AC 2010-2260: USE OF AN AUDIENCE RESPONSE SYSTEM FOR CONTINUOUS SUMMATIVE ASSESSMENT

William Guilford, University of Virginia

Will Guilford is currently an Associate Professor and Undergraduate Program Director in Biomedical Engineering at the University of Virginia. He received his B.S. in Biology and Chemistry from Saint Francis College in Fort Wayne, Indiana, and his Ph.D. in Physiology from the University of Arizona in Tucson. Will's research is the biophysical basis of cell movement and muscle contraction. He is also keenly interested in promoting research-based undergraduate education.

Use of an Audience Response System for Continuous Summative Assessment

Abstract

Audience response systems (ARS) have been used extensively for formative assessment and active learning in lecture-based courses. It is not known, however, if they can be successfully used in large classroom settings as the medium for delivering summative assessments. We used an ARS to deliver daily quizzes in lieu of exams to students in a course on cell and molecular biology for engineers. We found that ARS can be used for frequent assessment with instant feedback to the students and with minimal work by the instructor, with exactly the same learning outcomes as paper-based exams.

Introduction

Audience response systems (ARS, also known as "clickers") have been used extensively for formative assessment – helping students determine for themselves whether or not they understand the material, and breaking up rote lectures with an active learning activity. There is an extensive literature on their use and efficacy in these regards. While it is disputed whether ARS use improves student performance, there is evidence of improved retention as a result of using ARS in the classroom setting. Readers are referred to recent articles by Fies [1], and Crossgrove and Curran [2].

ARS have the added advantage of being able to assess large numbers of students simultaneously and rapidly. Paschal noted that ARS can overcome the problem of delays between issuing of formative homework and receipt of feedback by the student [3]. Reports on their use for summative assessment, however, are scant. It has been *proposed* to use ARS to deliver frequent low-stakes assessments rather than relying on infrequent, high-stakes testing [4]. However, the author is aware of no published reports of doing so at any educational level.

A worthwhile question is whether frequent low-stakes testing is equivalent in educational efficacy to infrequent high-stakes testing. It is accepted that testing students on the same material twice leads to improved retention of material. This is referred to as the "testing effect." More than just a psychological phenomenon, there is evidence that the testing effect can be used to enhance learning and retention of material in a classroom setting [5]. The testing effect is often noted in the context of frequent quizzes followed by an exam, or mid-term exams followed by a comprehensive final exam. However, even test-retest intervals of as little as three days can give a test effect and improve retention [6]. Thus an approach of more frequent summative assessment that overlaps material, such that students are required to review any individual learning component more than one time, might prove effective. This approach, however, could prove labor-intensive when it comes to grading; the number of assessments at minimum doubles. This is especially problematic in a time of increased college enrollment and an environment of large class size.

We therefore combined these educational practices, ARS and retesting of material, to deliver overlapping summative assessments to large numbers of students in an undergraduate course on cellular and molecular biology for engineers.

Implementation

We used a SMART Response audience response system from SMART Technologies (Calgary, AB, Canada) in concert with the manufacturer's presentation software – SMART Notebook (Figure 1A). SMART Notebook supports multi-question quizzes; it does not do so with third-party software like Microsoft Excel®. These are radio-frequency devices rather than infrared, so that they do not require line-of-sight to the receiver. We used two USB-interface receivers on a single laptop computer to deliver the assessments; this was necessary to accommodate the class size (96 students) since each receiver has a limited capacity.

The SMART Response handset (Figure1B) has features that make it especially useful for in-class summative assessment. A multi-line screen shows the available questions, the range of allowable answers (if T/F or multiple-choice), and the answer the student has entered for each. The keypad has cursor controls that allow the student to scroll back and forth through the questions, single-letter entry for multiple choice, and explicit Yes/No, True/False keys. Numeric entry is also allowed, including real numbers and fractions.

Students picked up a remote on the way into class. Each student was assigned a serial number that they used as a personal identifier to log onto the system so that their identity was independent of any specific remote. Remotes were returned to the instructor at the end of class.

