IMPLEMENTATION OF A TWO-STAGE EXAM IN FIRST-YEAR PHYSICAL CHEMISTRY COURSES OR INTERNATIONAL ENGINEERING STUDENTS

Roza Vaez Ghaemi¹ and Gabriel Potvin^{1,2}

¹Department of Chemical and Biological Engineering, University of British Columbia

²Applied Science Stream, Vantage College, University of British Columbia

ghaemi@chbe.ubc.ca, gabriel.potvin@ubc.ca

Abstract – *Group exams have been shown to improve student performance, retention of material, and teamwork and communication skills. This paper assesses the opinion of students regarding group exams, and their perception of potential benefits and impacts on their learning, before and after having participated in one. Both a traditional and two-stage exam were performed in first-year physical chemistry courses in the engineering stream of Vantage College at UBC, which means the participating cohort is entirely composed of international students with a range of English-language communication skills.*

The overall experience of students with group exams, based on survey responses, was positive, and the large majority of students indicated they would like to continue using this format of exam in the future. The perception of group exams improved before and after having written one, with the students initially overestimating the difficulty, stress, and level of conflict associated with this process. Some students indicated less confidence that peer learning helped them improve their performance after having written the exam, but further study to elucidate the significance and the causes of this result.

Keywords: Active Learning, Group Exam, Teamwork, International Students.

1. INTRODUCTION

Active learning and participation-based activities have been shown to lead to significant improvements in student learning, particularly for students enrolled in science, technology, engineering and mathematics (STEM) programs [1]. Collaborative tests or group exams, one form of these participation-based activities, have been shown to improve student performance [2-6], enhance student learning [7], promote teamwork and communication skills [8] and improve students' perception of the material [9]. These collaborative tests most often take the form of two-stage group exams, in which students first complete a traditional individual

exam, worth the majority of their exam mark, and immediately afterwards complete the same or a very similar exam in groups, during which time they have the opportunity to discuss the material and receive immediate feedback from their peers [10]. These discussions during the group portions of the exam have been shown to present valuable opportunities for students to share their knowledge [11], practice justifying their reasoning and thought-process for each answer [12], learn from their peers, and practice their communication skills [13][14].

This paper presents the implementation of a two-stage midterm exam in a first-year engineering chemistry course attended exclusively by international students as part of the Vantage College (VC) program at the University of British Columbia. VC was established with a mandate to offer innovative interdisciplinary first-year programs to international students that integrate core content courses with discipline-specific language instruction and support. These programs are offered to international students who are in outstanding academic standing but whose language proficiency level falls just below that set by UBC for direct entry into its programs. Participants take all their first-year courses with their Vantage cohort, and, upon successful completion of the program, are considered to have met UBC's academic and language proficiency requirements, and join their "direct-entry" counterparts in second year and for the remainder of their undergraduate studies. Given the particular language barriers that exist for these students, active discussions in randomly assigned multi-lingual groups during group exams may, in addition to the benefits listed above, provide an opportunity to students to practice their technical and argumentative communication skills.

Very little work has been reported on international students' perception of group exams. Given that all students in the engineering stream of VC must take the physical chemistry courses APSC 182 (Matter and Energy I) and APSC 183 (Matter and Energy II) in terms 1 and 2 respectively, and that these courses follow the same structure and are taught by the same instructor, this

provides a good opportunity to compare student performance on and student perception of traditional and group exams in a relatively uniform environment. This paper evaluates some of these aspects in the first term (APSC 182) for a traditional midterm and in the second term (APSC 183) for a group midterm, through a comparison of performance through exam scores and of student opinions of these exams through responses to surveys.

