AN ANALYSIS OF TECHNOLOGY-ENHANCED PEDAGOGY AND LEARNING: STUDENT RESPONSE SYSTEMS (CLICKERS) - TOOL OR TOY?

Justin Matus
Associate Professor
Wilkes University
84 West South Street
Wilkes-Barre, PA 18766
United States of America
E-mail: justin.matus@wilkes.edu, Telephone: (570) 408-4714.

Kristin Summa
Wilkes University
84 West South Street
Wilkes-Barre, PA 18766
United States of America

Robert Kuschke
University of Missouri – St. Louis
United States of America

Abstract

The purpose of this research was to investigate the effect of technology-enhanced pedagogy on student learning. Three teaching environments were examined including a traditional lecture method, an on-line environment and a student response system, more commonly called "clickers". A counterbalanced design using three intact classes of upper level college business students enrolled in a capstone business strategy course comprised the study population (n = 62). Student learning was measured via standardized textbook chapter tests. Statistical analysis was performed using SPSS. A 3 x 3 mixed factorial (repeated measures) ANOVA procedure did not detect any statistically significant differences between the three groups' learning, regardless of the teaching method.

Key Words: Technology-enhanced pedagogy, learning, clickers.

INTRODUCTION

The purpose of this research is to investigate the effect of student response systems (clickers) on student learning. These systems, also known variously as electronic voting systems (Kennedy & Cutts, 2005), classroom performance systems (Blackman et al., 2002) or computer assisted learning (Draper et al., 1996) are all generally variations of a classroom technology designed to encourage students to participate in class by recording their individual response (sometimes anonymously) electronically via a remote-control like device to an instructor posed question (usually displayed on a projector screen). The idea is that the instructor can then quickly collect data from an entire class, display the results of all the student responses, indicate to the students the correct answer and thereby provide immediate feedback, fill in any gaps in understanding and thereby increase learning. For a full discussion of the technology and instructor experiences, see Caldwell (2007). For this particular study, a counterbalanced quasi-experimental design using three different pedagogical approaches will be investigated. The approaches include: (1) traditional classroom lectures and traditional reviews of material taught, (2) traditional classroom lectures and reviews of material taught using student response systems (hereafter referred to as clickers) and (3) traditional classroom lectures and on-line practice quizzes of material taught.

As technology advances and becomes more prevalent in the environment, the careful and thoughtful integration of these technologies becomes more critical. There is a seemingly natural inclination for instructors and administrators to sometimes blindly embrace the latest gadget, especially against the backdrop of promised improvements to student learning. Often times, however, the actual results in the classroom fall short of the promises. It is against this backdrop that the researchers embarked upon an investigation into this seemingly popular technology – clickers.

A review of the literature has shown an increasing breadth and depth of research into classroom technology. Sample populations include primary, secondary and higher education settings as well as a variety of disciplines including math, science and nursing, to name just a few (Hatch et al. 2005; Brunce, VanderPlas, & Havanki 2006; Wood 2004; Kaleta & Joosten 2007). While it is difficult to generalize any one particular trend or substantive finding, there does appear to be a few emerging themes. Most notably is the theme that students appear to have a favorable disposition towards some technologies, including clickers and on-line environments, especially for quizzes and lower stakes activities (Caldwell 2007; Stuart et al. 2004; Draper et al. 1996;). Hicks et al. (2001) note the advantages to on-line teaching including that it is: "…fully contextualized; seamless for student; staff member engages in developmental processes; providing significant learning opportunities" (p. 149). Bunce et al. (2006) in their study comparing a clicker approach to an online environment found, "… that the proposed benefit of student response system was not realized in this study" (p. 493). Moreover, in this same study, the authors noted that the online environment offers availability outside the classroom that clickers do not. Quite simply students may review material at their leisure. However in the same Bunce et al. study (2006) the researchers did not detect any significant difference in outcomes (as measured by a standardized final exam) across four treatment groups including: (1) clickers; (2) on line; (3) clickers/online combined; (4) no clickers/no online.

At this point, the reason why students like or dislike a technology is the wrong question. The ***relationship*** of technology to individual learning is ***the fundamental question***. A student may enjoy finger painting but whether that activity in and of itself helps a student learn about spatial perspective is a different story. In a similar vein, many instructors have a favorable disposition towards technology (Draper & Brown, 2004; Brown, 2004) but here again, caution is the watchword. "Teachers at all educational levels, even those who use a computer for many tasks, still remain fundamentally uncertain about how to use technology to solve real, everyday classroom problems" (Coppola, cited in Edens 2006, p. 161). Stated more simply, teachers need to understand if, when and how a given technology will support a given learning objective.