Questions were projected from an overhead LCD in the classroom. Quizzes consisted of ten multiple choice or numeric value questions, projected in turn on approximately 1 minute intervals for simple recall questions, 2 minute intervals for comprehensive questions, and more as appropriate for analytical questions. The entire set of questions was repeated for review, and then requests for additional time on specific slides were taken from the class. When students were satisfied with their answers, selecting the "Finished" entry on their remote displayed their score on their individual clicker screen. At the end of the quiz, the instructor saved their scores, along with their individual answers to each question, in a spreadsheet. This was followed by a

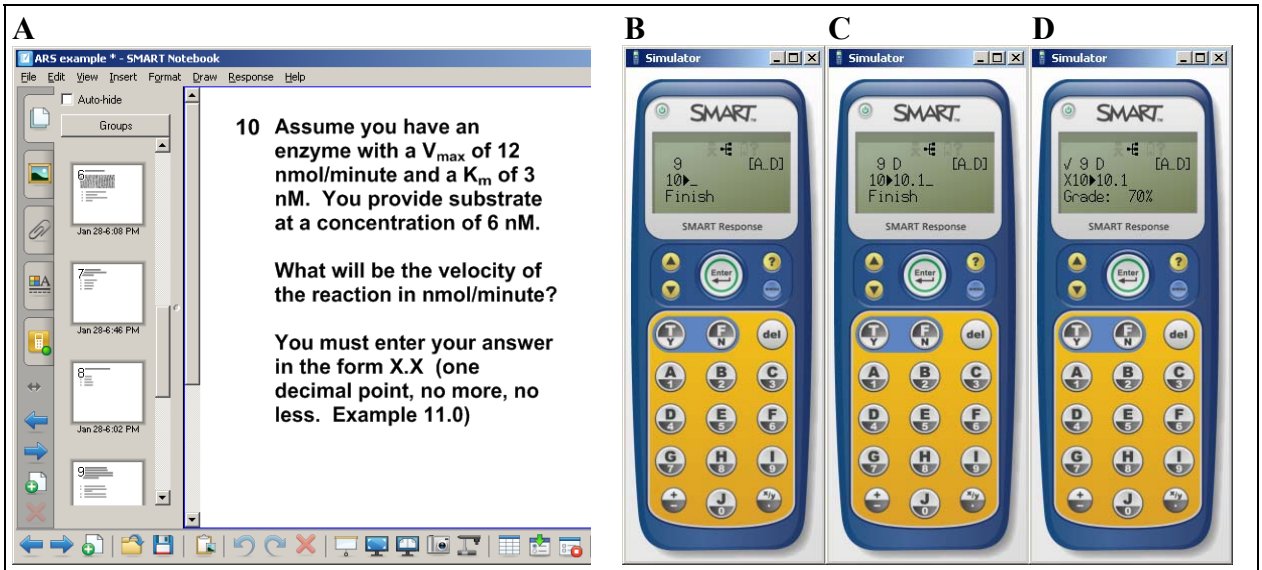

Figure 1: SMART Response software as projected on the classroom screen (A), and hand-held remotes (B-D). B: Remote screen before entering answers. C: During entry. D: After submitting answers, their individual scores are displayed.

very brief review session, where the built-in graphing features of the SMART Notebook software (Figure 2) were used to show the correct answers to each question, and discuss questions that proved problematic.

