The Use of Clicker Technology to Evaluate Short- and Long-Term Concept Retention

Adam J. Czekanski
David-Michael P. Roux
Department of Geography and Environmental Engineering

United States Military Academy

Abstract

Electronic student response systems ("clickers") offer instructors an alternative way to quickly query a student population's grasp of lesson material. We examined the use of clickers with 3rd-year undergraduates to determine if the anonymity provided by the clickers influenced student responses and to investigate clicker effectiveness on long-term concept retention. These objectives were accomplished in an introductory environmental science course using multiple-choice questions focused on key concepts at the end of 12 selected lectures. Three class sections (n = 48) answered the questions with clickers, while three sections (n = 48) answered by hand raising. Long-term concept retention was evaluated by linking exam question performance to these questions. We hypothesized that the anonymity offered by the clickers allowed each student to respond without feeling indirect pressure from peers to select a specific answer. Approximately 74.0% of the clicker population and 86.9% of the hand-raising population answered correctly. This illustrated that a lack of anonymity can greatly skew results by allowing a few intelligent students to sway the majority and give the instructor a false impression of class comprehension. Three sections (n = 55) exposed to no review questions acted as the control group. Results from two course examinations show that 79.9% of the hand raising population, 74.9% of the clicker population, and 75.3% of the control population answered correctly on these paired questions. A single-factor (population) analysis of variance (ANOVA) revealed that the students did not perform significantly better on the selected exam questions regardless of population type ($F_{2, 48} = 0.6165$; $F_{critical} = 3.191$; $P = 0.5441$). Our study showed that, despite their lack of effectiveness on long term retention, clickers are a valuable asset for short term concept evaluation. They provide immediate accurate feedback of student comprehension that enables the instructor to modify lesson material accordingly.

Key words: clickers, retention, evaluation

Introduction

Accurately assessing student comprehension of material in the classroom has always been a challenge for educators. Methods historically used by instructors have included calling on a broad range of students to answer questions, having the entire class answer questions through visible means, or using volunteers[1]. Although these methods have merit in moving toward a more active classroom learning environment, they all fail to truly give the instructor an accurate picture of how well or how poorly all students have grasped recently taught concepts. This downfall is primarily due to issues with the sample size, namely that the small minority can inherently mislead the instructor into believing that the majority either understands or misunderstands the concept being questioned[1]. Without the use of periodic quizzes or deliberately advertised examinations, the instructor has historically lacked a means to accurately assess the proficiency of the entire class. Electronic student response systems, or "clickers," are interactive remote answering devices that offer instructors a means to gain this accurate real-time assessment of student comprehension, and students the ability to assess their own grasp of material[2]. In a society that

demands instantaneous gratification, the ability to obtain immediate student proficiency feedback and an assessment of their abilities, is almost a necessary tool in the classroom. Clickers are a vehicle for assessing student understanding of learning objectives, but their growing popularity in the classroom in recent years has also been linked to their role in improving student preparation for class and supporting active learning[3]. The effective use of clickers has proven to be an excellent means to not only capture the attention of students, but also to hold that attention and give them a stake in the material being covered. Students who commit to an answer tend to become invested in the question and develop an increased interest in the discussion surrounding that question[1]. The benefit that clickers have over other "crude" methods historically used to query classroom populations is their ability to provide a level of anonymity to the student in answering a question posed in public. They allow students to provide input without fear of embarrassment in front of their peers and without having to be concerned about being lost in discussion surrounding the topic[4]. Without this anonymity, many students are likely to "withhold ideas or comments due to evaluation apprehension and conformance pressure[5]." This concept is true for both large and small classroom environments, but is seemingly more relevant in the smaller classroom due to the more intimate, and therefore more threatening, setting. This work seeks to investigate the effectiveness of clickers to evaluate short-term concept retention in a small classroom environment, and to assess whether the anonymity they provide plays a significant role in student response to review-based questions. The linkage between frequent clicker use during in-class lesson reviews and performance on examinations, or the long-term concept retention, was also evaluated.