RESEARCH DESIGN

Of paramount concern to the researchers in this study was a strong design, especially with a view to the experiment's internal validity (Campbell & Stanley 1963). As noted by Edens (2006), "Few of these studies, however, used a comparison group, and none placed statistical controls on the comparison group (Roschelle, Peneul, & Abrahamson, 2004 cited in Edens)". Edens (2006) goes on to say, "Empirical Evidence on the influence of particular pedagogical uses of the technology system on achievement thus is limited" (p. 163). Accordingly, the researchers in this study developed a counterbalanced design. As noted in Campbell & Stanley (1963), "Like all quasi-experiments, this one (the counterbalanced design) gains strength through the consistency of the internal replications of the experiment" (p. 52).

Another important research design consideration was the question of equivalency across the three intact groups. Issues include the general aptitude of the three classes as well as the general inclination of a student to participate or not. Of course, the preferred solution when forming equivalent groups is always random selection and random assignment, however this was not possible in a college environment were student scheduling requirements preclude any such arrangement. The solution therefore was to take a pre-test measure of all students along these two dimensions (aptitude, general inclination to participate). These two pre-tests will be discussed in detail below.

The study population consisted of 62 students enrolled in three sections of a college senior level capstone business strategy course. The course met for 15 weeks, three times a week. The course syllabus was designed to approximate three distinct and equal periods of time of approximately five weeks each (September, October, and November) to facilitate a counterbalanced design. Each of the three sections would be exposed to three distinct learning environments over each of the three periods. During each month students were exposed to the same text book material, the same lectures and the same practice review questions. At the end of each month, all sections were given an exam to assess their mastery of the material presented during the month. The three independent variables were: pedagogical approaches (traditional, clickers, and on-line); class (1, 2, 3); time (September, October, November). The dependent variable was the three exams. See Table 1.

Table 1: Schedule of pedagogical approaches

	September	**October**	**November**
Class A	Traditional $Exam_1$	Clickers $Exam_2$	On-line $Exam_3$
Class B	Clickers $Exam_1$	On-line $Exam_2$	Traditional $Exam_3$
Class C	On-line $Exam_1$	Traditional $Exam_2$	Clickers $Exam_3$

The traditional learning environment consisted of a traditional classroom lecture and traditional review of text book material including a power point review of practice quiz questions each week. During the review of quiz questions the instructor would reveal the question and a list of four or five distracters. Students would then be encouraged to respond orally with the correct answer. The clicker learning environment consisted of a traditional classroom lecture and student response systems (clickers). Students would be given a weekly review with the identical practice quiz questions as the traditional and on line environments, except instead of responding orally, the students would use their individual clickers. During the review of quiz questions the instructor would reveal the question and a list of four or five distracters. After revealing the question and distracters, the instructor would wait approximately one minute and then reveal via a screen projector the class results, allowing the students to see what percent of the class selected which of the possible answers. The instructor would finally inform the class as to which was the correct answer.

The third environment was on-line. Utilizing the Web-CT platform, students were exposed to traditional classroom lectures and on-line practice review quizzes. Once again, the quiz review questions were identical to the traditional and clicker environments. Students were able to access review questions and take on-line practice quizzes each week during the month the class was using the on-line learning environment. In this environment students were allowed to take the quiz up to three times and view their total overall score, although the correct answers to specific individual questions were never revealed. Therefore if a student took a quiz three times and scored 80% she would not know which specific questions were correct or incorrect. The researchers felt that it was necessary to guard against student's sharing the correct answers and therefore failing to make an earnest individual effort. A post hoc analysis revealed that the vast majority (95+%) of students ultimately scored 100% on these review quizzes.

Attitude towards class participation

In order to measure the student's attitudes towards participating in class the researchers administered a 14-item instrument which they had previously developed (Matus & Kuschke 2007). The purpose of this instrument is to measure a student's generalized tendency to participate in class. The instrument does not portend to predict a student's level of participation in a specific class. The instrument consists of 14 items measured on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree) and includes question such as, "I participate more in classes when I am comfortable with the subject". The 14 individual items comprise the final summative scores for each student hereafter described as the Class Participation Summary (CPS) scores. The higher a student's CPS score, the more likely it is that a student will participate in a class. The instrument has a reported reliability of .76 and the researchers feel it is a valid measure of the construct being measured. The CPS scores were used as a pre-test to establish if there were any significant differences across the three classes in terms of inclination to participate.

