journal of Holistic Nursing*, vol. 23, no. 3, pp. 348–355, 2005.
- [8] Ravi Mehta, Rui Zhu, and Amar Cheema. "Is Noise Always Bad? Exploring the Effects of Ambient Noise on Creative Cognition," *Journal of Consumer Research*, vol. 39, no. 4, pp. 784-799, 2012
- [9] Carol Pandey and Susan Kapitanoff, "The influence of anxiety and quality of interaction on collaborative test performance." *Active Learning in Higher Education*, vol. 12, no. 3, pp.163–174, 2011.
- [10] Antonio Russo and Susan H. Warren, "Collaborative test taking," *College Teaching*, vol. 47, no. 1, pp. 18–20, 1999.
- [11] Robert E. Slavin, "Synthesis of research on cooperative learning," *Educational Leadership*, vol. 8, no. 5, pp. 71–82, 1991
- [12] James L. Szalma and Peter A. Hancock, "Noise Effects on Human Performance: A Meta-Analytic Synthesis," *Psychological Bulletin*, vol. 137, no. 4, pp. 682–707, 2011
- [13] Z. Bilsky-Torra, ""Son of S.T.A.D.": A comparison of group quiz and individual test scores," *English Teachers' Journal*, vol. 46, pp.35-43, 1993.
- [14] Maryellen Weimer, "The Effects of Collaborative Testing," Available as of March 1, 2013 from <https://www.facultyfocus.com/articles/educational-assessment/the-effects-of-collaborative-testing/>