Methodology

The short- and long-term retention of critical course concepts taught in an introductory environmental science course designed for non-environmental science or engineering majors at the junior level was evaluated. To accomplish this, we designed a semester-long experiment in which we could quantitatively capture and evaluate student retention of concepts they recently learned and compare these results to examination scores. This study involved 96 students divided among six sections, each with an average size of approximately 16 students. The course was organized into four blocks of instruction (I through IV), each representing major course themes. Mid-term examinations were administered at the end of blocks I and III. We were each assigned three sections of students to teach over the course of the semester and selected ten (of 40) scheduled 55-minute lectures in which review questions were used to assess the students' grasp of course material presented during the lessons. Three review questions were added to each of the lesson plans and presented at the end of the designated lessons, each in a multiple-choice format with three or four answer choices. The review questions were based on the critical learning objectives for these lessons and designed to interact with the clickers using Turning Technologies TurningPoint® 2008 plug-in software for Microsoft® Office PowerPoint® 2003. Student responses to the clicker questions were recorded using a Turning Technologies TurningPoint ResponseCard® RF wireless response system (Figure 1) or by raising hands.

Figure 1. *Turning Technologies TurningPoint ResponseCard® RF wireless response system. The system includes a universal serial bus (USB) receiver and 25 response cards. The pen is shown is for size reference.*

Answering by using the clickers and raising hands were the methods used to capture student responses to the review questions. To assess the impact of response method on answer choice, we fixed the response method for each section within the lesson blocks. Treatment populations were defined by the aforementioned methods of response. The control group was selected to be a population of students (n = 55, three sections) for whom no review questions were presented during lectures and whom a different instructor taught. The control group did, however, receive a general review of material from their instructor at the beginning of each lesson. During block I, one instructor (n = 48, three sections) used raising hands as the method of recording student responses, while the second instructor (n = 48, three sections) used the clickers. In order to account for the difference in scholastic ability among the nine sections, we computed the average cumulative grade point average (CGPA) of the students under each instructor. Five lessons in block I included review questions. After the first mid-term examination, the instructors switched response methods for the five lessons in blocks II and III with embedded review questions. Short-term concept retention was measured by using the clicker software to compute the percentage of a section that selected an answer choice or by counting the number of hands raised for each answer choice. These percentages were compiled on a data table that allowed us to gauge quickly how well our students understood concepts presented to them during lectures and to determine if the anonymity provided by the clickers caused a notable difference in the percentage of each section that selected the correct answers. We concluded the short-term concept retention study by developing a column chart that compares the average review question scores for the clicker and raising hands populations. A two-sample student t-test was used to determine if the response method influenced the percentage of the clicker and raising hands populations that obtained the correct answer on the in-class review questions.

Long-term concept retention was evaluated by selecting mid-term examination questions that tested concepts similar to those presented in the review questions. The examinations selected for evaluation were the first and second mid-term examinations. Five multiple-choice and two discussion/problem solving questions were selected from the first mid-term examination. From the second mid-term examination, five multiple-choice and five discussion/problem solving questions were selected. Scores from a total of 17 mid-term examination questions were recorded. The scores were then tallied to determine the average percentage of each section that attained a correct answer (for multiple-choice questions) or the average percentage of the total score (for discussion and problem solving questions). The overall percentage correct for the examination questions related to the review questions were then calculated by averaging these scores for each population type (raising hands, clicker, and the control group). These percentages were then plotted on a column chart for the three populations. A single-factor (population) analysis of variance (ANOVA) was then performed on the 17 average scores to determine if the use of the embedded review questions led to different average test scores for the clicker and hand raising populations over the control group.

