Learning science and collaborative skills during summative testing

HELEN BREMERT, AMY STOFF, AND SARAH B. BOESDORFER

Scientists and other professionals across the globe require employees to collaborate, think critically, and solve problems effectively. To this end, the Next Generation Science Standards (NGSS) have a vision of ensuring that through inquiry, collaboration, and evidence-based instruction, students will have the necessary skills to be practical and rational thinkers upon graduation (NGSS Lead States 2013). Collaborative work among students is part of the vision of NGSS. For example, the Grade 9-12 Science and Engineering Practice (SEP) of Planning and Carrying Out Investigations states: “Plan an investigation or test a design individually and collaboratively to produce data to serve as the basis for evidence as part of building and revising models, supporting explanations for phenomena, or testing solutions to problems” (italics added, Appendix F, NGSS Lead States 2013).
Collaborative learning is a form of instruction that promotes an active classroom learning environment where the students form pairs or groups to accomplish tasks (Meseke, Nafziger, and Meseke 2010). The advantages of this method of instruction include increased conceptual understanding, retention, and problem solving and critical thinking skills (Gilley and Clarkston 2014). Furthermore, collaboration promotes heightened intrinsic motivation, interpersonal skills, and the ability for students to engage in evidence-based argumentation (Meseke et al. 2010) that are all necessary skills for the current workforce.

While many students collaborate in tasks and formative assessments in their courses, they generally take summative assessments individually (Siegal et al. 2015). Breedlove, Burkett, and Windfield (2004) point out that individual tests have have been shown at times to lower intrinsic motivation, use only information-recall questions, and increase students’ test anxiety.

Collaborative summative assessments can bring the benefits of collaborative learning to assessment, transforming it into a learning experience. Studies done at the undergraduate level have shown that collaborative assessments improve students’ depth of understanding, critical thinking skills, and exam performance (Gilley and Clarkston 2014), likely as a result of students engaging with their peers to discuss questions and answers, thereby filling in knowledge gaps (Vogler and Robinson 2016).
In this article, we describe how two teachers have used and continue to use collaborative assessments in their classes. We describe the impact on student learning and testing, and provide some tips and resources to implement collaborative assessments in any science classroom.

**Using collaborative assessments in the classroom**

**Alternating assessments**

Ms. Bremert used an alternative testing method with her AP Environmental Science (APES) course. For each exam, students did not know until they came to class if it was an individual exam or a collaborative group exam to ensure that all students studied. The format of the tests was the same for both types of tests; multiple choice and short-answer questions sourced from College Board APES exams, with a time limit of 50–55 minutes. The groups were changed for each test and the students were not told who was in their group until the testing day.

For the group exams, every student had to answer the multiple-choice questions; however, for the short response questions, only one member had to provide complete sentences; other group members could use bullet points. Enabling the students to hand in their individual multiple-choice answers gave them the ability to reject the group's answers if they felt the group was wrong. This rarely occurred. However, when it did, the students indicated that this gave them more control and reduced stress within the group and for themselves. All members of the group received the same short response grade. Student scores were determined by combining their multiple-choice scores with the responses to the short answer questions.

**Two-stage assessments**

Ms. Stoff used a two-stage assessment method in her AP Chemistry course. On exam day students first completed a multiple-choice portion of the chapter test individually. Once each student had completed the multiple-choice portion, the students were randomly placed in small collaborative groups of three to answer the collaborative portion of the exam.

Collaborative groups were changed for each exam, and students did not know their group until the exam period.

The same multiple-choice questions were used on both the individual and collaborative portions of the exam. The students worked together, discussed the material, and agreed on one answer to record for the group for each collaborative assessment. The students were given approximately 90 seconds for each multiple-choice question on both the individual and collaborative portions of the exam. Once the group finished the collaborative portion of the exam, each student completed a short answer portion of the exam individually. Students’ exam scores were calculated as 85% from their individual portions and 15% from the collaborative portion.

