In-Class Incentives That Encourage Students to Take Concept Assessments Seriously

By Michelle Smith, Katie Thomas, and Maitreya Dunham

Pre/post concept assessment testing is becoming increasingly common in college courses. Instructors use different approaches to give assessments, but few studies have examined how administration differences affect results. Here, we ask if administering a posttest on the final exam differs from administering it on the last day of class with extra final exam points for scoring 100% on the posttest as an incentive. To answer this question, we compared the performance of two courses on the Genetics Concept Assessment. We find comparable results between both courses, suggesting that instructors can choose the method of administration that resonates best with their teaching style.

In recent years, there has been a growing call to assess undergraduate student learning in an effort to transform college science classrooms (American Association for the Advancement of Science, 2009). Consequently, more instructors are using pre/post concept assessment as one instrument to assess students' learning gains. Concept assessments are typically composed of multiple-choice questions that target common student misunderstandings and include wrong answer choices that are based on extensive research (Adams & Wieman, 2011). Results from assessments help instructors identify common areas of student confusion, providing a measure for making decisions about how to revise course content.

There are a variety of concept assessments available that focus on various science subdisciplines (D'Avanzo, 2008; Knight, 2010; Libarkin, 2008; Redish, 2003). These assessment tools are developed through a rigorous process that includes validation through interviews with students and feedback from experts. The tools are also analyzed for reliability using data from hundreds of students from multiple institutions. Typically assessments are given on the first day of class as a pretest and then the identical questions are given again at the end of class as a posttest. The outcomes of the pretest are not discussed with students.

Two main approaches for administering posttest questions in class have been identified (Adams & Wieman, 2011; Redish, 2003). The first is to give the pretest in class on the first day and put the posttest questions on the final exam. When this approach is used, the number of participating students is maximized, the assessment is occurring after students have reviewed the course material, and students are likely to take the questions seriously. However, some instructors object to having assessment questions given in place of other possible final exam questions, because, for example, they are worried about asking the same questions on the final exam from one year to the next or they prefer to emphasize material taught in the last unit of the course. Thus, some instructors prefer a second administration approach, giving the pretest in class on the first day and giving the posttest during the last week of class for participation points. Instructors then typically review low-scoring posttest questions in preparation for the final exam. However, because the in-class administration method does not count for a grade, students may not have enough incentive to take the assessment seriously. Furthermore, one study found that students who take the Conceptual Survey of Electricity and Magnetism (Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001) posttest in class for participation points, without any incentive to do well, perform significantly lower than students who take the same assessment on the final exam (Ding, Reay, Lee, & Bao, 2008).

Here, we propose a method for

administering the posttest that allows instructors to administer the assessment in class, but also includes performance incentives. In this study, students in two different quarters of the same genetics course took the Genetics Concept Assessment (GCA; Smith, Wood, & Knight, 2008) as a pretest on paper during the first day of class. During one quarter, students answered the posttest questions on the final exam and during another quarter, students answered posttest questions for the first 30 minutes on the last day of class. Students who took the posttest on the last day of class were told that if they scored 100% on the posttest, points would be added to their final exam. In addition, if the posttest results revealed that a large percentage of students had a particular misunderstanding, extra information was added to the study guide released the weekend before the final exam. In this study, we sought to find out whether these in-class incentives provided enough motivation to encourage students to perform at a level equivalent to when questions are on the final exam. In this field test, we found that students in both scenarios performed equivalently. The results of this study can be used to help instructors make informed decisions about reliable ways to administer concept assessment posttests.

TABLE 1

Demographic information on students in the spring 2011 (in-class posttest) and summer 2011 (final exam posttest) courses.

	Spring 2011 (in-class posttest)	Summer 2011 (final exam posttest)
Class standing	2% junior, 97% senior	18% junior, 74% senior
Ethnicity	58% non-Caucasian	63% non-Caucasian
Sex	57% female	61% female
Major	70% biology 18% biochemistry	62% biology 23% biochemistry

Methods

Course background

This study was conducted in two undergraduate genetics courses for majors at the University of Washington taught in the Genome Sciences Department (spring quarter 2011 and summer quarter 2011). The spring quarter class was team taught by author Maitreya Dunham, course coordinator Michelle Smith, and one other instructor. The summer course was taught by Michelle Smith. Student demographic information is shown in Table 1.