2. METHODOLOGY

Students' perception and experience with two-stage group exams were assessed through surveys in both terms of the academic year. All respondents were first-year engineering students enrolled in APSC 182 and 183 as part of the VC program, and were part of a cohort of 90 students. APSC 182 and 183, taught by the same instructor, are both composed of lectures, interactive problem-solving sessions, and laboratory sessions. In both

terms, students received a problem-solving review session a few days before their midterm. A survey containing a series of questions classified into three categories (shown in Table 1) was sent to students after having written, but before receiving the marks for, a traditional individual midterm in term 1, and again after completing a two-stage group midterm in term 2. For the group part of the second term exam, students were randomly assigned into groups of 4, and were informed of their groups prior to the midterm. The group portion was worth 20% of the overall midterm mark, but students received a guarantee that if their individual scores were higher than their group scores, their individual mark would count for 100% of the midterm mark.

The students were encouraged to take the survey within two days of writing the midterm, with an incentive of receiving a 1% bonus mark on the exam if more than 70% of the class participated.

Table 1. Survey questions about the group exam impact

	Term 1 questions		Term 2 questions
	<i>One way to write midterms is to use a 2-stage exam approach, in which students take the exam individually and then retake the same or a very similar exam immediately afterwards in a group. Group members work together to answer each question. The individual part is worth ~80% and the group part is worth ~20% of the exam mark for an individual student. Based on the description of this 2-stage exam, please rank your agreement with the following statements on a scale of 1-5.</i>		<i>The individual part is worth ~80% and the group part is worth ~20% of your exam mark. Based on your experience with 2-stage exam, please rank your agreement with the following statements on a scale of 1-5.</i>
Team work	Q1	If I was writing the exam in a group, unanimity must be reached about each answer (all team members must agree on the final answer).	When I was retaking the exam in a group, unanimity was reached about each answer (all team members agreed on the final answer).
	Q2	Peer teaching (learning from my peers) during group exams could be an effective method for improving my understanding of the material.	Peer teaching (learning from my peers) during group exams was an effective method for improving my understanding of the material.
	Q3	While working on APSC 182 material, I often asked my classmates questions when I did not understand how to solve a problem.	While working on APSC 183 material, I often asked my classmates questions when I did not understand how to solve a problem.
	Q4	A group retake of exam would improve my grade	The group retake of the midterm is going to improve my grade
Potential negative impacts	Q5	Hostility can emerge between group members during group exams.	Hostility emerged between group members during the group exam.
	Q6	I would feel pressured by peers (for any reason) during group exams	I felt pressured by peers (for any reason) during the group exam.
	Q7	It could be stressful to try to establish consensus (all agree) among the group members.	It was stressful to try to establish consensus (all agree) among the group members.
Learning objectives	Q8	In general, the exam retake would help me retain (remember for a long time) the material.	In general, the exam retake helped me retain (remember for a long time) the material.
	Q9	In general, the exam retake would help me with comprehension of the course material.	In general, the exam retake helped me with comprehension of the course material.
	Q10	In general, the exam retake would help me apply the course material.	In general, the exam retake helped me apply the course material.
	Q11	I would like to try this type of exam.	I would like to do 2-stage exams more often.

Both surveys were completed anonymously, and students were explicitly asked to provide consent for their responses to be aggregated and used in this paper, with a clear indication that their participation still counted towards the participation threshold to receive the bonus mark, even if they refused that their data be used. In order to be able to directly compare each individual's attitude and perception of the two-stage exam, a unique user-generated 8-digit identifier code was produced by each student, which ensured the anonymity of the survey but enabled matching the responses from both surveys.

Prior to attending VC, none of the students had ever participated in a group exam. In the first term survey, the students were introduced to the concept of group exams and were asked to share their opinion on this type of evaluation. During the second term, after having actually written the group exam, the students were asked to reflect on their experience and answer similar questions about two-stage exams. The survey consisted of 5-point Likert-scale questions, which are listed in Table 1. Each survey also contained a quality assurance question specifying a correct answer, to ensure the validity of the data.

There were 3 categories of questions, each targeting different aspects of the group exam experience. The first category focuses on team work, the second addresses some of the potential downsides of group work during exams, and the third asks students to evaluate the impact of the two-stage format on their confidence remembering, understanding and applying the technical material covered in class. Each of these categories is discussed in turn.