**Students’ responses to collaborative testing**

In both teachers’ courses, working in collaborative groups improved assessment scores compared to completing the exam individually. (Note: Because these were different courses at different schools with different assessment methods, the same data was not collected in each classroom. The results presented here are a summary of what was collected.) For example, the alternating assessment method from Figure 1 shows the averages for the exams by type of exam (group or individual) and in the order they took them from Ms. Bremert’s AP Environmental Science Class.

In Ms. Stoff’s AP Chemistry class, two-stage assessments showed benefits for both low-achieving students and high-achieving students, with scores on the short answer section increasing after the collaborative portion. Three class periods after completing the collaborative portion of the chapter test, students in Ms. Stoff’s class took the same number and similar type of multiple-choice questions individually and in scrambled order to determine the impact of the collaborative portion of the exam.

**FIGURE 1**

**Individual and group test results.**

Results are for the three individual (Ind) and three group (Gp) tests, from AP Environmental Science course (n = 38). The results are displayed in order of when the students took the test.
Figure 2 provides the results for low-achieving students (≤ 84% on individual multiple-choice part of the assessment) and high-achieving students (≥ 85%). It appears that low achievers benefited more from the collaborative portion of the exam than high achievers; however, these results were not found to be statistically significant for high achievers, possibly due to small sample size.

Although the collaborative assessments did not lead to a large increase in scores on exams for most students, both teachers observed increased student engagement, improvement in critical thinking related to the assessment questions, and active discussion and debate among peers in the collaborative groups during the exams; again, skills promoted by NGSS and desirable in students.

Students in both classes completed surveys about their thoughts and opinions relating to collaborative assessment and anxiety related to test-taking. The majority of students in both classes indicated that collaborative testing lowered their anxiety levels, even though they prepared the same way as they would for individual tests. Figure 3 provides some sample student responses related to test-taking anxiety. Additional, students indicated that they believed working collaboratively to answer assessment questions enhanced their understanding, which they then applied to the AP exam. Figure 4 provides sample student responses about how collaborative assessment impacted their understanding.

Suggestions and tips

**Group formation**
The instructor should determine the student groups before students enter the classroom, taking group size and skill level of students into consideration. To ensure effective group size, both teachers placed the students in groups of three or four with a mix of genders and ability (heterogenous grouping techniques). Ms. Bremert used previous test scores and determined the groups based on high, middle, and low performing students. Ms. Stoff initially chose students for groups randomly, taking gender and numbers into consideration, but quickly learned that skill level needed to be taken into account when forming the groups.

Furthermore, addressing differences in student personalities should be considered when establishing the groups. Both teachers learned that although it is important for the students to learn to work with different people, there were some pre-existing personality conflicts that affected the results. For example, more forceful personalities might dominate the group, and less persuasive students may not be included in the discussions, or they feel pressure to accept an answer.

**FIGURE 2**

Low and high achievers results.

This figure summarizes the average scores and standard deviations for the low and high achieving students on individual section (dark) and the learning (lighter shade) section (3 days later) for each of the exams in Ms. Stoff’s Chemistry course (n = 17).

**FIGURE 3**

Sample of students’ survey responses related to test-taking anxiety.

- Doing an individual test definitely increased my anxiety because I had no one backing up my answer or contradicting me on why the answer I picked was wrong. I felt that taking a test individually stresses me out more because I could be the only one getting the bad grade.
- I tend to second guess myself, so having group members to take the test with and talk to them about the choices helps me relax more.
- I felt more confident when my whole group got to discuss our answer.
- Without being able to explain my thoughts out loud to a group of my peers, I felt less confident in my answers.
- They [Collaborative Exams] are less stressful than normal tests and they increase my engagement with the material.
- I would definitely rather work in groups because it’s less stressful. You have more people to bounce your ideas off of if you are struggling to pick an answer.
Finally, new collaborative groups should be created for each exam; again, groups should not be announced before exam day. Both teachers found this ensured that all the students studied for the test in the same manner as they would for any individual assessment. By not knowing who is in their group ahead of time, they do not know if there will be someone on whom they could depend for correct answers. For both teachers, anecdotally, the students indicated they did not feel that anyone “coasted” in their groups; they all participated equally, and the teachers have not observed significantly unequal distribution of thought in the groups.