Business Field Exam

The Business Field Exam (BFE) is described by the non-profit testing organization ETS as follows: "The Major Field Test for the Bachelor's Degree in Business contains 120 multiple-choice questions designed to measure a student's subject knowledge and the ability to apply facts, concepts, theories and analytical methods. Some questions are grouped in sets and based on diagrams, charts and data tables. The questions represent a wide range of difficulty and cover depth and breadth in assessing students' achievement levels" (www.ets.org accessed 15 March 2010). The reliability and validity of the BFE has been well established and is widely used by universities and colleges. The BFE exam scores were used as a pre-test to establish if there were any significant differences across the three classes in terms of aptitude.

Assessing student achievement on classroom material presented

During the course of this study each class section was administered an exam for the material presented in class for that month. All three class sections received identical exams, although exam questions and distracters were scrambled within and across classes in order to preserve the integrity of the exam process. All students completed a total of three exams over the three month period. Exams consisted of standard test bank questions from the text book publisher and were assumed to be valid and reliable measures of student mastery of the concepts being evaluated. Each exam contained 30 questions.

Research Questions

The following research questions were examined:

Are there differences in class participation attitudes between groups?
Are there differences in standardized business field exam scores between groups?
What is the effect of technology-enhanced teaching approaches on student learning?

RESULTS AND ANALYSIS

The first research question, “Are there differences in class participation attitudes between groups?” was investigated by performing an ANOVA on the Class Participation Summary (CPS) scores. Table 2 contains mean response scores for the CPS score. The average CPS score was 48.46. A higher score indicates a greater inclination to participate in a class.

Table 2: Mean Scores of Class Participation Summary

	N	Mean	Std. Dev.
Class A	22	49.00	6.21
Class B	21	48.04	6.71
Class C	19	48.31	7.65
Totals	62	48.46	6.74

An ANOVA analysis of the CPS scores is displayed in Table 3 below. The analysis indicated an F statistic of .111 and *p* value = .895 meaning there were no statistically significant differences between the three groups. This result gives some confidence that the three groups of students were approximately equal along this important dimension, the inclination to participate, thereby further strengthening the research design. Had these three groups had statistically significant different scores between them, it would be much more difficult to attribute one pedagogy’s effectiveness to the treatment itself rather than the class’s own inclination to participate more.

Table 3: ANOVA of Class Participation Summary

	Sum of Sqs.	Df	Mean Sq.	F	Sig.
Between Groups	10.37	2	5.189	.11	.895
Within Groups.	2767.05	48.89	46.89		
Total	2777.43				

The second research question, “Are there differences in standardized business field exam scores between groups?” was investigated by performing an ANOVA. The Business Field Exam (BFE) mean scores are displayed in Table 4.

Table 4: Mean Scores of Business Field Exam

	N	Mean	Std. Dev.
Class A	22	151.22	12.77
Class B	21	152.61	12.03
Class C	19	152.25	10.38
Totals	62	152.01	11.66

Results of the ANOVA analysis are displayed in Table 5. The analysis indicated an F statistic of .080 and a *p* value = .923 meaning there were no statistically significant differences between groups. This result again gives some confidence that the three groups of students were approximately equal along this second important dimension, aptitude, thereby further strengthening the research design.

Table 5: ANOVA of Class Participation Summary

	Sum of Sqs.	Df	Mean Sq.	F	Sig.
Between Groups	22.84	2	11.242	.080	.923
Within Groups.	8264.50	59	140.076		
Total	8286.98	61			

The third, and perhaps most important research question, "What is the effect of technology-enhanced teaching approaches on student learning?" posed a more challenging proposition in terms of the appropriate analysis. As Table 6 illustrates, the nine cells raise a few questions. Should the analysis be between months/exams (Sep, Oct, Nov; Exam 1, Exam2, Exam3) or pedagogy (clickers, on-line, traditional) or a combination of both?

Table 6: Mean Exam Scores by Month/Pedagogy

	September	**October**	**November**
Class A n = 22	68.32 (T)	77.69 (C)	74.50 (O)
Class B n = 21	71.72 (C)	80.65 (O)	74.52 (T)
Class C n=19	76.47 (O)	78.12 (T)	71.85 (C)
Legend **C**= Clickers **O**=On-line **T**=Traditional			

We initially felt that a 3 x 3 mixed factorial (repeated measures) ANOVA using the three exam scores by teaching methods as the dependent variable was the appropriate approach wherein you would effectively ignore the element of time (Sep, Oct, Nov as well as the Exams; 1, 2, 3). This analysis resulted in a finding of a *p* value = .109 meaning there were no statistically significant differences between the three groups, i.e the three different teaching methods. We pondered this result for some time, but realized that by ignoring the time periods/exams we were not taking into account these effects. In other words, we essentially bundled the exam scores by pedagogy, but not by time/exam, and more importantly we were never comparing the three time/exams to one another. We therefore ran a second 3 x 3 mixed factorial ANOVA which did yield a *p* value = .001, a statistically significant difference between the three groups (i.e., there was a difference in exam scores between the three months). We should also note that multiple analyses leaves us open to the criticism of familywise errors, and we are accordingly quite cautious in any conclusions we make (Field, 2006).