To further broaden the scope of our work, we elected to continue our study with a similar experiment the following academic semester. Two major changes to the design of our experiment were made to accomplish this goal. First, the course studied was an introductory environmental engineering course taught to non-environmental engineering majors. As a result, the composition of the statistical populations changed due to different student class rosters. Two instructors were each assigned three sections of the course to teach (n = 44 students per instructor). The instructors presented the review questions using the clickers as the response format for two sections and had their third section respond by raising hands. Having the instructors vary response formats among their own sections allowed for them to see first-hand variations in student response from one section to another. Additionally, having the instructors vary the response method may account for the difference in each instructor's teaching style and how well these teaching styles convey information to the students. Secondly, each instructor presented the review questions at the start of the class period instead of the end. This allowed for the

review questions to serve not only as a review of the previous lesson's material, but also as segue into the current lecture. Six lessons from the first block (ten lessons) of this course were selected for embedded review questions. The questions represented a variety of conceptual and problem-solving questions in a multiple-choice format. Similar to data collection from the previous semester, we recorded the percentage of each section responding to the answer choices from the review questions and compiled these results in both tabular and column chart formats. A two-sample student t-test on the data was used to determine if the percentage of students that obtained a correct answer was correlated to the response method. As of the time this report was written, we have not linked these in-class review questions to mid-term examination questions testing similar concepts nor have we computed new CGPAs for the treatment populations.

Results

The data for this study was organized into two categories based on the use of clickers for enhancing short- and long-term retention of course concepts. The short-term retention portion of the study sought to determine if the students were able to retain critical course concepts recently presented to them and to see if the condition of anonymity (as provided by the clickers) gave the instructors an accurate assessment of their students' level of comprehension. To account for the difference in scholastic ability between populations, we computed the average CGPA of the students under each instructor. Since the treatment populations were identified by response method, however, the average CGPA for the raising hands and clicker populations switched after block I (Table 1).

Table 1. *Average cumulative grade point averages (CGPA) for statistical populations in the introductory environmental science course organized by block of instruction. The control was defined as the population of students (n = 55) for whom no in-class review questions were given.*

	Cumulative Grade Point Average Block of Instruction	
Population	**I**	**II/III**
Raising Hands	2.633	2.916
Clicker	2.916	2.633
Control	2.673	2.673

The percentage of each treatment population that answered these review questions was computed. The population raising hands obtained a higher percentage of students obtaining the correct answer on the review questions. Approximately 86.9% of the population raising hands answered the review questions correctly, compared to 74.0% of the clicker population (Figure 2).

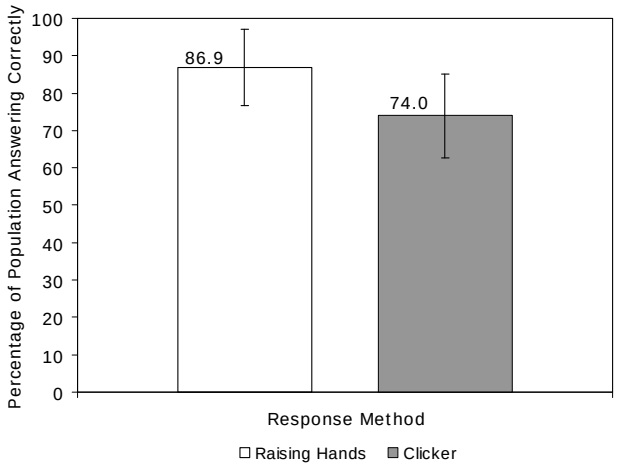

Figure 2. *Percentage of the raising hands and clicker populations that answered the review questions correctly for the introductory environmental science course. Data was averaged over ten lessons. The standard deviations from the mean for the raising hands population (n = 48) was 10.2% and 11.2% for the clicker population (n = 48).*

A two-sample student t-test was performed on the treatment population scores to test our hypothesis that the type of response method influences the percentage of students selecting the correct answer on the in-class questions. We obtained a $t_{statistic}$ of 2.934, which is greater than both the one- and two-tailed $t_{critical}$ of 0.008 and 0.017, respectively (Table 2). This indicates that, at a 95% confidence interval, we may statistically support the assumption that the method of response affects a population's average score on the in-class review questions.

