Interdisciplinary Journal of E-Learning and Learning Objects Volume 7, 2011

Facilitation of Formative Assessments using Clickers in a University Physics Course

David M. Majerich
Institute for Schools and Society, College of Education, Temple University, Philadelphia, PA, USA

majericd@temple.edu

Judith C. Stull
Department of Sociology, LaSalle University, Philadelphia, PA, USA

stull@lasalle.edu

Susan Jansen Varnum and Tiffany Gilles
Department of Chemistry, Temple University, Philadel-phia, PA, USA

susan.varnum@temple.edu; tgilles@temple.edu

Joseph P. Ducette
Department of Psychological Studies in Education, Temple University, Philadelphia, PA, USA

joseph.ducette@temple.edu

Abstract

This study provides an empirical analysis of the integration of clickers, used to facilitate formative assessments, in a university physics course. The sample consisted of students from two consecutive semesters of the same physics course, where one group used clickers and the other did not. Data included pre- and post-attitudinal and behavioral surveys, physics and mathematics pre-tests, two course examinations, and one cumulative final examination. The clicker group completed seven clicker episodes (weekly multiple choice questions and in-class discussion of results). On average, students who participated in clicker episodes achieved significantly higher scores on the cumulative final examination compared to the other group. Regression analysis was used to control for differences among the students and to quantify the effect of clicker use. The regression results indicate that controlling for all of the entered variables, for every one more clicker episode the student responded to, the final grade was raised by 1.756 points. Thus, if a student took all seven of the "clicker quizzes," the final grade would have been 12.3 points higher, a difference of a grade. Interestingly, how well the student did on these "clicker quizzes" never proved significant in the regression analyses. In an analysis of the residuals, grades were more important to those who performed better than expected as compared to those who performed less well than predicted. In sum, using clicker episodes appeared to result in improved achievement but more research is needed to support these findings.

Editor: Heinz Dreher

Keywords: student response system, formative assessment, clickers, learning objects, clicker episodes

Introduction

Faculty in the science education community are being charged to replace traditional methods of teaching in the large lecture hall with more learner-centered, student-engaged, interactive strategies informed by what is now known about how many students learn (Bransford, Brown, & Cocking, 2000). While the traditional methods of teaching have long been associated with disconnecting the students from both the instructor and course material, causing students to assume a more passive role in the learning process, encouraging memorization over conceptual understanding of course material, and treating students as if they learn course material at the same time and in the same way, these methods are common in many of today's lecture halls (Mintzes & Leonard, 2006; Sunal, Wright, & Day, 2004). In better preparing students for the skills needed for success in the 21st century (Floden, Gallagher, Wong, & Roseman, 1995; Partnership for 21st Century Skills, 2010), using new technologies during instruction that are interactive have shown to assist faculty in creating active learning environments whereby students learn by doing, receive feedback during the learning trajectory, construct new knowledge and improve skills, and continually refine their understandings of course material (Bereiter & Scardamalai, 1993; Hmelo & Williams, 1998; Mintzes & Leonard, 2006). While supporting research has shown increased student achievement and improved behavioral outcomes for students who are actively engaged with the course content and increased dialogue and interaction with the instructor and peers (Crouch & Mazur, 2001; Mintzes & Leonard, 2006; Slater, Prather, & Zeilik, 2006), "a key instructional implication from the research on learning is that students need multiple opportunities to think deeply and purposefully about the content and to gain feedback on their learning" (Ueckert & Gess-Newsome, 2006, p. 147). Options available to instructors that have been used to engage students and promote an active learning environment in the large lecture hall are Audience Paced Feedback, Classroom Communication Systems, Personal Response Systems, Electronic Voting Systems, Student Response Systems, Audience Response Systems, voting-machines, and zappers (MacArthur & Jones, 2008). Each of these systems has also been referred to as 'clickers' (Duncan, 2005; MacArthur & Jones, 2008). In the most fundamental sense, clickers are radio-frequency, battery powered, hand-held devices that are part of an electronic polling system. The predominant research about the clicker use has been shown to promote student discussion, increase engagement and feedback, and improve attitudes toward science (Cutts, 2004: Draper & Brown, 2004; Duncan, 2005; Latessa & Mouw, 2005). However, an extensive 2009 review of the literature revealed a paucity of empirical peer-reviewed evidence to support the claims that the technique can be used to improve student achievement (Mayer, et al., 2009). Although several research efforts report positive effects of clicker use on students' achievement (Addison, Wright, & Milner, 2009; Hoffman & Goodwin, 2006; Kennedy & Cutts, 2005; Watkins & Sabella, 2008), the empirical evidence suggested by Mayer et al. (2009) that is needed to corroborate existing results and substantiate any claims for using clickers requires additional studies. This study aims to provide evidence from university physics classes.