This leaves the reader to now wonder, what does it all mean? We feel in simple terms that we did not detect a difference in the various teaching methods, however we did detect a difference in the exams given. Quite simply it appears that the exam in October was the easiest and the exam in September was the most difficult. We also know that the average scores for all three exams were lowest for Class A, 73.50 and highest for Class B, 75.63 (see Table 7 below).

Table 7: Mean Exam Scores by Class Section

	Mean Exam Score (Combined scores of exams 1,2,3)
Class A	73.50
Class B	75.63
Class C	75.48

This also points to a not insignificant other problem, that of interaction effects. While the elegance of a counterbalanced design is intriguing if not enticing, especially for the strength of its internal validity, this also points up the very real problem of analysis and interpretation. Accordingly we offer two final tables, 8 and 9, for the reader to consider, each containing the mean score by class, by time/exam and finally be pedagogy. We leave further interpretation to the reader.

Table 8: Mean Exam Scores by Month

	Mean Exam Score (Combined scores of class A, B, C)
September (Exam 1)	72.16
October (Exam 2)	78.82
November (Exam 3)	73.63

Table 9: Mean Exam Scores by Pedagogy

	Mean Exam Score (Combined scores by pedagogy, ignoring class and month)
Traditional	73.42
Clicker	73.88
On-Line	77.19

Discussion

This study sought to carefully examine what impact a clicker-enhanced classroom would have on student learning. The research design intended to control for threats to internal validity through a counterbalanced design using three intact classes of senior level college students enrolled in a business strategy capstone course. Pre-test measures of student aptitude and inclination to participate indicated no statistically significant differences between the three groups. All three groups of students were exposed to the same text book, lectures and review material. Each group of students were exposed to the three distinct learning environments during three five-week time periods: traditional; on-line; clicker enhanced. At the end of each of the three five-week periods students were administered an equivalent exam to measure their learning of the material previously taught. A total of three exams were given to all three groups of students. A 3x3 mixed factorial ANOVA indicated that there were no statistically significant differences in the learning outcomes across any of the three groups during either of three time periods, regardless of pedagogical approach.

A second 3x3 mixed factorial ANOVA ***did*** indicated that there were statistically significant differences between the three sets of exam scores. This second finding points up to the significant challenge in deconstructing what is happening in the experiment. It raises many more questions than it answers. For example, if we were to assume for a moment that the clicker technology is actually very effective in improving learning, but that in the exact moment in this experiment, the class with the lowest aptitude and/or the lowest inclination to participate was also at the same moment taking the most difficult exam covering the most difficult material, but were also in fact being taught via clickers, is it possible that the effectiveness is being masked? Indeed, if we look carefully at our data, we know that the class B (highest BFE scores) in fact took the easiest test (test 2/October) during a month in which they were using what appears to be the most effective pedagogy (on-line, based on the highest average scores between the three methods; see Tables above) which may then suggest that this interaction perhaps intensified the effectiveness of on-line pedagogy. Conversely, we can also see from our data that class A with the lowest BFE scores took the most difficult test (test 1/September) using what appears to be the least effective pedagogy, traditional (see Tables above), again thereby perhaps exaggerating the relative observed effectiveness of the pedagogy. Finally across the entire duration of the experiment we have a multitude of possible interactions including the difficulty of the test, the pedagogy, the aptitude, the inclination to participate, student motivation, student fatigue, instructor motivation and instructor fatigue, all working in concert to possibly mask the effectiveness of any of the three approaches,

CONCLUSIONS

This main purpose of this study examined the effect of technology-enhanced teaching approaches on student learning. This study did not detect any statistically significant differences in the level of learning in any of the three groups, regardless of the pedagogical approach.

The researchers feel however that this finding should be interpreted with caution for several important reasons. First, the sample size across the three groups is rather small (n_1 = 22, n_2 = 21, n_3 = 19), therefore the failure to detect differences is not surprising. Second, the duration of each pedagogical approach was rather brief, about 5 weeks for each approach. The researchers speculate that for any of the approaches to take hold, so to speak, may indeed require a longer duration, perhaps a semester or more. Third, the learning outcomes measure (three 30-item exams) may not have been sensitive enough to tease out differences in the pedagogy being evaluated. Finally, the impact of instructor/researcher bias cannot be ignored. An on-going concern of the researchers was fear of "cheating" one section out of an opportunity to learn the material and be graded fairly. To the extent the researchers labored to maintain a sterile research environment in the classroom, there is nevertheless no denying that subliminal and subconscious effects may exist.

