A LIVELY ELECTRONIC COMPENDIUM OF RESEARCH, NEWS, RESOURCES, AND OPINION

Astronomy Education Review

Volume 5, Sep 2006 - May 2007
Issue 2

Different Reward Structures to Motivate Student Interaction with Electronic Response Systems in Astronomy

by **Patrick M. Len**
Cuesta College
Received: 05/17/06, Revised: 08/28/06, Posted: 10/19/06

The Astronomy Education Review, Issue 2, Volume 5:5-15, 2007

Abstract

Electronic response systems ("clickers") are used in introductory astronomy classes as a real-time assessment tool. Different reward structures for student responses to clicker questions are used to motivate individual participation or group collaboration before responding. The impact of two reward structures on student behavior and learning is investigated. This study finds that a success-bonus incentive (in which individual participation points are doubled when the class attains a threshold success rate) strongly motivated students to collaborate, whereas a participation-only credit (no success-bonus) incentive resulted in one-third of the students answering individually without collaboration. With a participation-only incentive, students who answered individually ("self-testers") were found to have more positive attitudes toward astronomy and science, and higher self-confidence in their learning than students who interacted before answering without a success-bonus incentive ("collaborators"). These collaborators experienced downward shifts in attitudes and self-confidence, in contrast to the static attitudes and self-confidence of self-testers. The implication is that students with little or no background in science prefer to answer collaboratively rather than independently and that these students are also negatively impacted by a one-semester introductory astronomy course.

1. INTRODUCTION

Cuesta College has implemented a live in-class electronic response system (Classroom Response System from eInstruction) for introductory astronomy classes. Students use individually purchased and registered clickers to answer questions and receive immediate feedback. A multiple-choice question is shown on an overhead projector, and each student responds using a clicker. Each clicker has a keypad with alphabetic choices (A–H) and sends its response and a unique identification number via an infrared beam to a central receiver. Signals from the receiver are compiled on a computer, and a digital projection screen shows which students have responded by highlighting their unique clicker identification number. Information on which answer each student has selected is not shown, but different color codes indicate whether a student has input or changed his or her answer, or selected an answer that is not allowed (e.g., pressing F–H on the clicker keypad when the multiple-choice question has only A–E answers). After the question is "closed" by the instructor or automatically by a countdown timer, student responses are tabulated for each multiple-choice selection. The correct answer is indicated, along with the success rate for the question and the cumulative success rate for the session. Use of this instruction mode in higher education has been discussed extensively (Judson & Sawada 2002), notably in introductory physics classes (Burnstein & Lederman 2001) and more recently in introductory astronomy classes (Duncan 2006).

In this study, two different reward structures for clicker questions were used. *Introduction* clicker questions (Appendix A) were intended to elicit student responses without the pressure to answer correctly. Students were given credit for responding to each introduction question regardless of whether their response was correct. *Review* clicker questions (Appendix B) were intended to encourage peer discussion (Nicol & Boyle 2003); students were credited for responding regardless of the correctness of their answers, as with introduction questions, but scores for a given review session would be doubled if the cumulative class success rate was 80% or higher.

The time allotted for introduction and review questions varied according to their purpose. Introduction questions were brief, avoided the use of technical terms (unless that was the concept explicitly asked about), and focused on the immediate subject at hand. The time allotted for these questions was no more than two minutes. In contrast, review questions were taken from actual quizzes and exams from previous semesters and focused on analytical thinking. The time allotted for these questions was two to three minutes, and more time was given when it seemed that students had not yet adequately discussed the question in depth. Allotting more time or giving hints raised student success rates if a review question was judged "on the fly" to be too difficult. Because the cumulative success rate was displayed after each review question, students were aware of their progress toward attaining the 80% threshold for doubling their participation credit. If by the last question it was not mathematically possible for the class to reach the 80% threshold but they were within a few percentage points of doing so, students were subsequently motivated by the instructor announcing that the success-bonus would be rewarded if the class was 100% correct in responding to the last review question.