Review of Related Literature

General Clicker Device Features and Uses

In general, clicker devices have a keypad (alpha, numeric, or alpha/numeric buttons) resembling a television remote control device or small electronic calculator. Using presentation software, the instructor poses a question (multiple choice or true-false formats). Students respond by selecting their answer choice and using the corresponding button on their devices. After students submit their responses, a receiver and related clicker software collect and tabulate the students' electronic signals. A computer is used to display the collective results graphically at the front of the classroom. Clicker systems save the responses from each student, and the data can also be exported to spreadsheet software. For some examples of the specific features of clicker devices, as well related technology and software compatibility-requirements, see Barber and Njus (2007) and Burnstein and Lederman (2003).

'Clickers' (Duncan, 2005) are classified as a type of instructional technology that affords students multiple opportunities to participate in their own learning processes, be actively engaged with the course content and the instructor, and receive frequent assessment feedback in real time. Both the students and the instructor can see the level of course material learned (Bergtrom, 2006; Duncan, 2005). Although clickers can be used for simply taking attendance or quizzing students to see if they have prepared for class, it has been suggested that they are most effective when used to challenge students to think about their understanding of the material under discussion (Barber & Njus, 2007; Duncan, 2005).

In a comprehensive review of the literature and empirical research on the use of clicker questions, Caldwell (2007) identified nine general strategies. Presented here in summative, clickers were used:

- to increase or manage interaction
- to assess student preparation and ensure accountability
- to find out more about students' opinions and attitudes
- for formative assessment (e.g., assess students' understanding, determine future direction of lecture)
- for quizzes and tests
- to do practice problems
- to guide thinking, review, or teach
- to conduct experiments on or illustrate human responses; and,
- to make the lecture fun (pp. 10-11).

When clickers were used as a type of formative assessment, the results revealed students' misunderstandings of the course material (Wood, 2004), determined students' readiness to continue after solving a problem (Poulis, Masses, Robens, & Gilbert, 1998), and afforded opportunities to self-assess their understandings at the end of class (Halloran, 1995). Although these studies have shown improved student learning from clicker use, comparisons among these studies and others (MacArthur & Jones, 2008) are compromised as researchers have used different methods or failed to report them at all. This study aims to contribute evidence from university physics classes when clicker-based quizzes were used as formative assessments, whereby a protocol for the formative assessment procedure is made explicit.

Learning Objects and Interoperability

"Learning object" (LO) is a broad conceptual term that can possess multiple meanings (Bergtrom, 2006; Thompson & Yonekura, 2005). LOs can include animation and video segments all the way to complete modules or lessons (Thompson & Yonekura, 2005). While a broad conceptualization of LOs advanced by Harman and Koohang (2006) focused on online discussion boards, Bergtrom (2006) extended their conceptualization to include clickers. Clickers "are learning objects in the same sense that performance art is art – they have learning (or in the case of art, aesthetic) value despite their transience" (p. 108), even though clickers are confined to the in-class presentation. For the purpose of this paper, the conceptualization of LOs advanced by Harman and Koohang (2006), and extended by Bergtrom (2006) to include clickers, is used.

Contrary to the general idea that learning objects are interacted with individually or with a group, the use of clickers requires instructor facilitation. For the present study, clicker quizzes were presented to the students as PowerPoint^{tm} slides. In the most fundamental sense, each clicker quiz slide could be an interactive LO mediated by the instructor (Bergtrom, 2006). Students were presented with a slide, asked to solve a physics problem, and then submit their answers using their InterWrite^{tm} Personal Response System clicker device. However, clickers are more than the hand-held device students use to send their "votes" (answers) to a computer-based analysis-report-display system. As the results are discussed by the instructor and the students, the instructor guides the students towards knowledge and skills that parallel the knowledge and skills used by experts in the field (Mintzes, Wandersee, & Novak, 1998). During the discussion, both the instructor and students see learning has occurred. For the purpose of this present study, a "clicker episode" begins with the presentation of a clicker question and ends when the students understand the principles underlying the question.

In this research, each clicker quiz was a multiple choice question from a specific topic. When clicker questions are used, they "allow the assembly/disassembly of broad subject matter into component structural elements, ideas, concepts, and ways of thinking" (Bergtrom, 2006, p. 2). The content for each of the clicker quizzes is self-contained, i.e., instructional message aligned to a specific learning objective, and could be used at any time during the teaching-learning sequence to assess students' understanding of material presented to them. Furthermore, the 'chunking' of broad topics into multiple subtopics creates learning objects of fine granularity (Bergtrom, 2006; Fournier-Viger, Najjar, Mayer, & Nkambou, 2006). Since clickers are meant to be managed and administered throughout the learning process to support student learning and inform how instruction needs to be changed in order to accommodate students' needs, their use can facilitate formative assessment (Bergtrom, 2006; Crumrine & Demers, 2007). Clickers have an educational purpose (McGreal, 2004) and have pedagogical value, are learner-centered, and can be contextualized by the students (Duncan, 2005; Harman & Koohang, 2005). In addition, the LOs can be encapsulated, stored, and reused in appropriate contexts.

