

Examinations That Support Collaborative Learning: The Students' Perspective

By Georg W. Rieger and Cynthia E. Heiner

We used surveys and classroom observations to examine student reactions to two-stage exams, where students first do the exam individually and then redo it collaboratively. Our results show why both students and instructors appreciate this examination format: Two-stage collaborative examinations are relatively easy to implement, have a high potential for learning, and support the collaborative learning approach used in many sciences classes. A look at survey data from an introductory physics class shows that a vast majority of students (76%) had a positive opinion of this exam format (expressed in 236 comments) whereas only 10% expressed an overall negative opinion in 30 negative statements. Most of the positive comments relate to how this benefits learning. In this article, we describe how to implement two-stage exams, discuss advantages and disadvantages, and present the students' view.

University instructors increasingly use interactive engagement and social/collaborative learning methods in their science classes to achieve better learning outcomes (National Research Council, 2012). Such methods result in deeper engagement by the students and incorporate more formative assessment to support learning. A number of research-based methods, such as peer instruction (Mazur, 1997), think-pair-share (Johnson, Johnson, & Smith, 2011), and cooperative group problem solving (Heller & Hollabaugh, 1992; Heller, Keith, & Anderson, 1992), share some basic features that are recognized to support learning across a wide range of contexts. These features include intense engagement by students, collaborative learning where students develop their thinking, and immediate feedback through the interactions with their peers (National Research Council, 2012). In this article we discuss an exam format—two-stage exams—that uses these same features.

Frequently, collaborative learning and formative assessment will be used in classroom instruction, but the course exams will remain in the traditional format in which students solve problems in isolation and only receive feedback several days later. Exams send very powerful messages, and such an exam format does not support the message that collaborative learning is important.

Moreover, although individual exams produce a uniquely intense engagement with the material, that engagement provides little or no contribution to learning—defined as acquiring new ideas—because of the lack of timely and useful feedback (Black & Wiliam, 1998).

The two-stage exam is a relatively simple way to solve these problems. In a two-stage exam, students first complete and turn in the exam individually and then, working in small groups, answer the exam questions again. This makes the exam itself a valuable learning experience while also sending a consistent message to the students as to the worth of collaborative learning. We see indications that the use of this exam format goes beyond ensuring consistency across course components, in that it positively impacts how students approach the other collaborative components in the course. The two-stage exam accomplishes this while still providing summative assessment of individual performance.

Collaborative tests have been used for some time in a variety of formats (see summaries in Leight, Saunders, Calkins, & Withers, 2012; Zipp, 2007). The two-stage format discussed in this article (sometimes referred to as *group test* Cortright, Collins, Rodenbaugh, & DiCarlo, 2003; or *pyramid exam*, Cohen & Henle, 1995) has also been used in the past, in particular in team-based learning as part of the readiness assurance process (see

e.g., <http://www.teambasedlearning.org>). This process, which uses scratch-and-win type testing cards during the group part to reveal the answers to all questions, follows up on the assigned reading, and provides a low-stakes way to ensure that students have the background knowledge necessary for the problem-solving activities that follow. However, administering high-stakes examinations such as midterm or final examinations in a two-stage format is still relatively rare. Stearns (1996), for example, mentions increased student performance on the (individual) final exam in a research method and statistics class after taking the midterm exams in a two-stage format, as well as decreased dropout rates, higher enjoyment of the course, and increased collaborative skills. Only a few studies have attempted to measure the benefits of two-stage exams on learning in science: In a recent study, Gilley and Clarkston (2014) reported knowledge gains (increases in student learning, i.e., the original acquisition of knowledge by students) due to the collaborative part of the exam in a science course on natural disasters, whereas other studies in biology (Leight et al., 2012) and physiology (Cortright et al., 2003) have focused on the retention of content. A positive impact on student motivation, reduced test anxiety, increased collaborative skills, and improved perception of the course were also mentioned in a number of other studies (see references in Gilley & Clarkston, 2014; Leight et al., 2012; Zipp, 2007). Potential limitations of two-stage exams are a reduced number of questions on the tests (to make time for the group portion) and a slightly higher administrative effort. In addition, differences in group composition may limit the effectiveness of this approach in groups with one dominant student or in groups with

free-riders (see discussion in Zipp, 2007). Our survey results, however, indicate that this occurs only in a small number of groups.