3. RESULTS AND DISCUSSION

3.1. Survey Responses

For the cohort of 90 students, the response rate was 76% in term 1 ($n = 69$) and 71% in term 2 ($n = 63$). From the collected data, 56 completed surveys in term 1 and 39 responses in term 2 met both the validation criteria (correctly answered the quality assurance question) and gave their permission for the data to be used, with the rest of the data being discarded. Out of the usable data, based on the unique 8-digit identifiers, only 13 respondents correctly completed both surveys and gave permission to use their data both times.

For the first category of questions, students were asked about their perception of group discussions, peer teaching, and the impact of 2-stage exams on their grades. In the second category of questions, they were asked to rank their perception of, or their experience with, potential downsides of group exams, including peer-pressure or stress. Finally in the third category, students were asked questions on course learning objectives at different levels of Bloom's Taxonomy, and whether they believed the portion of the exam would or did make any difference in their level of content retention or understanding.

3.2. Team Work

In term 1, 66% of students reported a belief that peer collaboration during a group exam would be an effective method to enhance their understanding, and after writing a group exam in term 2, that number increase to 74% ($n=39$). Results are shown in Figure 1. On average peer teaching received a score of 4.05 ± 0.6 (out of 5) after the second midterm, which represents a 6% increase compared to the first midterm. Similar upward trends were reported in other work [13]. On average, 71% of students reported the belief that unanimity was reached/could be reached during a two-stage exam. The unanimity score was 3.9 ± 1 with a 0.8% decrease after the group retake of the exam. Given the small number of respondents, this difference is not statistically significant, but in the future it would be interesting, given the composition of this particular cohort, to investigate whether communication skills and proficiency with the English language impacted this perception. Before taking the two-stage exam, 73% of the students thought that the group retake of the exam could potentially improve their grades, but after the group exam 89% of them reported that they found the group retake to be a helpful strategy to enhance their performance. This agrees with previous studies that suggest a lower level of test anxiety in group exams as a reason for improved performance [11]. Interestingly, and contrary to other reports of group exams, after compensating for marking discrepancies between markers, just over 10% of students did better on the individual portion of the exam than on the group portion (data not shown). In informal follow-ups with these students, many indicated a lack of confidence in their answers, and a tendency to defer to other students claiming to know how to correctly answer a given question. One student also reported that although she

Fig. 1. The weighted average of the impact of group two-stage exam on the team work. Error bars shows weighted standard deviation.

suspected an answer provided by a team mate was wrong, she was unable to articulate why the answer was incorrect, and she was overruled by her team. Further work is of course necessary to determine whether the later situation is simply anecdotal, or if the wide range of English-language proficiency in the VC program

could explain why strong performance on individual portions fails to translate to the group portion in some cases.

3.3. Potential Negative Impacts

It is interesting to note that during the group portion of the exam, 38% of the students reported group conflict being an issue, with the same proportion of students reporting everything went well, which explains the average ‘hostility’ score of 2.8 ± 2.02 . This value represents a 9% decrease compared to the number of students indicating conflict may be an issue in the term 1 survey. Results are shown in Figure 2. Only 15% of students reported being under pressure from their teammates during the second midterm, which was less than half the number that reported this may be an issue during the first survey. The perception that establishing consensus would create stress decreased considerably, going from 3.2 ± 1.2 to 2.4 ± 1.0 . Only 12% of the students found it stressful to have to reach an agreement between all group members during the two-stage exam, compared to 32% believing it would be an issue before trying it. Although here again, the small number of respondents makes it difficult to extract statistically significant data from the results, an overall decrease in the “Negative Aspects” categories suggest that the experience of the group midterm was more positive than was initially perceived. It would be interesting, in the future, to explore questions of group work anxiety in this setting, whether in general or for this cohort in particular.