Science Education Reform and Assessment

Discussions of the role of assessments frequently take center stage in the arena of science education reform debates. As delineated in the *National Science Education Standards* (NRC, 1996), "assessments provide an operational definition of standards, in that they define in measurable terms what teachers should teach and students should learn" (pp. 5-6). Furthermore, "when students engage in assessments they should learn from those assessments" (p. 6). Extended to colleges and universities undergoing undergraduate science education reform (Siebert & McIntosh, 2001), this perspective suggests that teaching, assessment, and curriculum are mutually reinforcing and need to be aligned in order to optimize learning experiences and maximize student learning outcomes. While the curriculum is already established in many college and university courses, and if assessment and learning are two sides of the same coin (NRC, 1996), it would seem reasonable that administering frequent assessments, analyzing their results, and sharing them with students, could inform changes to instruction needed in order to accommodate learners' needs for continued learning.

As generally understood, assessment is used by most instructors to determine what learning has occurred when compared to course expectations and is the basis for the assignment of grades to overall achievement. This type of assessment is *summative* and is the measurement of achievement at the end of a teaching-learning sequence. Assessment is *formative* when frequent evidence during the students' learning process is gathered and analyzed, where the results inform changes needed to instruction in order to meet students' needs, and provide students with feedback about their learning (Black & Wiliam, 1998). The results of formative assessments have been described as providing 'snapshots' of what students know and can do at specific junctures in their learning processes (Treagust, Duit, & Fraser, 1996). Where traditional assessments have been criticized for merely gauging what students know instead of probing what students know (McClymer & Knoles, 1992; McDermott, 1991; Mintzes & Leonard, 2006; Pride, Voos, & McDermott, 1997), feedback from formative assessments can align teaching with learning (Black & Wiliam, 1998; Yorke, 2003). The feedback to the instructor illuminates both changes needed in instruction and the degree to which instruction was successful. Feedback to the students helps highlights problem areas and provides reinforcement for continued learning. Since formative assessment has been identified as a key predictor of student achievement (Black & Wiliam, 1998: Bransford et al., 2000), its use has been recommended for integration into curriculum as part of the learning process whereby students can self-regulate their own learning (Nicol & Macfarlane-Dick, 2006).

Six-Stage Model of Formative Assessment

The model of a formative assessment that informs this present study is derived from the theoretical perspective described by Yorke (2003). The model is dynamic, recurs throughout the teaching-learning sequence, and has been modified and used elsewhere (Stull, Schilller, Jansen Varnum, & Ducette, 2008). The model is conceptualized as having six stages. Specifically, the instructor develops a lesson and related assessment based on the students' preparedness and prior knowledge (Stage 1). The instructor presents the lesson (Stage 2). The instructor administers an assessment (Stage 3). Together the instructor and students consider the assessment results (Stage 4). Dialogue between the instructor and students begins (Stage 5). Thereafter, the instructor determines if reinstruction is warranted for the previously taught lesson or proceeds to the next lesson (Stage 6). In this model, formative assessment is theorized and the connections between roles of the instructor and students are made explicit. While the instructor determines when and how often to administer formative assessments and modify instruction to optimize learning experiences to accommodate students' needs, the key is to provide sufficient informative feedback to students so that students can chart their development, adjust their learning styles, and maximize their learning (Yorke, 2003). While some recommended that formative assessment must be continuous (Brown, 1999; Heritage, Kim, Vendlinski, & Herman, 2009), it has been suggested that it "can be very occasional, yet still embody the essential supportiveness towards student learning" (Yorke, 2003, p. 479).

It has been argued that the time has come for formative assessments to receive greater prominence in the learning process (Black & Wiliam, 1998; Bransford et al., 2000; Layng, Strikeleather, & Twyman, 2004); however, in reality, a combination of formative and summative assessments should be incorporated into the overall teaching-learning sequence (Black & Wiliam, 1998). In doing so, the instructor must switch roles from being a supporter of learning to judging the students' overall achievement (Ramsden, 1992). It is through the analysis of frequent formative assessment results and continued dialogue with the students that the instructor will gain a sense of when the shift in roles needs to occur.