This article was inspired by seeing both the success of two-stage exams and how popular they have been with both students and instructors across the Faculty of Science at the University of British Columbia (UBC). This exam format was first introduced in the UBC Faculty of Science 3 years ago and is now being used in at least 20 science courses. The faculty members value the widespread intense engagement by their students during the second stage of the exam, and as discussed below, students see them as valuable learning experiences. Next, we describe how to implement two-stage exams, discuss their benefits, and present the students' view.

Implementation of two-stage exams

The particular format of the two-stage exam we use is relatively easy to implement and has worked well in UBC science courses.

- Stage 1 (individual, between 3/4 and 2/3 of the examination time): This is a standard formal examination students complete working alone.
- Stage 2 (small groups, remainder of the examination time): The group portion begins after all individual exams are collected. Students work in groups of three or four students on (mostly) the same problems as in the individual portion (Figure 1). They must come to a consensus on the answers and hand in one copy with the names and student ID numbers of all group members. Because students have already seen each problem during Stage 1, solving the same problems again in Stage 2 usually takes much less time than in Stage 1, includ-

ing the time for discussions and agreeing on a solution.

As an example, the two-stage exam given in our introductory physics course ($N = 178$) had a total duration of 90 minutes that was split as 55 minutes for individual effort (Stage 1) and 30 minutes for group effort (Stage 2), with 5 minutes in between for making the switch from Stage 1 to Stage 2. During the switch, instructors and teaching assistants first collected the individual exam copies, and then students were instructed to sit with their predetermined group members (3–4 students per group). In some courses, these groups are preformed (e.g., same as collaborative groups in class or groups put together by the instructor), whereas in other courses, students are free to choose their groups. Once the groups were assembled, the second part of the exam was distributed. Generally the switch can be done in less than 5 minutes—even in large classes, if there is at least one instructor or teaching assistant for 50 students.

A two-stage exam in a 50-minute lecture time slot is doable, but having a 90-minute time slot is easier. In some courses, instructors have replaced their 50-minute in-class midterm exams with 90-minute evening exams, so that similar content can be covered. Concerns about the length of an exam can be addressed by repeating only the conceptual questions of the individual part in the group portion and/or by turning short-answer questions of the individual part into multiple choice or ranking tasks in the group portion; see Figure 1 for examples.

Grades from the individual and the group portion are combined for the total examination mark, weighted between 75% to 90% for the individual portion and 25% to 10%, respectively, for the group portion. The group exam score has no effect on the differentiation between stu-

dents (i.e., a student's performance relative to the class), yet even the small weight of the group portion provides sufficient motivation for students to take this part seriously. For example, an 85/15 (individual/group) split used in our physics class

resulted in an average increase of the midterm mark due to the group portion of 3.3% and an average increase in the final exam score due to the group portion of 1.6%. The resulting impact on the average course grade of the group part of the exams was

0.5% from the midterm and 0.7% for the final exam, where the standard deviation of course grade distribution was 9.7%.

On the basis of the collective experience at UBC across the science disciplines of physics, chemistry,

FIGURE 1

Examples of questions taken from a two-stage exam for physics.

Most questions will be the same for the individual and the group part. If questions are modified, it is usually to reduce the number of detailed calculations, which do not promote discussions, and replace with prompts to "explain your reasoning." Additionally, one or two more challenging questions may be added.

INDIVIDUAL PART

A train is approaching the train station at velocity $v_0 = 15$ m/s relative to the ground in still air. The train operator sounds the train whistle, which emits a note with frequency $f_0 = 2500$ Hz.