Students were relatively quiet when responding to introduction (participation-only) questions, with discussion occurring only among neighboring students. Behavior during review (participation with success-bonus) questions was similar at the start of the semester, until just before the end of the first or second weekly review session. At that point, students collectively realized the importance of the success-bonus incentive in answering correctly. From that point through the rest of the semester, student interaction during review questions was much more animated, with a marked increase in discussion among students and within small groups. Some students shouted out for assistance from the rest of the class; others attempted to coach the rest of the class on how to answer, indicating their answer on the overhead projector using fingers, on the screen using laser pointers, or vocally. The level of interaction and contentiousness during discussion was related to how close the cumulative correct response rate of the class was to matching or exceeding the 80% threshold, which would double students' participation credit.

This study was conducted to measure how the different reward structures of introduction and review clicker questions affected classwide student collaborative behavior. Also investigated was the degree to which students preferred to collaborate, and its correlation with their attitudes toward science.

2. METHODS

2.1 Course Description

Astronomy 10 is a one-semester survey course at Cuesta College. It emphasizes descriptive and conceptual understanding without explicit use of math, although Math 23 (elementary algebra) is the prerequisite. Enrollment ranges from 30 to 90 students per class. Clickers have been used in this course since summer 2004; this study was conducted in fall 2005, with an enrollment of 39 students.

Other teaching modes include traditional lecture, multimedia presentations, and peer-instruction worksheets. The laboratory component of this class (Astronomy 10L) is optional, a separate adjunct course to Astronomy 10.

2.2 Procedures and Student Demographics

Student responses were recorded for a total of 211 clicker questions. A total of 111 introduction questions were asked intermittently throughout each lecture during this semester. In addition, 100 review questions were asked during this semester, with 5–10 questions at the end of each lecture prior to a quiz or exam.

To measure the amount of collaborative behavior that occurred for introduction and for review questions, students participated in a Self-Report of Clicker Use Survey (Appendix C) on the last day of class, categorizing themselves either as *self-testers* ("I usually answer on my own without talking/listening to others") or *collaborators* ("I usually take into consideration what other students are saying before answering") for introduction (participation-only) questions and for review (participation with class success-bonus) questions.

Student attitudes were assessed using the Survey of Attitudes Toward Astronomy (SATA), a 34-question five-point Likert scale questionnaire that measures four attitude subscales (Zeilik & Morris, 2003):

- *Affect:* positive student attitudes toward astronomy and science
- *Cognitive competence:* students' self-assessment of their astronomy/science knowledge and skills
- *Difficulty:* reverse-coded such that high difficulty corresponds to a rating of 1, low-difficulty assessment of astronomy/science corresponds to a rating of 5
- *Value:* students' assessment of the usefulness, relevance, and worth of astronomy/science in personal and professional life

The SATA was administered as a pretest on the first day of class and as a posttest on the last day of class.

Student demographics were determined using the Astronomy Diagnostic Test (ADT, version 2.0). The ADT is a 33-question survey, with 21 questions on a broad range of astronomy concepts based on K–12 science curriculum standards, and 12 demographic questions (Hufnagel 2002). The ADT was administered as a pretest on the first day of class, and without the demographic questions as a posttest on the last day of class.

Out of the enrollment of 39 students, only 36 students (21 male, 15 female) were present for all pretests and posttests (Self-Report of Clicker Use, SATA, and ADT). Results from the three students who had taken the pretests but had not taken the posttests were not reported.

Of these 36 students, 28 students were 20 years of age and younger, five students were aged 21–23, and three students were aged 24–30. Because of the nonexclusivity of ethnicity response categories on the ADT, the majority of the class (28 students) reported themselves as white (non-Hispanic), with three reporting as Asian, one as Hispanic, and five as multicultural/other (for a total greater than 36 students). Similarly, because of the nonexclusivity of degree major (or interest) categories, a plurality of the class (11 students) described themselves as humanities, social sciences, or arts majors, and six students reported themselves as science, engineering, or architecture majors. Four students reported themselves as business majors, four as education majors, and ten as other/undecided, for a total greater than 36 students.

An online survey (Appendix D) was administered midsemester, in which students ranked the relative efficacy of introduction and review questions on a three-point scale, compared with other learning and assessment modes. Of the 36 students who had taken all pretests and posttests, 24 students completed this online survey.