Fig. 2. The weighted average of the potential negative impacts of group two-stage exam. Error bars shows weighted standard deviation.

3.4. Learning Objectives

Before trying the two-stage exam, 66% of the students reported thinking that this type of exam would be a helpful tool to improve their content retention, and after writing the second midterm, 87% of respondents indicated that it had that effect. Results are shown in Figure 3. 92% of the respondents also indicated, after having written the group exam, that it improved their comprehension of the

course material. 87% of the students found the two-stage exam effective at helping them understand how to apply their learning to solve problems in the course, compared to 64% who indicated they thought it would have that effect in the first survey. Application capability received a score of 4.05 ± 0.6 in the ranking scale which shows 0.3 point increase from the initial perception. This is in line with many studies that have reported an enhancement of retention and learning as a result of two-stage exams [2].

Fig. 3. The weighted average of the impact of group two-stage exam on students' learning. Error bars shows weighted standard deviation.

3.5. Individual Comparison

As previously discussed, the following directly compares individual responses for the first and second midterm, with responses identified by their 8-digit code. Only 13 respondents successfully completed both surveys, using the same identifying code, and provided permission to use the data. The average difference between the responses between the first and second surveys is shown in Fig. 4.

Interestingly, after the group midterm there was a 46% increase in the number of students reporting that reaching unanimity as a group was important, compared to the answers provided in the first survey. This highlights the difference in perception before and after having gone through the group exam, as results suggest either the students underestimated the discussion that would arise during this process, or simply did not give the survey responses much thought. Also of particular interest is a decrease of 15% in the perceived value of peer teaching to improve one's score, compared to the initial survey. Although this is a phenomenon worth studying further in the future, it could simply be a result of the small sample size. Given the data analyzed here is from students who agreed to complete both of the optional surveys, correctly answered the quality assurance questions both times which suggests they took the time to read the questions and provide thoughtful answers, and correctly followed the instructions to generate the 8-digit identifying code, it is conceivable there is a bias towards students on the

stronger end of the spectrum in the provided responses. If this is the case, it is possible these students led their team during the group portion of the exam, or simply did not get as much out of their group experience than a weaker student benefitting from the guidance of a stronger one, which could lead to a less enthusiastic rating on the second survey. 38% of respondents found group exams more helpful than they thought to improve their grades, which, in conjunction with the previous point, suggests a separation of ‘grade’ and ‘understanding’ in the minds of the students.

The responses for the questions related to the negative impacts of the group exam all showed a decrease in scores in the second survey, suggesting once again that students may have overestimated the difficulty of working with random teammates in this type of setting. 30% of individuals had drastic (2 points or higher) change in their opinion about the potential hostility in the group exam after experiencing it. Similarly, 30% of individuals reported experiencing less pressure and stress than they previously thought.

Lastly, modest increases of 0.3, 0.4 and 0.1 points on the group exam’s contribution to retention, comprehension and application of the material were observed. This is interesting given the decrease in the perception of the impact of peer teaching on their grade. This suggests that perhaps students saw value in repeating the same exam twice, and attributed an increase in learning to that, and not the collaborative aspect. This will be investigated in a future focus group.

Fig. 4. Average shift in individuals' perception after experiencing two-stage exam

4. CONCLUSION AND FUTURE WORK

The survey responses to the questions evaluating students' perception of group exams suggest that in general they found the group exam helped improve their retention, comprehension and application of the course material. It is important to note that these capabilities were not assessed directly, and the analysis is only done on the basis of what students felt after the midterm.

79% of the students indicated they would want to try this format of group exam again in their future courses, and 87% of them found it to be a helpful educational tool. Considering the international demographic of students in the Vantage college at UBC, it would be interesting to see how these group exams would help these students further develop their communication skills. It would be also interesting to have a more comprehensive evaluation of the impact of group exam on student confidence with course material as well as student perception of the course in general.