In light of the above, the larger question remains, what did this study teach us, if anything? We feel that on the plus side the research design was powerful and went to great lengths to build in a high level of internal validity. On the negative side, as mentioned earlier, we feel that a larger sample size, a longer duration for the experiment, and a more sensitive outcomes measure, would all or in part work in concert towards perhaps teasing out the much nuanced relationships between learning and pedagogy. While there appear to be a great many perceived benefits to using clickers, we agree with Caldwell's comment, "The reviews of the literature, however, also agree that much of the research so far is not systematic enough to permit scientific conclusions about what causes the benefits (p. 13, 2007). Moreover, while we did not detect any benefits to clickers per se, we specifically do not dispute earlier findings of other researchers. Quite simply we can only conclude that more research is needed. Indeed, we leave it to the reader and the practitioner to decide if a clicker in the classroom is a tool or simply a toy.

Because of our own uncertainty the researchers plan on a second round of research incorporating the lessons learned from this first round. The researchers are especially intrigued by the methodology described in Bunce (2006) wherein various concepts are emphasized within a class period while using a given pedagogy, i.e. clickers. This tactic allows the researcher, for example, to teach the four stages of the product life cycle, but focus say on the first two stages with a clicker emphasis, while teaching the last two stages without a clicker emphasis. This type of reflexive design is especially useful when faced with a typical classroom environment where manipulating the treatments or establishing a control group is often times problematic. This research team looks forward to the continued examination of this intriguing technology.

REFERENCES

Blackman, M. S., Dooley, P., Kuchinski, B., & Chapman, D. (2002). It worked in a different way. *College Teaching, 50*(1), 27-28.

Brown, K. (2004). Technology: Building interaction. *Tech Trends, 48*(5), 36-38.

Bunce, D. M., VanderPlas, J. R., & Havanki, K.. L. (2006). Comparing effectiveness on student achievement of a student response system versus online Web-CR quizzes. *Journal of Chemical Education, 83*(3), 488-493.

Caldwell, J. E. (2007). Clickers in the large classroom: current research and best-practice tips. *CBE-Life Sciences Education, 6,* 9-20.

Campbell, D. T. & Stanley, J.C. (1963). Experimental Design and Quasi-Experimental Designs for Research. Houghton Mifflin Company, Boston.

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

Draper, S. W., Brown, M. I., Henderson, F. P., & McAteer, E. (1996). Intergrative evaluation: An emerging role for classroom studies of CAL. *Computers & Education, 26*(1-3), 17-

32.

Draper, S. W., Cargill, J., & Cutts, Q. (2002). Electronically enhanced classroom interaction. Aust. J. Educ. Technol. 18(1), 13-23.

Edens, K. M. (2006). The interaction of pedagogical approach, gender, self-regulation, and goal orientation using student response system technology. *Journal of Research on Technology in Education, 41*(2), 161-177.

Field, Andy (2006). Discovering Statistics Using SPSS, 2nd. Ed. Sage. London.

Hatch, J., Jensen, M., & Moore, R. (2005). Manna from heaven or "clickers" from hell. *Journal of College Science Teaching, 34*(7), 36-39.

Hicks, M., Reid, I., & George, R. (2001). Enhancing on-line teaching: Designing responsive learning environments. *The International Journal for Academic Development*, *6*(2), 143-151.

Kaleta, R., & Joosten, T. (2007). Student response systems: A University of Wisconsin System study of clickers. *EDUCAUSE Center for Applied Research, 2007*(10), 2-12.

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.

Matus, Justin and Kuschke, Robert (2007). *Determinants of class participation: A preliminary investigation.* Conference Proceedings, Northeastern Association of Business, Economics and Technology, 30th Annual Meeting, State College, PA.

Roschelle, J., Penuel, W.R., & Abrahamson,L. (2004) *Classroom response and communication stems: Research review and theory*. Paper presented at the 2004 Meeting of the American Educational Research Association, San Diego, CA.

Stuart, S. A. J., Brown, M. I., & Draper, S. W. (2004). Using electronic voting systems in logic lectures: One practitioner's application. *Journal of Computer Assisted Learning, 20*(2), 95-102.

Wood, W. B. (2004). Clickers: A teaching gimmick that works. *Developmental Cell, 7,* 796-798.