Pearson chi-square tests and student t-tests (paired or independent) were used to evaluate the statistical significance of results. A $p < 0.05$ result was considered significant, whereas a $p < 0.01$ result was considered highly significant. A result for which $p \geq 0.05$ was considered not significant.

3. ANALYSIS AND RESULTS

3.1 Self-Report of Clicker Use Survey

As shown in Table 1, of the 36 students who reported their clicker use and who had taken all pretests and posttests, 13 were self-testers and 23 were collaborators for introduction questions. For review questions, there was only 1 self-tester and 35 collaborators, where all 13 introduction self-testers became review collaborators, and 1 introduction collaborator became a review self-tester. Because a near-unanimity of the class became collaborators for the review questions, for the purposes of further discussion, students will be referred to as self-testers or as collaborators based on their introduction question behavior. The distribution of self-tester or collaborator students (based on introduction questions) with respect to gender or quiz/exam course points was not found to be statistically significant using chi-square analyses. No other demographical constructs were tested for statistical significance because of the small size of this class.

Table 1. Clicker Question Type for Self-Testers and Collaborators		
	Group	
Construct	Self-Testers	Collaborators
Clicker Question Mode [a]		
Introduction (participation-only)	13	23
Review (participation with success-bonus)	1	35

[a] Chi-squared test results: $X^2(1,36) = 12.77$, $p < 0.001$.

The clicker response rate for each student was calculated by dividing the number of introduction or review responses received from that student by the total number of clicker questions of that type (111 introduction questions and 100 review questions), making this rate a raw measure of student attendance and attentiveness. The clicker response rate for the class was determined by the arithmetic mean of the individual student clicker response rates. No statistical significance was found between self-tester and collaborator response rates for either introduction or review questions.

The correctness rate for each student was calculated by dividing the number of correct responses made by that student by the total number of responses made by that student. The correctness rate for the class was determined by the arithmetic mean of the individual student correctness rates. Table 2 shows that the class as a whole had higher correctness rates for review questions than for introduction questions; this result was highly significant [paired *t*-test results: $t(1,35) = –31.4$, $p < 0.0001$]. No statistical significance was found between self-tester and collaborator correctness rates for either introduction or review questions. Note that the mean correctness rate for review questions was just above the required 80% success threshold for the class to double their participation credit.

Table 2. Clicker Response and Correctness Rates		
	Whole Class	
Construct	*M*	*SD*
Response Rates		
Introduction (participation-only)	80%	13%
Review (participation with success-bonus)	80%	14%
Correctness Rates		
Introduction (participation-only)	45% [a]	7%
Review (participation with success-bonus)	82% [a]	7%
[a] Paired *t*-test results: $t(1,35) = -31.4$, $p < 0.0001$.		

3.2 Survey of Attitudes Toward Astronomy

Review of Table 3 shows that collaborators and self-testers did not differ significantly on the pretests for each of the four SATA subscales. Furthermore, self-testers did not show a significant change in attitudes from pretest to posttest for each of the four SATA subscales. However, downward shifts in the mean scores for collaborators were highly significant for both cognitive competence [paired *t*-test results: $t(1,22) = 3.66$, $p < 0.001$], and for value [paired *t*-test results: $t(1,22) = 3.80$, $p < 0.001$].

Although self-testers had slight upward shifts in their mean scores for the same two subscales (cognitive competence and value), these pretest-to-posttest gains were not statistically significant. As a result of both the slight upward shifts in self-tester scores in cognitive competence and value, and the (significant) downward shifts for collaborators in these two subscales, the differences between self-tester and collaborator posttest scores were significant for cognitive competence [independent *t*-test results: $t(1,34) = 2.51$, $p < 0.05$], and highly significant for value [independent *t*-test results: $t(1,34) = 2.93$, $p < 0.01$].