The spring and summer quarter genetics courses were very similar. The classes met twice a week for a 90-minute lecture and also included a 50-minute recitation section. During the recitation section, students worked in small groups to solve genetics problems. The spring- and summer-quarter students worked on identical recitation section questions. Both quarters also included weekly homework and clicker questions in each lecture and shared a set of learning goals that cover the first 75% of the course. There was also a commitment among instructors to avoid memorization questions. Furthermore, the spring exam questions were often the summer homework questions, and the spring homework questions were typically the summer exam questions with changes to surface features.

Assessment administration

The GCA (Smith et al., 2008), which consists of 25 multiple-choice questions, was administered on the first day of class as a pretest in both courses. Students completed the assessment during the last 30 minutes of class using bubble sheets. Students were asked to put forth their best effort and were informed that their scores would have no bearing on their grade, but that the instructors would use the results of the pretest to help determine which topics in genetics were particularly challenging for this group of students. The results of the pretest were never discussed with the students.

At the end of the spring 2011 course, the identical 25 questions were given during the first 30 minutes of the last day of lecture. The students were told that if they got all the questions correct, 5 points would be added to their final exam (the final exam was worth 100 points out of a total of 320 points for the quarter). Two students got 100%. The instructors also talked to students in the class about taking the assessment seriously even if they were not confident that they would score 100%. They encouraged all students to use the opportunity to practice answering timed multiple-choice questions before the final and told students that information about concepts from questions with the poorest performance would be added to a final exam study guide. For example, the following statement was added to the study guide: "Nondisjunction can occur during meiosis I or II, make sure you practice drawing the differences." After students took the posttest, the instructor lectured on a concept not covered on the GCA (association mapping). The final exam primarily focused on concepts covered in the last 3 weeks of the course: LOD score analysis, association mapping, and cancer genetics. In the subsequent sections of this paper, this class will be referred to as the "in-class posttest course."

In the summer 2011 course, students were given the GCA questions on the final exam; they were worth half of the total points on the exam.

TABLE 2

Mean pretest, posttest, and normalized change scores for students in the in-class posttest and final exam posttest courses.

Course	Pretest (%)[a]	Posttest (%)[a]	<c>[a,b]
In-class posttest	46.8 (1.4)	71.9 (1.5)	49.6 (2.0)
Final exam posttest	46.2 (1.5)	71.4 (1.6)	48.8 (2.4)

[a]The SEM is shown in parentheses.

[b]See Methods section for the normalized change formula.

Students did not know the assessment questions would be on the final exam, so they could not study for them specifically. The remaining questions on the test were short-answer questions that focused on the last unit of the course: LOD score analysis, cancer genetics, genetics counseling case studies, and population genetics. For the remainder of this paper, this class will be referred to as the "final exam posttest course." Only students who completed both the pretest and the posttest are included in this study (n = 136 in-class posttest course, n = 107 final exam posttest course).

Data analysis

The change in learning between pretest and posttest was computed for each individual student using a modified version of the Hake normalized gain formula (Hake, 1998) known as normalized change <c> (Marx & Cummings, 2007). Normalized change values provide a measure of how much a student's score increases compared with that individual's maximum possible increase. When calculating the normalized change between pretest and posttest, the following formula was used when an individual's posttest score was equal to or higher than the pretest score (135 out of 136 cases in-class posttest course, and 106 out of 107 cases final exam posttest course): <c> = 100[(posttest – pretest) / (100 – pretest)]. Alternatively, if an individual's pretest score was higher than the posttest score, <c> = 100[(posttest – pretest) / pretest)], was used. There were no cases in which both an individual's posttest and pretest scores equaled either 100 or 0. The standard error measurements on reported <c> values are used to provide a depiction of the spread of values (Marx & Cummings, 2007).

To examine the correlation between posttest performance and course performance, total exam points were calculated for each student. For the in-class posttest course, student final grades were determined by the three out of four best exam grades, so this value was used in the correlation analysis. For the final exam posttest course, students took three exams that all counted toward their final grade. However, for each student, we subtracted the posttest score from the exam points so we were not correlating posttest scores with values that also included posttest points. All statistical analyses were performed with SPSS (SPSS, Chicago, IL) or Excel (Microsoft, Redmond, WA).