Table 2. *Student t-test results for the raising hands and clicker population in-class review questions from ten lessons in the introductory environmental science course.*

Statistical Parameter	Raising Hands	Clicker
Mean Score (%)	86.9	74.0
Standard Deviation (%)	10.2	11.2
Observations (Lessons)	10	10
Degrees of Freedom (df)	9	
$t_{statistic}$	2.934	
$P(T \leq t)$ one-tail	0.008	
$t_{critical}$ one-tail	1.833	
$P(T \leq t)$ two-tail	0.017	
$t_{critical}$ two-tail	2.262	

We were also interested in determining if the in-class review questions enhanced our students' long-term retention of critical course concepts. Average scores from 17 mid-term examination questions were recorded for both treatment populations as well as a control group that was not subjected to in-class review questions. The average score for the population raising hands was 79.9%, 74.9% for clickers, and 75.3% for the control (Figure 3). Standard deviations from the mean were largest for the control (16.2%), 14.6% for clickers, and smallest for raising hands (13.1%). These statistics are graphically shown on Figure 3.

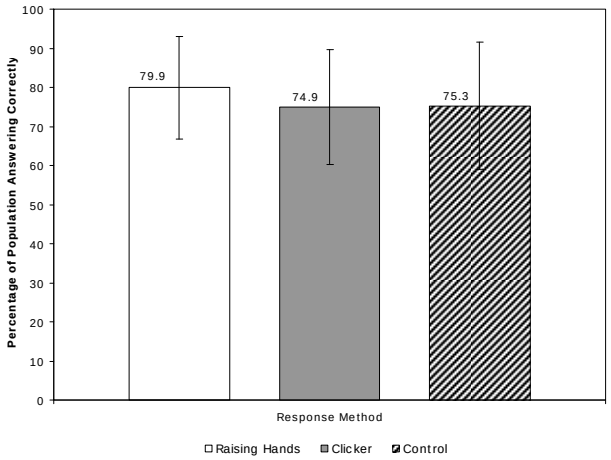

Figure 3. *Average examination scores (in percentage) for each population on 17 linked mid-term examination questions. Standard deviations from the mean were 13.1% for raising hands, 14.6% for the clickers, and 16.2% for the control population.*

In order to statistically compare the treatment populations with the control group, we performed a single-factor ANOVA on the average scores from the mid-term examination questions. We obtained an F-statistic of $F_{2, 48} = 0.6165$ (Table 3). Since this is significantly smaller than the $F_{critical}$ of 3.191 at a 95% confidence interval, no significant correlation could be drawn between population type and examination score.

Table 3. *Single-factor (population) ANOVA results on average examination scores for the raising hands, clicker, and control populations. Seventeen examination questions (observations) testing similar concepts to that of the review questions were used to calculate these values.*

Source of Variation	df	$F_{2, 48}$	P-value	$F_{critical}$
Between Populations	2	0.6165	0.5441	3.191
Within Populations	48			
Total	50			

We accounted for the difference in scholastic ability among the statistical populations by comparing the CGPAs listed on Table 1 with the average examination scores recorded for the mid-term examinations. This was accomplished by converting the average mid-term examination scores for each population (shown on Figure 3) into grade point averages on a 4.0 scale (Table 4). We found no clear trend between CGPA and examination performance on the 17 selected test questions, but we did note that the *method of response* among the treatment populations made a difference. In block I, the raising hands population (GPA = 2.633) achieved the highest GPA on the first mid-term examination (3.496). On the second mid-term examination (covering blocks II and III), however, the raising hands population (GPA = 2.916) earned the highest mid-term GPA (2.988). Thus, the raising hands population had the tendency to score highest on examination questions linked to the in-class review questions. Additionally, we found that the population raising hands earned an average mid-term GPA (for the 17 linked questions) of 0.863 points above their average CGPA of 2.633 on the first mid-term examination (Table 4). On the second mid-term examination, however, the control group earned the highest point difference (0.223 points) greater than their average CGPA (2.673).