Table 3. Mean SATA Pretest and Posttest Scores for Self-Testers and Collaborators				
	Self-Testers		**Collaborators**	
SATA Subscale	*M*	*SD*	*M*	*SD*
Affect (low = 1, high = 5)				
Pretest	3.6	0.9	3.6	0.6
Posttest	3.4	0.9	3.4	0.8
Cognitive Competence (low = 1, high = 5)				
Pretest	3.6	0.8	3.5 [a]	0.5
Posttest	3.8 [b]	1.0	3.1 [a,b]	0.6
Difficulty of Subject (high = 1, low = 5)				
Pretest	2.8	0.6	2.7	0.4
Posttest	2.9	0.7	2.7	0.5
Value of Subject (low = 1, high = 5)				
Pretest	3.9	0.6	3.8 [c]	0.5
Posttest	4.0 [d]	0.5	3.4 [c,d]	0.7

[a] Paired *t*-test results: $t(1,22) = 3.66$, $p < 0.001$. [b] Independent *t*-test results: $t(1,34) = 2.51$, $p < 0.05$.
[c] Paired *t*-test results: $t(1,22) = 3.80$, $p < 0.001$. [d] Independent *t*-test results: $t(1,34) = 2.93$, $p < 0.01$.

3.3 Astronomy Diagnostic Test

Table 4 summarizes findings from the demographic survey portion of the ADT, which was administered only for the pretest. Because of the overlapping structure of math course sequences at Cuesta College, responses for question 24 (highest completed math course) could not be mapped to reflect a consistent cumulative math background. Students' math and science experiences were instead inferred from their qualitative level of math and science confidence upon entering this course. Self-testers had rated themselves higher than collaborators as being good at math (question 30), good at science (question 31), and expecting that astronomy would be less difficult (question 32, similar to the SATA difficulty subscale). However, only for question 31 (good at science) was there found to be a significant difference between self-testers and collaborators [independent *t*-test results: $t(1,34) = 2.06$, $p < 0.05$].

Table 4. ADT Pretest Responses for Self-Testers and Collaborators

	Self-Testers		Collaborators	
Construct	*M*	*SD*	*M*	*SD*
Pretest Questions (low =1, high = 5)				
How good are you at math?	3.5	0.9	3.0	0.8
How good are you at science?	3.5 [a]	0.9	3.0 [a]	0.7
What best describes the level of difficulty you expect from this course? (1 = *extremely difficult*, 5 = *very easy*)	3.0	0.8	2.7	0.5

[a] Independent *t*-test results: $t(1,34) = 2.06$, $p < 0.05$.

3.4 Online Survey of Learning/Understanding Efficacy

Of the 36 students in the class, 24 students (8 self-testers, 16 collaborators) took this online survey, rating the degree to which introduction and review questions helped with understanding the course material (where 1 = *not at all helpful*, 2 = *somewhat helpful*, 3 = *very helpful*). The rating for traditional lecture by instructor was included for comparison. As shown in Table 5, self-testers gave higher ratings than collaborators for lecture, introduction questions, and review questions. The only statistically significant result was in the higher rating given to lectures by self-testers, compared with the lower rating from collaborators (independent *t*-test results: $t(1,23) = 2.10$, $p < 0.05$).

Table 5. Online Survey Results for Self-Testers and Collaborators				
	Self-Testers		**Collaborators**	
Construct	*M*	*SD*	*M*	*SD*
Helped Learning (low = 1, high = 3)				
Lecture by instructor	3.0[a]	0.0	2.6[a]	0.5
Introduction questions (participation-only)	2.9	0.4	2.6	0.5
Review questions (participation with success-bonus)	2.9	0.4	2.8	0.5
[a] Independent *t*-test results: $t(1,23) = 2.10$, $p < 0.05$.				

4. SUMMARY AND CONCLUSIONS

4.1 Summary of Key Results

The success-bonus reward structure of review questions—credit for participation doubled if the cumulative class success rate was 80% or higher—strongly affected student clicker behavior. A near-unanimity of students became collaborators for review questions, up from nearly two-thirds for introduction questions, and the mean success rate for review questions increased to just over the 80% success-bonus threshold, up from just under 50% for introduction questions.

Students identifying themselves as collaborators for introduction questions had significant downward pretest-to-posttest trends for cognitive competence and value attitudes, and their posttest cognitive competence and value scores were significantly lower than the posttest scores for self-testers.