Acknowledgements

The authors would like to express their gratitude to the facilitators of the *Certificate in Advanced Teaching and Learning (CATL)* program at UBC's Center for Teaching Learning and Technology (CTLT) for their support and insightful feedback. Special thanks are offered to Dr. Eric Cytrynbaum for his kind advice and for sharing his experiences in the implementation of group exams.

References

- [1] B. Bruno *et al.*, "Two-Stage Exams: A Powerful Tool for Reducing the Achievement Gap in Undergraduate Oceanography and Geology Classes," *Oceanography*, vol. 30, no. 2, 2017.
- [2] B. H. Gilley and B. Clarkston, "Collaborative Testing: Evidence of Learning in a Controlled In-Class Study of Undergraduate Students.," *J. Coll. Sci. Teach.*, vol. 43, no. 3, pp. 83–91, 2014.
- [3] K. Knierim, H. Turner, and R. K. Davis, "Two-Stage Exams Improve Student Learning in an Introductory Geology Course: Logistics, Attendance, and Grades," *J. Geosci. Educ.*, vol. 63, no. 2, pp. 157–164, 2015.
- [4] K. T. Wissman and K. A. Rawson, "Why does collaborative retrieval improve memory? Enhanced relational and item-specific processing," *J. Mem. Lang.*, vol. 84, pp. 75–87, 2015.
- [5] R. A. Green, T. Cates, L. White, and D. Farchione, "Do collaborative practical tests encourage student-centered active learning of gross anatomy?," *Anat. Sci. Educ.*, vol. 9, no. 3, pp. 231–237, 2016.
- [6] J. E. Lindsley, D. A. Morton, K. Pippitt, S. Lamb, and J. M. Colbert-Getz, "The two-stage examination: A method to assess individual competence and collaborative problem solving in medical students," *Acad. Med.*, vol. 91, no. 10, pp. 1384–1387, 2016.
- [7] J. F. Zipp, "Learning by Exams : The Impact of Two-Stage Cooperative Tests Author (s) : John F . Zipp Published by : American Sociological Association Stable URL :

- <http://www.jstor.org/stable/20058530>
LEARNING BY EXAMS: THE IMPACT OF TWO-STAGE COOPERATIVE TESTS *,” vol. 35, no. 1, pp. 62–76, 2017.
- [8] D. Rumble *et al.*, “Data on Continental Carbonate Precipitates To Ground-Truth Flux-Balance Models,” *Terra Nov.*, no. January, pp. 122–124, 2009.
- [9] B. G. W. Rieger and C. E. Heiner, “Examinations That Support Collaborative Learning: The Students’ Perspective,” *J. Coll. Sci. Teach.*, vol. 43, no. 4, pp. 41–48, 2014.
- [10] J. Ives, M. de J. Van Lier, N. K. Sumah, and J. B. Stang, “Examining student participation in two-phase collaborative exams through video analysis,” *2016 Phys. Educ. Res. Conf. Proc.*, pp. 172–175, 2016.
- [11] K. A. Fournier, J. Couret, J. B. Ramsay, and J. L. Caulkins, “Using collaborative two-stage examinations to address test anxiety in a large enrollment gateway course,” *Anat. Sci. Educ.*, vol. 10, no. 5, pp. 409–422, 2017.
- [12] G. J. Sinner, J. C. Briggs, F. T. Stevenson, and S. J. Nazian, “Group Testing in Medical Education: An Assessment of Group Dynamics, Student Acceptance, and Effect on Student Performance,” *Med. Sci. Educ.*, vol. 23, no. 3, pp. 346–354, 2014.
- [13] R. Nafziger, J. K. Meseke, and C. a Meseke, “The Effect of Group Formation Process on Overall Student Performance L,” *Agenda*, vol. 25, no. 1, pp. 11–15, 2011.
- [14] D. P. Larsen, A. C. Butler, and H. L. Roediger, “Test-enhanced learning in medical education,” *Med. Educ.*, vol. 42, no. 10, pp. 959–966, 2008.