The sound of the whistle is heard by different observers:

The train operator hears a frequency f_A ;

a person standing on the station platform watching the train approach hears a frequency f_B ;

the operator of a second train approaching the station from the other direction with velocity $v_2 = 10$ m/s hears a frequency f_C .

What are the frequencies f_A , f_B , and f_C ?

GROUP PART

(Changed to ranking) A train is approaching the train station at velocity v_0 relative to the ground in still air. The train operator sounds the train whistle, which emits a note with frequency f_0 .

The sound of the whistle is heard by different observers:

The train operator hears a frequency f_A ;

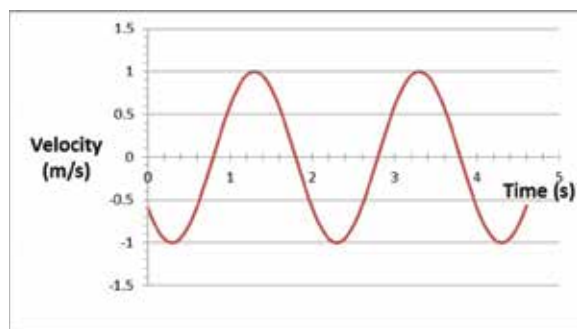
a person standing on the station platform watching the train approach hears a frequency f_B ;

the operator of a second train approaching the station from the other direction with velocity v_2 hears a frequency f_C ;

a passenger traveling on a slower train that has just been overtaken by the first train (and sees the first train move farther away) hears frequency f_D .

Rank the frequencies heard by the observers (f_A, f_B, f_C, f_D) in order from the highest to the lowest frequency.

The graph shows the velocity vs. time graph of a harmonic oscillator.



Determine

- a) the angular frequency
- b) the maximum displacement
- c) the phase constant and the equation describing the position as a function of time.

(Replace part c)

- a) same
- b) same
- c) Sketch the potential energy curve as a function of time. Assume that we have a **horizontal** harmonic oscillator.

biology, math, statistics, Earth and ocean sciences, computer science, and land and food systems, we would further recommend the following:

- Students are told on the first day of classes that examinations will be conducted in this format and, more important, why this is done in this way.
- A policy is implemented that the group score cannot be lower than the individual mark. This will address concerns about fairness. In practice, it affects only a few high-performing students as groups perform equal or better than individual students in almost all cases.
- Clear instructions are given during the individual-to-group transition. For example, students should remain seated while their individual exam copies are collected. Remind and check that all names and student numbers are listed on the group exam.
- Students are discouraged from working on their own during the group portion and all members are encouraged to be involved in discussing every problem. Teaching assistants and instructors can help with forming groups and encouraging collaborative work, but this is seldom needed.

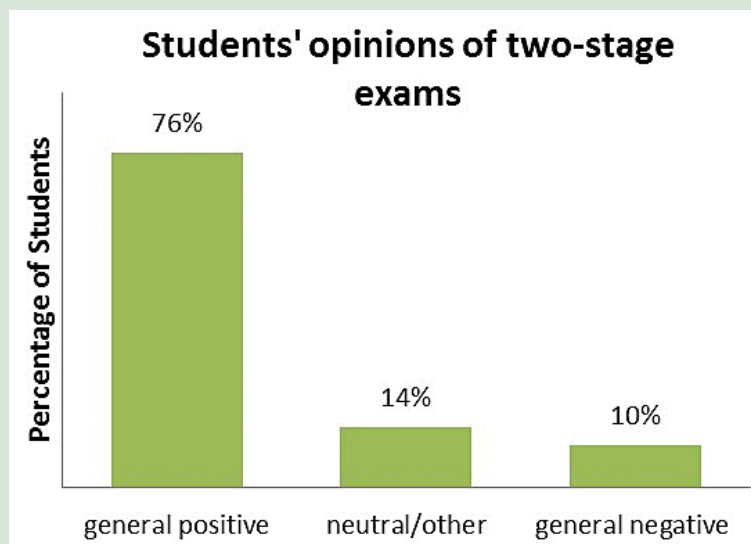
Overall, it does not take much effort to run a two-stage exam. From our experience, creating the group portion of the exam is easy because it is largely identical to the individual exam, and the additional marking time of the group copies is minor because most solutions are correct. To our knowledge, no instructor at our institution who has tried two-stage exams has abandoned this approach.