Table 4. *Mid-term examination average scores on 17 selected questions related to the in-class review questions. The scores were converted to grade point averages on a 4.0 scale. The block(s) of instruction each mid-term examination covered is/are in parenthesis. The differences between the mid-term scores and the population CGPAs from Table 1 are also shown.*

Population	Mid-Term 1 GPA (Block I)	Difference from CGPA	Mid-Term 2 GPA (Blocks II/III)	Difference from CGPA
Raising Hands	3.496	**0.863**	2.988	**0.072**
Clicker	3.221	**0.305**	2.840	**0.207**
Control	3.176	**0.503**	2.896	**0.223**

Six lessons from the first block of instruction were selected for embedded review questions. The average percentage of the population answering the review questions correctly was 66.8% for raising hands and 59.3% for clickers (Figure 4), which follows a trend similar to what we encountered during the environmental science course. The more notable aspect of the results, however, is the standard deviation computed for these averages. The standard deviation from the mean for the population raising hands was 16.2%, which is greater than the 5.4% computed for the clickers (Figure 4).

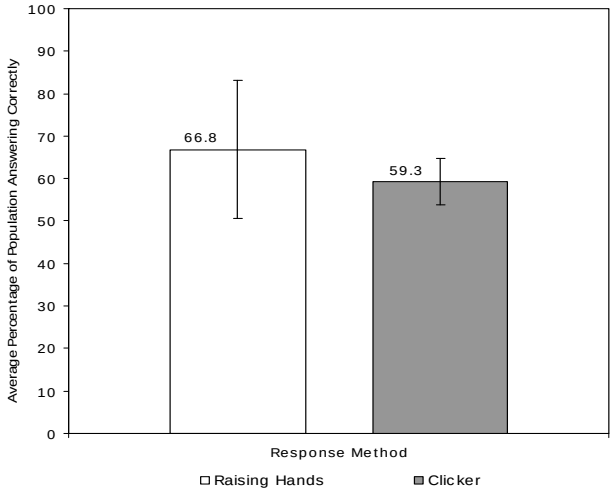

Figure 4. *Percent of the clicker and raising hands populations that answered the review questions correctly during the in-class review questions for the introductory environmental engineering course. Data was collected over 6 lessons. The standard deviations from the mean for the raising hands population was 16.2% and 5.4% for the clicker population.*

Performing a two-sample student t-test for the treatment populations resulted in a $t_{statistic}$ of 1.215, which is less than both one- and two-tailed $t_{critical}$ of 2.015 and 2.571, respectively (Table 5). This indicated that statistically there is not a significant difference in the percentage of the population attaining the correct answer on the review questions based on population type.

Table 5. *Student t-test results for the raising hands and clicker population in-class review questions. These results were generated from responses to review questions in the introductory environmental engineering course.*

Statistical Parameter	Raising Hands	Clicker
Mean Score (%)	66.8	59.3
Standard Deviation (%)	16.2	5.4
Observations (Lessons)	6	6
Degrees of Freedom (df)	5	
$t_{statistic}$	1.215	
$P(T \leq t)$ one-tail	0.139	
$t_{critical}$ one-tail	2.015	
$P(T \leq t)$ two-tail	0.279	
$t_{critical}$ two-tail	2.571	