Methodology

Learning Environment

This study was conducted to determine the effect that increased feedback from clicker episodes (formative assessment) had on students' physics achievement (summative assessment) for students who used clickers when compared to students who were nonusers.

Students who enrolled in this physics course were mostly science and health profession majors and took this course to fulfill either a university core requirement or a major requirement. Taught in the large lecture hall, enrollment numbers generally range between 150-250 students per course. While all students are taught together during the lecture by the same instructor, the students are required to register for recitation and laboratory sections which generally have 25-40 students and are taught by other instructors. The lecture textbook is used as a curriculum guide and is the source for course content and assigned problem sets. Lecture examinations require that students recall knowledge (facts, concepts, theories, laws, principles, generalizations, models), but mostly solve problems. The cumulative final examination requires that students also recall knowledge, but mainly to solve problems.

Methods and Subjects

This study was conducted at a large, public, urban university in the mid-Atlantic region. Data were obtained from two fifteen-week introductory physics courses that met twice a week for 80 minute periods over two semesters taught by the same instructor. In the fall and spring semesters of the course, respectively, 157 and 152 students participated. The fall semester course was traditionally taught and the following spring semester course had clicker episodes (formative assessments) integrated into the instruction. Each learning object episode began with a multiple-choice question associated with a specific course topic, followed by a discussion of the results. The results of the clicker-based questions were collected, tabulated, and results displayed for students at the beginning of the next scheduled class. Problems areas were identified and provided the topic for discussion for the instructor and students. Based on the discussion, the instructor made appropriate adjustments to the instruction. In the end, the spring semester students (clicker group) completed a total of seven formative assessments during weeks 5-7, 9-11, and 13.

In addition to the “clicker quizzes”, attitudinal and behavioral surveys were administered at the beginning and the end of the semester as well as pre/post tests which included physics and mathematics questions (week 1). Among the attitudinal data collected were the students’ perceptions about the usefulness of class activities (group work and group grade, student-lead whole class discussions, model making, descriptions of reasoning, investigations, presentations, self-evaluation of learning, decisions about course activities, assessments results modifying what is taught) and hours spent on activities (watching television, playing computer games, socializing, doing jobs at home, playing sports, working, attending classes, and doing homework). Students took two course examinations (weeks 6 & 11) and a cumulative final examination (week 15). The cumulative final examination contained a set of questions representative of the major course topics discussed over the semester. The fall semester students comprised the control group. All protocols in this study were approved by the university’s institutional review board.

Results

Equivalent Groups

Both groups suffered the loss of students. The attrition rates for the control and clicker groups were 20.4% and 23.0%, respectively; however the difference of proportions was not significant. It is expected that the more challenged students have a higher probability of withdrawing from the class. Accounting for self-selection bias, it is acknowledged that the groups’ content and skill sets should be better at the end of the course than at the beginning.

Maximum possible points for the physics and mathematics pretests were 7 and 25 points, respectively. Pretest scores were determined by applying a two-point rubric (1=correct solution; 0=incorrect or no solution) to students’ solutions. Points for each pretest were summed separately. Pretest scores were converted to percentages based on the number of correct solutions divided by the total number of questions. Results of the pretests given at the beginning of the semester revealed the control group's pretest physics percentage scores (M=31.4%, SD=11.3%) were higher than the clicker group (M=30.7%, SD=11.3%), but the difference was not statistically significant. The clicker group's pretest mathematics percentage scores (M=57.3%, SD=23.1%) were higher than the control group (M=56.8%, SD=21.5%), but again were not statistically significantly different. Based on these results, the groups were equivalent.

Regression Analyses

Regression analysis does assume interval level variables. However, categorical variables are accommodated. In the case of a nominal level variable as the dependent variable, discriminant function is the appropriate form to use. If the dependent variable is ordinal, then logistic regression analysis is the correct form. In the case of categorical variables (either nominal or ordinal) as independent or predictor variables, the appropriate technique is to decompose the variable into "dummy" variables where the option is either "present" or "not present." For example, if social class categorized into "upper," "middle," and "lower," three separate variables would be made and then, not to encounter the statistical problems (such as a variable being "forced" out of the analysis), up to two could be entered into the analysis. The coefficient is interpreted as in comparison to the group or groups that have been excluded from the analyses (Gujarati, 1988).

While the use of ANOVA in this context will determine if one group is different from another, that is all it will tell (Knoke, Bohrnstedt, & Mee, 2002). Regression analysis allows for testing a more complex model, one in which confounding differences between or among groups can be addressed as predictor and control variables can be entered. Also the size of the effect, i.e., what a one unit change in the predictor resulted in the dependent variable, can be captured. Unstandardized coefficients are necessary as that is the only means of capturing these "effect size" contributions of the variables. The standardized coefficients give only strength of relationship. There may be a very strong relationship between two variables, but the size of the effect may be very small.