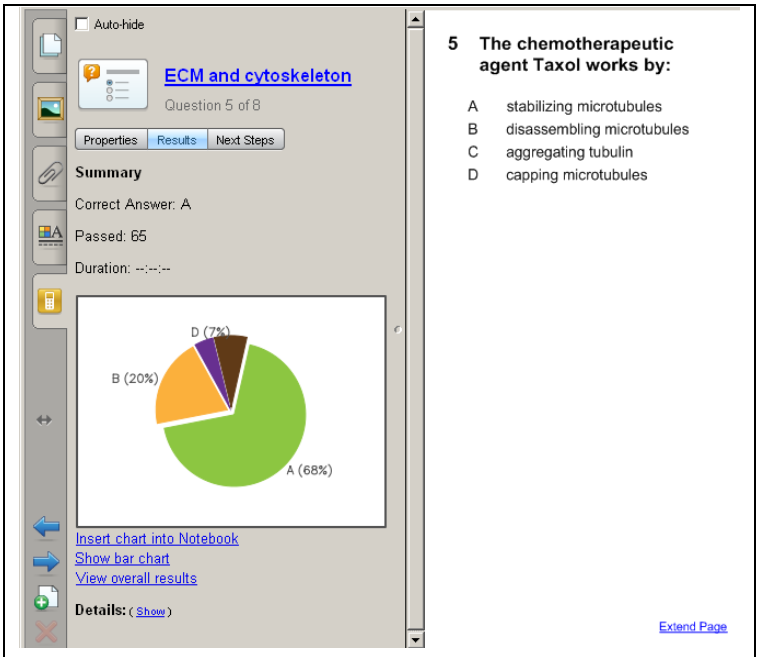

Figure 2: Review of answers at the end of each quiz included a pie-chart of class answers to identify problem areas.

Quizzes were given daily during the first 15 minutes of the 75 minute class session. Each quiz covered the previous *two* lectures material, ideally forcing students to twice study each lecture. This amounted to 24 individual quizzes that accounted for 50% of their class grade. The lowest two quiz scores were dropped.

The only other graded elements of the course were centered on a large writing assignment in which students work in teams to write a review article and engage in peer review. This educational approach has been published elsewhere [7]. The final exam likewise was focused on the written assignment rather than on a comprehensive retesting of the lecture material. The final exam period was also used to gather student feedback on the ARS/quiz approach and to test their retention of the lecture material.

The lecture content and method of delivery was unchanged from the previous academic year, which was used as an experimental control.

Results

Student evaluations showed that they only marginally preferred the daily quiz format over traditional exams (Table 1). However, when offered the choice of paper-based delivery with slow turnaround of grades, or ARS delivery with instant grading, students vastly preferred the ARS delivery. Students also perceived that they learned more and retained more information as a result of this format.

To determine if their perceptions of enhanced learning was correct, a subset of quiz questions were matched to exact equivalents delivered using a traditional "bubble sheet" exam format from the previous year. A non-parametric Wilcoxon test showed no difference between the two methods of delivery and testing ($p=0.68$). Comparing course grades to the previous year using ANOVA, scores overall increased by 2 percentage points, though the difference was not quite significant by conventional assumptions ($p=0.08$).

As a measure of whether retention was improved as a result of the ARS/quiz approach, 10 questions from previous quizzes randomly selected from across the semester were posed during

the final exam period *without* preparation by the students. These were answered with a median accuracy of 53%. Though we lack a control group for this test, this is a similar though slightly lower level of retention than measured by others 4 months after content delivery [2]. Those authors showed no difference in retention amongst majors in the discipline whether or not ARS were used, though retention among non-majors was significantly improved.

One might suspect that the daily quiz approach promotes less studying by students and "cramming" for short times right before the quiz, as opposed to studying for long hours right before a traditional exam. Students self-report spending 1.7±0.1 hours preparing for each of the 24 daily quizzes, for an aggregate 41 hours. For comparison, a course in physiology taught to the same students the same semester encompassed three exams. Students reported spending 16±2 hours preparing for each of them, for an aggregate 48 hours. Thus in total, students spend approximately the same amount of time in out-of-class study regardless of the approach.

Table 1: Student evaluations of the ARS/quiz approach to summative assessment. Mean Likert scale response (1= strongly agree, 3 = neutral, 5=strongly disagree) ± standard error of the mean.

The daily quiz format force me to study more frequently	1.48±0.08
The ARS was easy to use	1.37±0.06
The quizzes covered a reasonable cross-section of the lecture content	2.04±0.08
I learned more as a result of the daily quiz format	2.1±0.1
I retained more as a result of the daily quiz format	2.6±0.1
I wish more classes would use daily quizzes rather than exams	2.8±0.1
I would sacrifice some lecture content in order to have daily quizzes rather than exams	2.3±0.1
The ARS delivery was better than paper-based quizzes	2.5±0.1
I would rather have a paper-based quiz than the instant feedback of the ARS	3.8±0.1

Discussion

Audience response systems provide a viable approach for rapidly assessing large numbers of students for grading purposes while providing instantaneous feedback. The author is aware of no other quantitative or qualitative studies of the use of ARS as the sole method of delivering summative assessment in a classroom. Our data are consistent with other studies of ARS use in delivering formative assessment; there was no significant improvement in student learning.