4.2 Conclusions

Regardless of the differences between self-tester and collaborator attitudes toward science and learning (as discussed in the next section), both types of students had identical success rates in answering introduction questions. The success-reward structure of review questions very strongly encouraged students to interact and cooperate while responding; this resulted in much higher success rates for both self-testers and collaborators in answering review questions, in contrast to less cooperation and lower success rates for introduction questions. This is a striking result for such a simple incentive and demonstrates that students can be persuaded to interact cooperatively in responding to clicker questions.

4.3 Implications for Teaching

Self-testers reported a higher pretest proficiency in science and maintained positive attitudes toward science, in contrast to collaborators, who reported a lower pretest proficiency in science and subsequently experienced a negative shift in their attitudes toward science. The implication is that this one-semester astronomy course had a great impact on the formation of nascent science and learning attitudes of students with little or no background in science. This is consistent with earlier findings by Zeilik and Morris (2003), who reported no change in attitudes for students in an astronomy course with college-level physics as a prerequisite or corequisite, noting that one semester of astronomy (following an extensive amount of exposure to science) had little effect on their attitudes toward science. A possible explanation is that students with more exposure to science would be more motivated to "self-test" their knowledge during inquiry questions. It is plausible that self-testers (with more exposure to learning science than collaborators) would have well-formed positive attitudes toward astronomy. In contrast, students with less exposure to science would be less comfortable answering on their own and would collaborate more before answering.

Students already regard clickers as an effective mode of instruction and assessment (Judson & Sadawa 2002). The challenge is in maximizing the impact of clickers in fostering beneficial collaborative behavior, while contributing to positive shifts in attitudes toward science and learning. The downward shifts in attitudes demonstrated by the collaborators in this study could, it is hoped, be minimized or even reversed in future classes by first using SATA pretest scores to identify students with negative attitudes and lack of exposure to science. Because these students would be predisposed toward collaborative behavior, engaging students in this mode using clickers should be made as constructive and meaningful as possible in order to stimulate positive analytical discussion, especially during review questions, rather than passive "herd mentality" behavior.

4.4 Implications for Future Research

Subsequent investigations should involve accumulating data from many consecutive semesters in order to generate a larger student population for statistical analysis. Student interviews to elicit reasons that students engage in self-testing or collaborative behavior could validate the self-reporting of their clicker response behavior and determine if students are "hard-wired" self-testers and/or collaborators, or instead track how they "evolve" into these roles.

Future research questions to investigate are the correlation and/or causation between pretest attitudes and collaborative behavior while answering introduction questions, and whether the implementation of clickers in this class has any impact (positive, negative, or none) on students with low cognitive competence attitudes toward science.

Acknowledgments

This research was supported by the 2005-2006 @ONE Carnegie Scholar Program hosted by Evergreen Valley College (John Whitmer, Mary Wong-Kauzlarich, and Judy Rookstool). Michelle Barton (Palomar College) and Darla Cooper (Oxnard College) provided valuable input and feedback regarding the design and analysis of this research project.

References

Burnstein, R. A., & Lederman, L. M. 2001, "Wireless Keypads in Lecture Classes," *The Physics Teacher*, 39(1), 8.

Duncan, D. 2006, "Clickers: A New Teaching Aid with Exceptional Promise," *Astronomy Education Review*, 5(1), 70. http://aer.noao.edu/cgi-bin/article.pl?id=194.

Hufnagel, B. 2002, "Development of the Astronomy Diagnostic Test," *Astronomy Education Review*, 1(1), 47.

Judson, E., & Sawada, D. 2002, "Learning from Past and Present: Electronic Response Systems in College Lecture Halls," *Journal of Computers in Mathematics and Science Teaching*, 21(2), 167.

Nicol, D. J., & Boyle, J. T. 2003, "Peer Instruction Versus Class-wide Discussion in Large Classes: A Comparison of Two Interaction Methods in the Wired Classroom," *Studies in Higher Education*, 28(4), 458.

Zeilik, M., & Morris, V. J. 2003, "An Examination of Misconceptions in an Astronomy Course for Science, Mathematics, and Engineering Majors," *Astronomy Education Review*, 2(1), 101.

Appendixes

Click here to access Appendixes A–D in a single PDF document.

[URL: http://aer.noao.edu/auth/len-appendixes.pdf]