Regression analysis was used to control for differences among students and to quantify the effect of clicker use. In the model for predicting the students' physics achievement, the dependent variable was the student's final examination score and the independent variables were the physics/mathematics pretest score, the number of clicker quizzes taken, whether the course was a required one, the number of different types of assessments the student had previously experienced, the number of hours per week the student reported working, and whether the student was male. In specifying the model, the percentage of correct answers on any quiz were entered, but never proved significant. These percentages were also averaged over all of the quizzes and then entered into the regression analyses. This never proved significant as well. They were dropped from the analyses as a result. Table 1 summarizes the means and standard deviations.

Table 1. Means and Proportions of Central Tendencies and Dispersion for Variables Predicting Students' Physics Achievement (N=329)

Variable	M	SD	Minimum	Maximum
Physics/Mathematics pretest score, 0-200	87.90	27.96	12.50	166.67
Number of clicker episodes taken, 1-7	1.79	2.42	0	7.00
Was this a required course? (1=Yes, 0=No)	.92	.27		
Number of types of assessments student had experienced, 0-9	8.18	2.12	0	9.00
Number of hours students works at a job per week, 0-40	5.05	6.13	0	40.00
Male student dummy (1=Yes, 0=No, female student)	.63	.48		

Of these variables the pretest score was included to control for differences in students' abilities. Whether the course was a required one was included to control for student interest variations. The number of different types of activities experienced by the student in class that they thought were helpful was included. The student's gender was also included as a control as research has shown that males have a greater interest in physics than do females (Halpern, Benbow, Geary, Gur, Hyde, & Gernsbacher, 2007; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Seymour & Hewitt, 1997). Lastly, the number of hours per week the student worked was included to control for SES differences and the amount of time the student could devote to studying. Table 2 summarizes the results of the regression analysis.

Table 2. Summary of Regression Analysis for Variables Predicting Students' Physics Achievement

Model	Unstandardized Coefficients	Standardized Coefficients		
	B	Beta	t	Sig.
Physics/Mathematics pretest score		-.059	-.952	ns
Number of clicker episodes taken	1.756	.230	3.298	.001
Was this a required course?	-1.799	-.030	-.485	ns
Number of types of assessments student had experienced	-1.881	-.413	-5.895	.000
Number of hours students works at a job per week	-.109	-.036	-.594	ns
Male student dummy	-.624	-.017	-.277	ns
(Constant)	58.142		9.567	

The R Square equaled 33 indicating that the included variables explained 33% of the variation in the dependent variable. In the end, two variables proved significant – the number of clicker episodes and the number of different types of assessments the student had experienced. In all, there were seven clicker episodes. The regression results indicate that controlling for all of the entered variables, for every one more clicker episode the student took, the final grade was raised by 1.756 points. Thus, if a student took all seven of the "clicker quizzes," the final grade would have been 12.3 points higher, a difference of a grade. Interestingly, how well the student did on these "clicker quizzes" never proved significant. The number of different types of assessments the student has experienced (e.g., group work done with one grade assigned to the group; participation in whole-class discussions where instructor talked less than the students; descriptions written of student's own reasoning; investigative activities performed including data collection and analysis; presentations designed and made by students to learn class concepts; the extent of student's own learning evaluated; decisions about course activities voiced by students; student assessment results modified what was taught and how) negatively related to how well they did on the final exam. Perhaps what is needed is consistency in assessing learning.

Different instructional strategies resonate with students. To understand who benefitted more from the inclusion of these clicker episodes into the course, a residual analysis was performed. First, the actual final course score was subtracted from what was predicted in the regression analysis and then the students were separated into those who did well above what was predicted (70 percentile and above), those in the middle of the distribution, and those who performed well below expectations (30th percentile and below). It should be noted that those in the "better than predicted" group need not have done well. They may have actually been in the lower part of the grade distribution. What is important is that they did significantly above what was predicted in the regression analyses. Also, students in the "worse than predicted" group may have earned good grades, but their grades were not as high as what was predicted in the regression analyses. In using the regression in this manner, differences in presenting abilities are "controlled for." The coefficients capture the "value added" by what was done in the course. See Figure 1 for further details.