Benefits of two-stage exams

Here we offer some thoughts on why collaborative exams can increase learning and retention and

FIGURE 2

Percentage of students with generally positive, negative, or neutral opinions about two-stage exams ($N = 123$ students). General positive means students found the exam format to be good or helpful for learning. Neutral/other means that students did not express a clear positive or negative opinion, or commented on other things. General negative means that student had overall negative comments about the exam format.



add a few observations from several science classes at UBC.

During the high-stakes environment of an examination, students are heavily invested in figuring out the correct answers. After the individual portion, all students are well prepared to discuss their approach in a group. In these discussions, students get immediate feedback on their solutions from their peers, which might help them clarify their thinking (Cortright et al., 2003, Gilley & Clarkston, 2014, Rao, Collins, & DiCarlo, 2002). Weaker students could benefit from the explanation that is targeted to their difficulties, higher achieving students might benefit from explaining concepts to others, and everyone may well benefit from critically evaluating others' ideas.

One may argue that these same benefits are also present in "normal" in-class collaborative learning activities; so why do this on an

exam? Even a casual observation of the two situations reveals the difference: We routinely see nearly 100% engagement during the group part of the exams, presumably because of the high-stakes situation of an examination. As confirmed through both observations and student self-reports, most groups discuss the questions until all members agree on an answer; even during open-book, two-stage exams, it was very unusual to see students looking through the book to find the answers instead of discussing them and figuring them out themselves. Those students who are usually too shy to speak up during in-class activities will defend their answers vigorously during the second stage of the exam. By comparison, the discussions during normal in-class activities, such as clicker question discussions, do not have nearly the same intensity. This is probably because the stakes are

lower, and it is not necessary for students to reach an agreement because they usually submit their own (individual) answers. The students also know that they will receive expert feedback from the instructor following the discussion, so they don't have to evaluate as carefully what their colleagues are saying. Finally, students are better prepared to carry out peer discussions in a two-stage exam than they are during lecture because (a) they have studied for the exam, and (b) each student is forced to think deeply about the questions during the individual portion of the exam before the discussion starts in the group portion.

In our introductory physics class, we noted an additional beneficial effect of a two-stage midterm exam: It increased the engagement during in-class collaborative activities following the examination. Although students regularly participated in peer discussions of clicker questions and worksheet problems before the midterm and the instructors explained the benefits of collaborative learning, it appears that the two-stage exam convinced the students (more) of the value of peer discussions. It is also possible that, after the midterm exam, students think of the in-class activities as more directly related to exams.

Impact on student opinions

For illustrative purposes, we examine in detail how two-stage exams impacted student opinions in one course; however, these results are similar to what has been seen in other science courses.

We gave both the midterm and final exams in a two-stage format in the aforementioned calculus-based introductory physics course. The students filled out a 20-question online survey after the final exam; four questions probed their views on the exams. Of the 179 students, 123 completed the survey. Eighty-seven percent supported the use of the two-stage exam format for midterms,

TABLE 1

Coding scheme and results as applied to students' written comments regarding their experience with two-stage exams in Physics 101.

Overall code	Detailed code	Description of code	No. of times mentioned (N = 123 students)
General positive (Total: 236)	G-E	Good, enjoy, benefit, great, liked, useful, OK, interesting	56
	H	Helpful	30
	C	Increased confidence	9
	LE	Good learning experience, good way to review exam	21
	LE-D	Learning from: discussions with others, hearing other approaches, comparing with others, explaining yourself, collaborating	48
	IF	Immediate feedback: good to know if right or wrong	34
	IF-LM	Immediate feedback: learning from mistakes	16
	GD-pos	Positive mention of group working together, group members, meeting friends, group preparation, cooperation, and references to grade boost	22
Neutral/ other	Misc	Random comments not fitting into the above categories as well as suggestions	15
General negative	NEG-gen	Negative mention of group not working so well together, not everyone pulling their own weight, hard to explain to others, and concerns about unfair grade boost to weaker student, not fair for the individual	15
	NEG-em	Dislike, frustrating, not helpful, feeling sad or depressed, less confident	15