Improvements in material retention are also in doubt. This was somewhat surprising. Since each quiz covered the previous two lectures material, it was assumed that retesting was effectively enforced and that the testing effect would give improved outcomes. This does not appear to be the case, though again this portion of the study was not sufficiently well controlled to draw firm conclusions. Nonetheless, it may be necessary to combine the ARS/quiz approach with a traditional comprehensive final exam to realize the benefits of the testing effect.

The true benefit of this approach is in reducing instructor workload and student complains about delayed grades. Students receive their grades *instantaneously* at the end of each assessment using our ARS of choice, though an ARS without a screen would have simply delayed receipt of grades for a few hours – the time necessary to upload scores to the class web site. This is an improvement even over the traditional "bubble sheets" (Scantron) used in large classrooms, which must be delivered to a central location, scanned, and typically generate a small number of misidentified exam results.

The benefits of no manual grading are obvious and cannot be overstated.

This approach is not without its limitations. ARS of the sort we used are expensive, with each handset costing approximately $100 US. While screen-less ARS are available at much reduced cost, they lack the immediate feedback to the student, visual confirmation of answers, and capability of responding to multiple questions at a time.

A principle complaint of the students regarding this approach was that the questions were projected on a screen at the front of the classroom. This forced students to spend roughly equal amounts of time on each question. The students prefer to allocate their time to questions individually. In the next offering of this class, students will be provided with an ARS remote and a paper copy of the quiz and given a time limit. They will respond to the quiz questions through the ARS and will still receive instant feedback, though they may otherwise divide their time among questions as they see fit.

A final issue with ARS is software integration. Not every ARS manufacturer is equal in its integration with common operating systems (e.g. Microsoft Windows®) and campus-wide online instructional tools (e.g. Blackboard). Entering quizzes into SMART Notebook, for example, took a significant amount of time. Further, getting the quiz results into an appropriate format for upload was non-trivial, though the time investment was not large. However, because quiz results are reviewed immediately at the end of the quiz rather than being handed back to students in paper form, the same quizzes can typically be used year after year. Thus over time the effort involved in assessment should be steeply reduced.

Daily quizzes cut significantly into class time. The time to deliver a daily quiz was greater than the time cost of assessing the same material periodically using a traditional exam, even though the time estimate we made in advance put them as being roughly equal. Instructors interested in adopting this approach are cautioned to consider the time necessary to set up the ARS, save results, and review results with the students.

Conclusion

ARS provide expanded options for summative assessment. Quizzes or exams can be delivered more frequently to larger numbers of students with no lag between delivery and feedback. Over the long term instructor effort can be greatly reduced. These benefits are realized without any measurable negative impact on learning or retention. However, while there is no *negative* impact on learning or retention, there is also no evidence either that ARS or low-stakes quizzing, in and of themselves, *improve* learning or retention.

Bibliography

1. Fies C, Marshall J. Classroom Response Systems: A Review of the Literature. *Journal of Science Education and Technology*. 2006;15(1):101-109.

2. Crossgrove K, Curran KL. Using Clickers in Nonmajors- and Majors-Level Biology Courses: Student Opinion, Learning, and Long-Term Retention of Course Material. *CBE Life Sci Educ*. 2008;7(1):146-154.

3. Paschal CB. Formative Assessment in Physiology Teaching Using a Wireless Classroom Communication System. *Advan. Physiol. Edu.* 2002;26(4):299-308.

4. Burnstein R, Lederman L. The Use and Evolution of an Audience Response System. In: Banks DA, ed. *Audience response systems in higher education*. Idea Group Inc (IGI); 2006:40-52.

5. McDaniel MA, Anderson JL, Derbish MH, Morrisette N. Testing the testing effect in the classroom. *The European Journal of Cognitive Psychology*. 2007;19:494-513.

6. Kang SHK, McDermott KB, Roediger HL. Test format and corrective feedback modify the effect of testing on long-term retention. *The European Journal of Cognitive Psychology*. 2007;19:528-558.

7. Guilford W. Teaching peer review and the process of scientific writing. *Adv.Physiol Educ.* 2001;25(1-4):167-175.