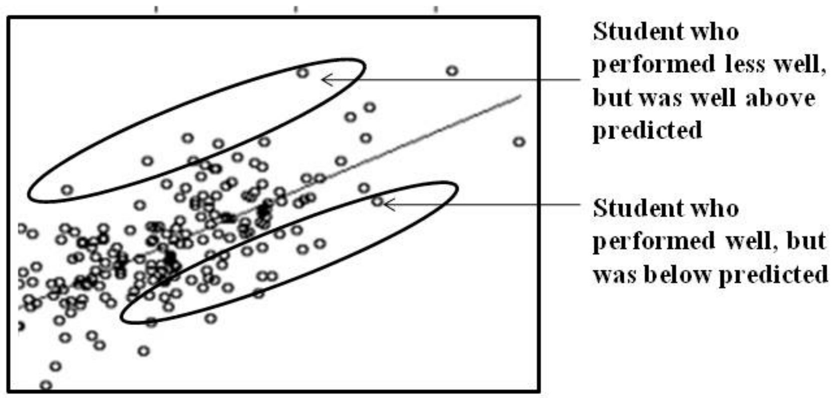


Figure 1. Regression plot

On both the pre-attitudinal and behavioral surveys, the students were asked about how they allocated their time in an average week. The categories were watching television and videos, playing computer games, socializing or talking with friends outside of school, doing jobs at home, playing sports, working, attending classes, and doing homework or other class related activities. Chi square tests indicate no statistically significant difference between those who did better than predicted and those who did worse than predicted. The students were also asked about how important grades were to them. On this variable, "How important are good grades to you?," students selected among the following responses: "Not important, Somewhat important, Important, or Very important." The two groups did differ on this variable; grades were more important to those who did better than expected than those who did not as shown in Figure 2.

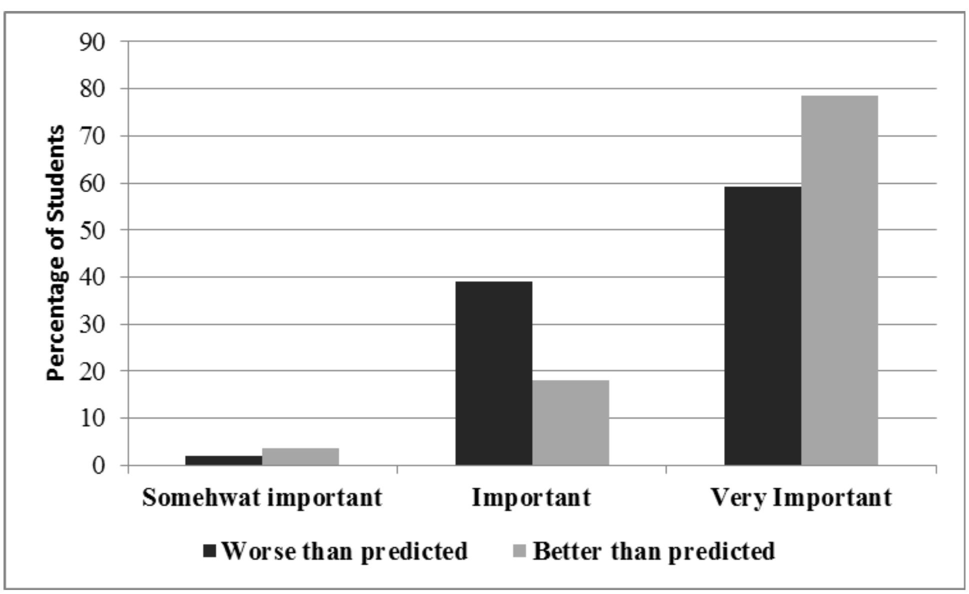


Figure 2. Distribution of Student Responses to How Important Grades Are by Whether They Were Well Below Expectations or Were Well Above Expectations

Conclusion

While there is an abundance of anecdotal information that advocates the use of clickers to improve student achievement in the science classroom, this study offered results to substantiate the claim. It is apparent that integrating clicker episodes, in this case weekly formative assessments consisting of multiple choice questions with in-class discussion of results, had a significant effect on student achievement. On average, students who used clickers achieved significantly higher scores on the cumulative final examination compared to the other group. The regression results quantified the effect. In sum, using clicker episodes did prove to be positively associated with improved achievement, but this is offered with caution as learning is a complex process and more data are needed on students' attitudes and behaviors.

However, there are some unresolved issues still to be addressed. While the students in both classes performed equally as well on the first examination without using clickers, lower scores on the second examination were obtained by the students who used clickers. To what extent are there delayed effects on students' learning and their metacognitive learning when using clicker episodes? How lasting are the effects of clicker use? Does clicker use apply equally well to all learning situations? These issues need to be studied as additional empirical evidence is gathered to support the use of clickers to improve student achievement and to corroborate anecdotal information about their use. In using clickers, instructors can uncover more about their students and what their students know about themselves during the learning process, and can be better informed about changes to instruction needed to promote student learning and achievement in the science classroom.