Discussion and Conclusions

The results from the initial semester clearly illustrate a correlation between percentage of correct answers on lesson review questions and anonymity provided to the student. The provision of anonymity enabled the student to place self-doubt and fear of embarrassment aside in order to answer according to her own understanding[6, 7]. This condition produced results that were most often distributed in a manner expectant of a class containing students of varying abilities and levels of comprehension. In contrast, sections using the method of hand raising, and therefore affording the student no anonymity, most often experienced results inaccurately skewed in favor of the correct response. Although such results provide the instructor with the satisfaction of feeling as though he was successful in educating his students, the reality is that many of those students never really grasped the concept. Unfortunately, until the next quiz or exam, there was no way for the instructor to fully realize the inaccuracy of his perception of class comprehension. The clickers gave the instructor the valuable real-time picture of student comprehension of recently taught concepts. The critical piece was not necessarily this real-time knowledge of where the class stood, but rather the opportunity provided to the instructor to take necessary steps to correct student misconception about a concept. Most instructors will acknowledge that when more than 80% of students answer a question correctly, there is little need to review the details of the concept covered in that question very thoroughly. Correct answer success rates in the hand raising sections of nearly 87% and a standard deviation that regularly brought the correct response beyond the 90th percentile reinforce the point that the instructor not only is misled as to the comprehension of his students, but also that he inherently feels little or no need to talk through the concept as it relates to the correct answer. The accurate breakdown given by the clickers exposed the truth, as ugly as it was at times, and provided the "red flag" of reality that let the instructor know that maybe he should take a few extra minutes not only to reveal the correct answer to the class, but to refresh everyone on the details of the concept linked to the question[8]. The clicker section success rate of only 74.0%, even with a standard deviation of 11.2%, shows that the opportunity to give the class the review they really needed was afforded much more often when using the clickers than with raising hands.

The clicker as a tool for long-term concept retention was proven to be of negligible value as illustrated by student performance on mid-term examinations. This was not overly surprising due to the manner in which we integrated the technology in the classroom. The value of the clicker was recognized to be its ability to provide the instructor with accurate real-time assessment of the level of concept comprehension of the class shortly after the concept was taught. No matter what level of success the instructor achieved during that particular lesson in the conveyance of material, the knowledge gained is perishable without the proper amount of review prior to the examination. The material covered in review questions given most closely to examinations was still close to one week old by the day of the examination, with the

concepts grasped minutes after being taught having been lost in the daily grind of classes, homework and tests for a full load of classes. The lack of any significant difference in examination scores between sections demonstrated that examination performance does not necessarily correlate with the quality of the knowledge base established throughout the course of the semester (as measured by lesson review sessions). This is further reinforced by comparing the difference, increase or decrease, between average CGPA and examination GPA. This comparison further showed that the means of review had virtually no bearing on students' long-term retention of concepts; those sections conducting review through hand-raising saw the greatest success during the first mid-term, but the least success during the second examination. Long-term concept retention and comprehension is instead more of a product of the style and quality of which concepts were taught initially[1,4], but more importantly the direct product of the time invested by the individual student to study and fully understand the material immediately prior to the examination.

The clicker itself is neither a tool to directly teach concepts, nor is it meant to replace quality lesson preparation and planning. The clicker is a powerful tool to augment and enhance active learning in the classroom, and is most importantly a means to provide accurate situational awareness to the instructor[1,7]. Despite the lack of effectiveness in long-term concept retention, the value of the clicker in fulfilling the short-term assessment role was proven in its first semester of use. One weakness in the framework of initial research was that it was difficult to isolate the degree to which an instructor's particular teaching style or effectiveness would have on this short-term grasp of concepts. The method of having one instructor pose review questions to all of his sections using clickers, while the other instructor asked the same questions while receiving responses by raising hands, introduced the potential for results to be either reinforced or moderated depending on the teaching style in place. The modification of this method to enable both instructors to have sections using both methods during the second semester sought to control this variable. The fact that the average success rates for correct answers was statistically insignificant following this modification suggests two possible conclusions. The first is that maybe the anonymity provided by clickers really does *not* have a significant impact on how students will choose to answer questions. The second (and more likely reason) is that the content of the question may also weigh heavily into the students' perception of how important it is to answer correctly in front of peers. The questions given during the first semester were more *qualitative* in nature and focused on understanding concepts that often involved processes and systems familiar to everyday life. The questions during the second semester were much more *quantitative*, asking students to carry out calculations that most had just learned for the first time the previous class. The fact that most of the students acknowledge that the material is still relatively foreign to them, and that it is perceived as "difficult" due to its association to this engineering course (for non-engineers), provides a potential blanket of security that eliminates the fear of being labeled a failure in front of one's peers. The overall decrease in correct answer response using both methods is likely attributable to the shifting of the review questions from the end of class, immediately after exposure to material, to the beginning of the following class. The introduction of lag time between initial exposure to the material and review of that material clearly illustrates how quickly concepts can be discarded once the student is distracted with other course work and obligations.