**Addressing concerns with collaborative exams**

An issue with collaborative testing is *coasting*—that students will rely on the smartest in the group. In Ms. Bremert’s class, coasting was discussed with students after collaborative testing. Students indicated that they felt few students coasted in their groups; this was also observed by the teacher moving around the classroom observing students engaged in the exam. For Ms. Stoff, comparing grades from the first semester (collaborative assessments not implemented) to the second semester (collaborative assessments implemented), the majority of students maintained the same letter grade for the course while some of the students’ averages decreased by less than 5%. Based upon the students’ semester scores, it is believed that the students used the collaboration time to better understand the material and did not coast as they still needed to demonstrate their understanding on the test and on the semester final exam.

Both teachers found success by not announcing the groups or the type of test until exam day. By not announcing the type of test, individual or collaborative, Ms. Bremert found that all students were prepared and contributed to their group’s responses. This could be the plan throughout the academic year, or a teacher could announce that it may happen on occasion to keep students honest, and then use it less frequently.

Another possibility would be to give a learning assessment a few days after the collaborative portion of the exam to ensure students are participating and learning in the group. When doing the two-stage assessment, Ms. Stoff used the learning assessment to determine student retention of materials, but teachers could also use it to keep individuals engaged in the group exam process.

An additional concern for teachers may be a lack of support for this technique from administrators and parents. For both courses, administrators and parents were informed of the use of the collaborative assessments in the courses and the rationale for their use prior to implementation. In both instances, each group was supportive. Parents were given the opportunity to ask questions to help allay their concerns, and most indicated support for the process. However, there were concerns from other teachers; part of the purpose of this article is to demonstrate the merits of this technique and to help reduce these concerns.

One other concern may be whether this method could be used across grade levels, as both courses described here were advanced courses for upper-level students. Ms. Bremert also used collaborative testing in a sophomore mixed-ability biology class, but not to the extent used in the AP course. The students did not collaborate as effectively as the senior students, and they seemed to have difficulty listening to each other’s viewpoints. In part, this was likely due to lack of exposure to collaborative group work. Ms. Bremert will use the method again with other levels but establish better culture and practice for collaborative work in day to day activities as well as exams.

**Other methods for collaborative assessment**

Ms. Stoff added individual short-answer questions after the group portion of the two-stage collaborative exam. The two-
stage method could also be used with the added step of the class completing the assessment together (Giuliodori et al. 2008), which would allow the students to process the questions on their own first, then receive feedback from their peers. Finally, parallel assessments could be used as another variety of collaborative assessments. Each student in a group would receive a version of the assessment with similar (but not the same) questions, so students would have to apply their own knowledge to their assessment questions, while also collaborating with their peers (Reiser 2017).

Students’ assessment grades can be determined in numerous ways—as averages of the two parts, weighted to a learning assessment result, or other ways the teacher sees fit. However, a teacher should be careful not to undervalue the collaborative portion of the exam; undervaluing it will lead to students not taking it seriously, not using it as a learning experience, or making it a competitive environment rather than the collaborative one it is meant to be.

Final thoughts
Collaborative summative assessments can be used in a high school science class to enhance student engagement, critical thinking skills, and understanding of the concepts taught. Teachers who have used collaborative assessment report that it encourages students to work together, discuss content concepts, and enhance their knowledge. Furthermore, collaborative assessments lessened students’ anxiety over testing, enabling them to focus on the questions and demonstrate their knowledge. Collaborating also enables students to develop positive interactions with each other. Collaborative assessments are a great tool for any science teacher to use in their classroom to promote student learning in all activities they require of students.

REFERENCES