whereas 74% supported the use for both midterm and final exams. A possible reason for the difference could be that students view the midterm as being part of learning and perhaps feedback on their studying, whereas they see the final exam as a kind of “certification,” similar to many instructors. Many students see this course as their final exposure to physics, so although students may see the second-stage feedback on what they did wrong for the midterms as productive, they may not appreciate it as much for the final, where there is no hope of using the feedback for future improvement. To explore this further, one would need to conduct interviews with students.

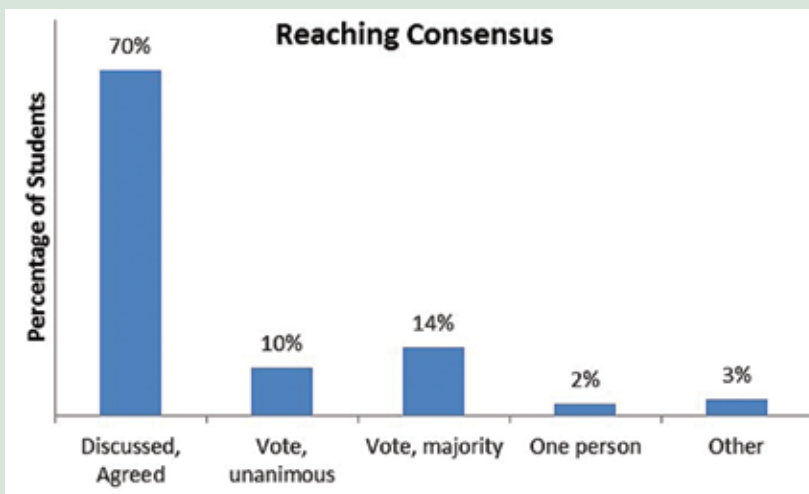
The survey included a question in which students were asked to describe their experience with the group exam in one or two sentences. All students who completed the survey answered this question. As shown in Figure 2, most students had a generally positive opinion.

The detailed analysis and coding scheme we developed for classifying the comments is shown in Table 1. Many students’ responses fell into multiple categories; from the 123 students, we coded 283 comments. The comments were coded independently by each researcher and then compared. The interrater reliability for the comments was 95%, with differences in the coding being discussed until an agreement was reached. A few examples of student comments and coding are as follows:

- Student A: “It was a good experience since going over the exam with my peers reassured me about my answers. As well, I was able to learn from my mistakes through the group exam.” (G-E; C; IF-LM)
- Student B: “It was surprisingly very helpful. I would say I contributed as much as I could. When I got a different answer I

FIGURE 3

Student survey results on group decision making (N = 123). Students were asked: “During the group exam, my group usually ____.” Full answer choices (left to right): “discussed EACH question until ALL members agreed on an answer and explanation,” “took a VOTE and if unanimous moved on, otherwise discussed the question until all members agreed on an answer,” “took a VOTE and used the MAJORITY to determine the answer,” “USED the answers from the ONE PERSON in the group who knew the most physics.”



always commented why I chose the answer that I did and our group would discuss it. I think I was also very lucky to meet kind people during lecture.” (H; MISC; LE-D; GD-pos)

- Student C: “It was sort of depressing to know what you got wrong right after writing the exam. I think it ends up being worth it, though, because you learn from your mistakes and the way classmates explain things could be easier to understand at times than the way it’s explained in the textbook.” (NEG-em G-E; IF-LM; LE-D)
- Student D: “The group exam was interesting and a good opportunity to go over the answers and talk about the questions. Just did not like when some members did not do anything.” (G-E, LE, NEG-gen)

These comments give us insight into why students generally value two-stage exams: they felt it was a good learning experience, and a good way of reviewing the exam, they learned from discussing with other students and hearing other students’ approaches, they enjoyed working together, and they valued the immediate feedback in the group part.