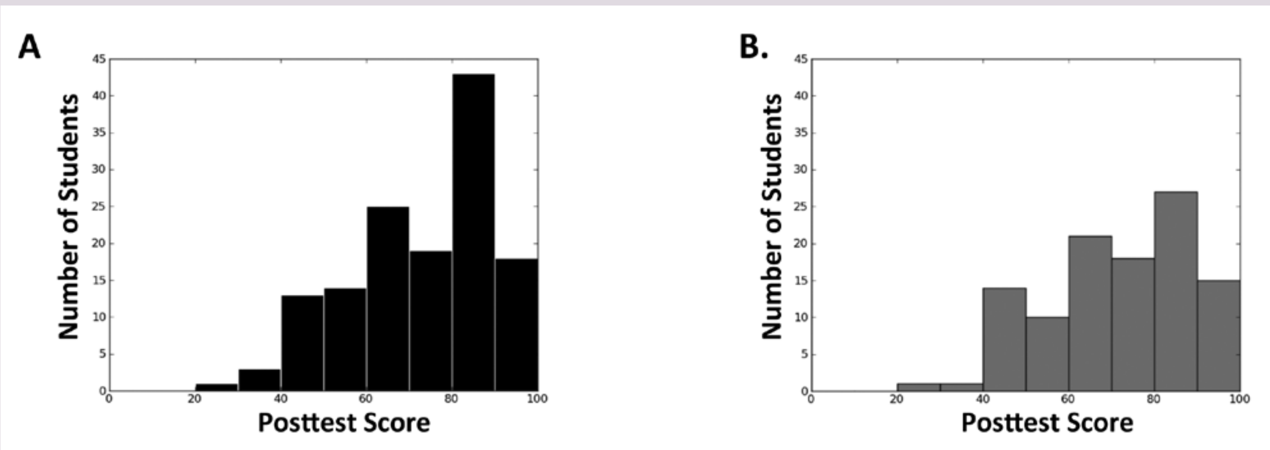
Approval to evaluate student pretest and posttest responses (exempt status, Protocol No. 39014) was granted by the Institutional Review Board, University of Washington.

Results

There is no significant difference between mean pretest scores in the in-class posttest and the final exam posttest courses (t-test, $p > .05$), indicating that the two courses are comprised of students who have similar

FIGURE 1

Distribution of posttest scores in the in-class posttest course (A) and final exam posttest course (B).



levels of genetics knowledge at the beginning of the course (Table 2).

The distribution of posttest scores for both courses is shown in Figure 1. There is no significant difference between mean posttest scores (t-test, $p > .05$ in all cases), and the mean normalized change (<c>) scores from pre to post are nearly identical (Table 2).

We also examined the correlation between student posttest score and their overall exam points. Figure 2 shows a strong positive and significant correlation between exam points and posttest score for both courses (in-class posttest, $r = 0.77$, $p < .05$; final exam posttest, $r = 0.76$, $p < .05$).

Finally, we examined the number of omitted posttest answers in both courses and found three omitted answers for the final exam posttest course and 21 omitted answers for the in-class posttest course. However, 18 of the 21 omissions came from two students (out of 136) who were likely not taking the in-class assessment seriously.

Discussion

As more instructors use pre/post concept assessments in their courses, it is important to know whether the incentives given to students to take assessments seriously are meaningful enough to gather reliable data. Although placing assessment questions on the final exam provides a powerful incentive for students to do their best, many instructors prefer not to administer assessments this way. Previous work on a different assessment, Conceptual Survey of Electricity and Magnetism (Maloney et al., 2001), revealed that when students take the posttest in class for only participation points, they perform significantly lower than students who take the same assessment on the final exam (Ding et al., 2008). Here, we show that adding a modest amount of performance incentives, offering five points to students who get 100% and a study guide with added extra information about the concepts in lower scoring posttest questions, is enough to encourage students to perform at a level comparable to answering the questions on the final exam (Table 2). In addition, we find a strong positive and significant correlation between exam points and posttest score for both courses (Figure 2), indicating that the in-class posttest is accurately capturing student ability. Finally, an investigation of the number of questions with omitted answers supports our conclusion that the vast majority of students who are taking the posttest in class are taking it seriously.