In the future we will continue to assess the effects of both question content and anonymity on student response to review questions by continuing to use both means of response. Although question content may potentially affect student response in a way we did not initially foresee, the quantity of data collected for the quantitative questioning is only half that collected last semester, making it premature to definitively accept the correlation. We will also determine if shifting the review to the beginning of the following class, although producing a lower success rate in the short term, translates into greater overall concept retention during testing. Testing and grading for the second semester occurred after the deadline for submission of this publication and, therefore, could not be captured to make this assessment. The benefit may be that conducting the review at the beginning of the following lesson creates an entirely

separate event that reintroduces concepts already familiar to the students, as opposed to the review immediately after class that essentially becomes part of the same event introducing the material for the first time. We intend to build a more organized feedback mechanism into the end-of-course critiques in order to capture student evaluation of review techniques. This will give further insight into preference, acceptance and effectiveness of these techniques from the students' vantage point.

References

1. Caldwell, J. (2007). Clickers in the Large Classroom: Current Research and Best-Practice Tips. *Life Sciences Education*, 6, 9-20.

2. Petr, D. (2005). Experience with a Multiple-Choice Audience Response System in an Engineering Classroom. *Proceedings of the 35th ASEE/IEEE Frontiers in Education Conference*, 19-22 October 2005, Indianapolis, Indiana.

3 Auras, R., and Bix, L. (2007). WAKE UP! The Effectiveness of a Student Response System in Large Packaging Classes. *Packaging Technology and Science*, 20, 183-195.

4. Martyn, M. (2007). Clickers in the Classroom: An Active Learning Approach. *EDUCAUSE Quarterly*, 2, 71-74.

5. Lowry, P., Romano, N., and Guthrie, R. (2006). Explaining and Predicting Outcomes of Large Classrooms Using Audience Response Systems. *Proceedings of the 39th IEEE Hawaii International Conference on System Sciences*, 4-7 January 2006, Kauai, Hawaii.

6. Siau, K., Sheng, H., and Nah, F. (2006). Use of a Classroom Response System to Enhance Classroom Interactivity. *IEEE Transactions on Education*, 49(3), 398-403.

7. DeBourgh, G. (2007). Use of Classroom "Clickers" to Promote Acquisition of Advanced Reasoning Skills. *Nurse Education in Practice*, in press.

8. Siegel, J., Schmidt, K., Cone, J. (2004). INTICE – Interactive Technology to Improve the Classroom Experience. *Proceedings of the 2004 American Society for Engineering Education Annual Conference*, 20-23 June 2004, Salt Lake City, Utah.

Biographies

ADAM J. CZEKANSKI is an instructor in the Department of Geography and Environmental Engineering at the United States Military Academy at West Point, New York. He teaches introductory courses in environmental science, environmental engineering, and hydrogeology. Mr. Czekanski's academic and research interests include engineering education and drinking water treatment in developing nations.

DAVID-MICHAEL P. ROUX is an instructor in the Department of Geography and Environmental Engineering at the United States Military Academy at West Point, New York. He teaches introductory courses in environmental science, environmental engineering, and air pollution control. Mr. Roux's academic and research interests include engineering education and physicochemical/biological treatment in environmental engineering.