References

Addison, S., Wright, A., & Milner, R. (2009). Using clickers to improve student engagement and performance in an introductory biochemistry class. *Biochemistry and Molecular Biology Education, 37*(2), 84-91.

Barber, M., & Njus, D. (2007). Special feature: Clicker reviews. *CBE-Life Science Education, 6,* 1-20.

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise.* Chicago, IL: Open Court Publishing.

Bergtrom, G. (2006). Clicker sets as learning objects. *Interdisciplinary Journal of Knowledge and Learning Objects, 2*, 105-110. Retrieved from http://ijello.org/Volume2/v2p105-110Bergtrom.pdf

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Brown, S. (1999). Institutional strategies for assessment. In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education: Choosing and using diverse approaches* (pp. 3-13). London: Routledge Press.

Burnstein, R. A., & Lederman, L. A. (2003). Comparison of different commercial wireless keypad systems. *Physics Teacher, 41*, 272-275.

Caldwell. J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Science Education, 6*, 9-19.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics, 69,* 970-977.

Crumrine, T., & Demers, C. (2007). Formative assessment: Redirecting the plan. *The Science Teacher, 74*(6), 64-68.

Cutts, Q. (2004, November/December). *Using an electronic voting system to promote active reflection on coursework feedback.* Paper presented at the 4th International Conference in Education, Melbourne, Australia.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20*(4), 81-94.

Duncan, D. (2005). *Clickers in the classroom: How to enhance science teaching using classroom response systems*. San Francisco: Pearson/Addison-Wesley.

Floden, R., Gallagher, J., Wong, D., & Roseman, J. (1995, April). *Seminar on the goals of higher education.* Paper presented at the annual conference of the American Association of Science, Atlanta, GA.

Fournier-Viger, P., Najjar, M., Mayers, A., & Nkambou, R. (2006). A cognitive and logic based model for building glass-box learning objects. *Interdisciplinary Journal of E-Learning and Learning Objects, 2*, 77-94. Retrieved from http://www.ijello.org/Volume2/v2p077-094Fournier-Viger.pdf

Gujarati, J. (1988) *Basic econometrics*. New York, NY: McGraw-Hill.

Halloran, L. (1995). A comparison of two methods of teaching computer managed instruction and keypad questions versus traditional classroom lecture. *Computers in Nursing, 13*(6), 285-288.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in Public Interest, 8*, 1–51.

Harman, K., & Koohang, A. (2005). Discussion board: A learning object. *Interdisciplinary Journal of Knowledge and Learning Objects, 1,* 67-76. Retrieved from http://ijklo.org/Volume1/v1p067-077Harman.pdf

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24-31.

Hmelo, C., & Williams, S. M. (1998). Learning through problem solving. *The Journal of the Learning Sciences, 7*(3 & 4).

Hoffman, C., & Goodwin, S. (2006). A clicker for your thoughts: Technology for active learning. *New Library World, 107*(1228/1229), 422-433.

Kennedy, G. E., & Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning, 21*, 260-268.

Knoke, D., Bohrnstedt, G., & Mee, A. (2002). *Statistics for social data analysis*. Belmont, CA: Thomson/Wadsworth.

Latessa, R., & Mouw, D. (2005). Use of an audience response system to augment interactive learning. *Family Medicine, 37*(1), 12-14.

Layng, T., Strikeleather, J., & Twyman, J. (2004, May). *Scientific formative evaluation: The role of individual learners in generating and predicting successful educational outcomes.* Paper presented at the National Invitational Conference on The Scientific Basis of Educational Productivity, sponsored by the American Psychological Association and the Laboratory for Student Success, Arlington, VA.

MacArthur J. R., & Jones, L. L. (2008). A review of literature reports of clickers applicable to college chemistry classrooms. *Chemistry Education Research and Practice, 9,* 187-195.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397–416.

Mayer, R. E., Stull, A., DeLeeuw, K., Ameroth, K., Bimber, B., Chun, D., . . . Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Education Psychology, 3,* 51-57.

McClymer, J. F., & Knoles, L. S. (1992). Ersatz learning, inauthentic testing. *Excellence in College Teaching, 3,* 33-50.

McDermott, L. C. (1991). Millikan lecture 1990: What we teach and what is learned – Closing the gap. *American Journal of Physics, 59*(4), 301-315.

McGreal, R. (2004). Learning objects: A practical definition. *International Journal of Instructional Technology and Distance Learning, 1*(9). Retrieved from http://www.itdl.org/journal/sep_04/article02.htm

Mintzes, J. J., & Leonard, W. H. (2006). *Handbook of college science teaching.* Arlington, VA: NSTA Press.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (1998). *Teaching science for understanding: A human constructivist view.* San Diego, CA: Academic Press.