Students also expressed concerns about the exam format in 30 negative comments, half relating to group work and the other half to the emotional impact of getting immediate feedback. However, of the 15 students who criticized the group work, only six students rated their overall experience on the two-stage exam as negative. Three of the six students mentioned concerns about “weak students unfairly gaining marks.” Nine students were critical of the group work but

still had an overall positive experience (as Student C). Fifteen students commented that it was “sad” or “depressing” to learn about their mistakes, but nine students still had an overall positive view.

In the survey students were specifically asked about how their group reached consensus. The results are presented in Figure 3. Clearly, most students worked on the group exam in the intended collaborative way. Only three students, two of whom commented about bad dynamics in their group, claimed to have “used the answers from the one person in the group who knew the most physics.” These responses support our observations of classwide participation in the second stage of the exam and the intensity of the physics discussions that ensue.

Summary

Two-stage exams are valuable instructional tools that offer a combination of formative learning and assessment. They can easily be implemented in many courses and are popular with students and faculty members who use them. Surveys show that this exam format is popular with students for the right reasons—students recognize the value of immediate feedback that takes place and the learning that results. The exam format is similar to the collaborative in-class activities and therefore strengthens the link between exams and the peer instruction activities in class. We have noted an increase in engagement during in-class peer activities after a group midterm exam. Further studies are necessary to establish that this is mainly a result of the two-stage exam. It would also be interesting to find out if students acquire better group skills through participation in the group part of the exam. The experience in our science faculty has shown that the two-stage approach contributes to

the overall coherence of any course that is using techniques of collaborative learning and formative assessment, as well as allowing students to learn while completing the exam. We therefore highly recommend this exam format to any instructor looking to add a formative element to their summative assessments. ■

Acknowledgments

The authors gratefully acknowledge the Carl Wieman Science Education Initiative (CWSEI) for funding and support. We thank Carl Wieman for assistance with the preparation of the manuscript. We also thank all the CWSEI Science Teaching and Learning Fellows for providing information on their experiences with two-stage exams, in particular Brett Gilley and Bridgette Clarkston.

References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5, 7–74.
- Cohen, D., & Henle, J. (1995, July). The pyramid exam. *UME Trends*, 10, 2, 15.
- Cortright, R. N., Collins, H. L., Rodenbaugh D. W., & DiCarlo, S. T. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education*, 27, 102–108.
- Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching*, 43(3), 83–91.
- Heller, P., & Hollabaugh, M. (1992). Teaching problem-solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, 60, 637–644.
- Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part

1: Group versus individual problem solving. *American Journal of Physics*, 60, 627–636.

- Johnson, D. W., Johnson, R. T., & Smith, K. A. (2011). Lecturing with informal cooperative learning groups. In J. Cooper & P. Robinson (Eds.), *Small group learning in higher education: Research and practice* (p. 46). Oklahoma City, OK: New Forums Press.
- Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE—Life Sciences Education*, 11, 392–401.
- Mazur, E. (1997). *Peer instruction: A user’s manual*. Upper Saddle River, NJ: Prentice Hall.
- National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press. Available at http://www.nap.edu/catalog.php?record_id=13362
- Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education*, 26, 37–41.
- Stearns, S. A. (1996). Collaborative exams as learning tools. *College Teaching*, 44(3), 111–112.
- Zipp, J. F. (2007). The impact of two-stage cooperative tests. *Teaching Sociology*, 35, 62–76.

Georg W. Rieger (rieger@phas.ubc.ca) is an instructor 1 in the Department of Physics and Astronomy and a member of the Carl Wieman Science Education Initiative (CWSEI), University of British Columbia (UBC), British Columbia, Vancouver, Canada. **Cynthia E. Heiner** was a teaching and learning fellow in the CWSEI at UBC at the time the article was written and is now at the Free University Berlin in Germany.