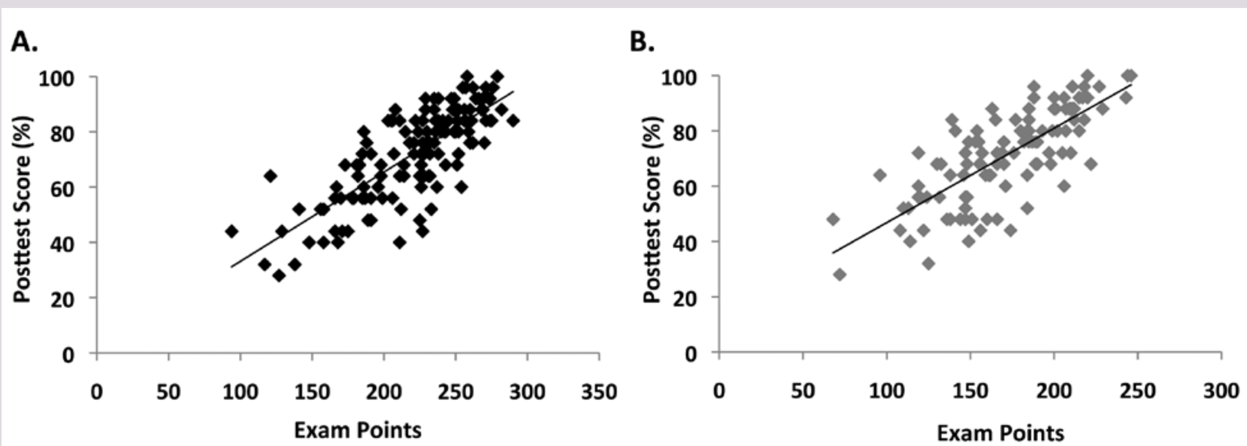
Although there are many ways to administer a posttest, it is important that students take the assessment seriously in order for the instructors to get accurate results. Our experience with this field test provides evidence that meaningful posttest results can be gathered in-class with a modest amount of incentives. This conclusion gives flexibility to instructors who do not want to make the posttest questions part of the final exam. We hypothesize that instructors who give concept assessments in other subjects will see similar results and encourage this line of investigation. Furthermore, taking separate and anonymous measures of student self-motivation while they are taking assessments in different testing conditions could provide additional information on how students view various incentives. ■

Acknowledgments

We thank Josh Akey, Colin Manoil, and Willie Swanson for giving the Genetics Concept Assessment in their genetics courses and Jason Patterson for help with the student demographic information. Special thanks to Alison Crowe and Rebecca Price for their comments on this manuscript. Maitreya Dunham is a Rita Allen Scholar and supported by National Science Foundation Grant 1120425.

FIGURE 2

Correlation between posttest score and exam points in the (A) in-class posttest course (y = 0.32x + 0.84, R^2 = 0.60) and (B) final exam posttest course (y = 0.34x + 12.64, R^2 = 0.57).



References

Adams, W., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education, 33,* 1289–1312.

American Association for the Advancement of Science. (2009). *Vision and change in undergraduate biology education: A view for the 21st century* (Conference homepage). Available from http://www.visionandchange.org

D'Avanzo, C. (2008). Biology concept inventories: Overview, status, and next steps. *Bioscience, 58,* 1079–1085.

Ding, L., Reay, N. W., Lee, A., & Bao, L. (2008). Effects of testing conditions on conceptual survey results. *Physics Education Research, 4,* 010112.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66,* 64–74.

Knight, J. K. (2010). Biology concept assessment tools: Design and use. *Microbiology Australia, 31,* 5–8.

Libarkin, J. C. (2008, October). *Concept inventories in higher education science*. Manuscript prepared for the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2, Washington, DC.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics, 69*(Suppl. 1), S12–S23.

Marx J., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*, 87–91.

Redish, E. R. (2003). *Teaching physics with the Physics Suite*. Hoboken, NJ: Wiley.

Smith M. K., Wood, W. B., & Knight, J. K. (2008). The Genetics Concept Assessment: A new concept inventory for gauging student understanding of genetics. *CBE Life Sciences Education, 7,* 422–430

Michelle Smith (michelle.k.smith@maine.edu) *is an assistant professor in the School of Biology and Ecology and the Maine Center for Research in STEM Education at the University of Maine in Orono.* ***Katie Thomas*** *is a graduate student in the Molecular and Cellular Biology PhD program and* ***Maitreya Dunham*** *is an assistant professor in the Department of Genome Sciences, both at the University of Washington in Seattle.*