NRC (National Research Council). (1996). *National science education standards.* Washington, DC: National Academy Press.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education,* 31(2), 199-218.

Partnership for 21st Century Skills. (2010). *21st Century Learning Environments.* Retrieved from http://www.p21.org/documents/le_white_paper-1.pdf

Poulis, J., Masses, C., Robens, W., & Gilbert, M. (1998). Physics lecturing with audience paced feedback. *American Journal of Physics, 66*(5), 439-441.

Pride, T. O., Vokos, S., & McDermott, L. C. (1997). The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems. *American Journal of Physics, 66*(2)*,* 147-157.

Ramsden, P. (1992). *Learning to teach in higher education.* London: Routledge.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO: Westview Press.

Siebert, E. D., & McIntosh, W. J. (2001). *Pathways to the science education standards.* Arlington, VA: NSTA Press.

Slater, T. F, Prather, E. E., & Zeilik, M. (2006). Strategies for interactive engagement in large lecture science survey classes. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 45-54). Arlington, VA: NSTA Press.

Stull, J., Schiller, J., Jansen Varnum, S., & Ducette, J. (2008). The use of formative assessment in university level mathematics courses. *Journal of Research in Education, 18*, 58-67.

Sunal, D. W., Wright, E. L., & Day, J. B. (2004). *Reform in undergraduate science teaching for the 21st century.* Greenwich, CT: Information Age Publishing.

Thompson, K., & Yonekura, F. (2005). Practical guidelines for learning object granularity from one higher education setting. *Interdisciplinary Journal of Knowledge and Learning Objects, 1,* 163-179. Retrieved from http://ijklo.org/Volume1/v1p163-179Thompson.pdf

Treagust, D.F., Duit, R. & Fraser, B (Eds.). (1996). *Improving teaching and learning in science and mathematics.* New York: Teachers College Press.

Ueckert, C., & Gess-Newsome, J. (2006). Active learning in the college science classroom. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 147-154). Arlington, VA: NSTA Press.

Watkins, E., & Sabella, M. (2008). Examining the effectiveness of clickers in promoting learning by tracking the evolution of student responses. *Proceedings of the 2008 Physics Education Research Conference, Edmonton, Alberta, CA*, 1064, 223-226. doi: 10.1063/1.3021260

Wood, W. B. (2004). Clickers: A teaching gimmick that works. *Developmental Cell, 6*(6), 796-798.

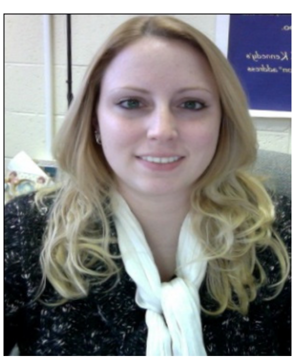
Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 4*(54), 477-501.

Biographies

David M. Majerich, Ed. D., is a science education researcher for the Institute for Schools and Society and is an adjunct professor of science education at Temple University. His research interests include improving the way that science is taught using demonstrations in the large lecture hall and studying the effects of infusing research-based models of teaching in science methods courses has on beginning teachers' understanding of science content and their perceptions about their ability to teach science. He holds an Ed.D. from Temple University.

Judith C. Stull, Ph. D., is an Associate Professor of Sociology at LaSalle University. Her teaching has been concentrated on statistics and research methodology. For the past 15 years she has been involved in research on improving teaching and learning. She has been a PI and co-PI on federal and state funded education grants. She holds a Ph.D. from Boston College.

Susan Jansen Varnum, Ph. D., is a Professor of Chemistry at Temple University. She has published widely in scientific journals and has produced 13 Ph.D. students. An analytical chemist, Dr. Varnum has directed numerous federally and locally funded projects designed to improve the educational opportunities of diverse student populations. She holds a Ph.D. from the University of Missouri.



Tiffany Gilles received her B.S. degree in Chemistry in 2005 and her M.Ed. degree in Educational Psychology in 2009. She has worked with science and science education faculty in developing experimental protocols for formative assessment in university level science classes. In addition, she has participated in multiple curriculum design projects that include developing activities for chemistry teacher professional development and a city-wide after school program. She is interested in science instruction and has significant experience in laboratory research. She has two publications in the science literature and has numerous professional presentations in science and science education.

Joseph P. Ducette, Ph.D., is a professor in the Department of Psychological Studies in Education, Temple University. He holds a B.A. in Psychology from Wisconsin State University and a Ph.D. in Experimental Psychology from Cornell University. His research interests include: Attribution Theory, Diversity, Urban Education, Early Intervention, and Learning Disabilities.