Testing a new voting machine question methodology

N. W. Reay,[a)] Pengfei Li, and Lei Bao
Department of Physics, The Ohio State University, Columbus, Ohio 43210

(Received 15 April 2007; accepted 9 November 2007)

A new question methodology has been developed and used with voting machines in large physics lecture classrooms. The methodology was tested by comparing student performance in voting machine and non-voting machine lecture sections during three consecutive electricity and magnetism quarters of introductory calculus-based physics. Data from The Conceptual Survey of Electricity and Magnetism and common examination questions indicates that students using voting machines achieved a significant gain in conceptual learning, and that voting machines reduced the gap between male and female student performances on tests. Surveys indicated that students were positive about the use of voting machines and believed that they helped them learn. The surveys also suggested that grading voting machines responses and/or overusing voting machines may lower student enthusiasm. © *2008 American Association of Physics Teachers.*
[DOI: 10.1119/1.2820392]

I. INTRODUCTION

Lectures are cost effective but not learning efficient, so educators continue to search for methods that enhance student participation in this traditionally passive environment. In-class electronic polling systems, also called clickers or voting machines, are being used to generate this participation at an increasing number of colleges and universities. These devices are even being offered by several publishing companies as an adjunct to their textbooks.

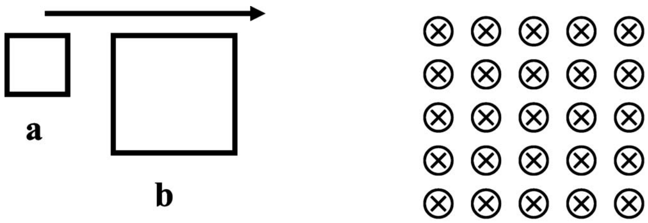
Continued voting machine hardware[1] and software development has eliminated most of the technical and economic barriers. Discussion of voting machine use can be found at The Ohio State University TELR and Vanderbilt websites.[2] Most voting machine companies also have websites.[3] Voting machines now provide an inexpensive, reliable, and easy-to-use interactive delivery system that is anonymous, in contrast to showing hands or raising numbered cards. Today's focus should be on developing and testing questioning methodologies that improve student learning without dramatically increasing the workloads of lecturers.

Existing materials and methodologies for using voting machines include Peer Instruction[4–6] and published articles.[7,8] There is increasing evidence that using voting machines engages students.[9,10] However, it also is important to focus on the effectiveness of using voting machines and to evaluate how specific methods in voting machine implementation impact student learning.

Voting machines generally have been used in a one question per concept format with a single set of surface features. Single questions, especially the excellent ones from Mazur[4–6] are available. In a few cases physics educators have given lectures entirely around the use of questions or sequences of questions that can be answered using voting machines, raising cards, or worksheets.[7,8,11] Electronic voting appears to encourage a higher percentage of participation and increased attendance, especially in larger classes. Using many short questions as intermediate building blocks toward developing a concept has been shown to be an effective approach, but may require an expanded set of instructional skills. However, even after a concept has been developed, students must recognize its range of applicability in a variety of situations. It is well recognized that learning is context dependent; students who have learned to apply a concept in one context might not be able to recognize the same concept in a different context, even when the two cases are considered to be isomorphic by experts.[12,13] Single questions provide limited assessment of whether students are able to make desired connections and transfer their understanding across contexts. Van Heuvelen[11] and others have used worksheets to present concepts in a variety of contexts. The use of voting machines provides a way to accomplish similar goals, with the added advantages of 90% participation and the ability to see instantaneous voting summaries.

A new methodology has been created that is based on using a sequence of questions, each displaying the same concept in a different context. This methodology is based on a constructivism paradigm widely used in active engagement curricula developed in physics education research,[14,15] but applies that paradigm within a much shorter time frame during lectures. This paper briefly reviews the question-sequence material reported in Ref. 10, and reports results from a quantitative study in which the question-sequence methodology was used during the electricity and magnetism (E & M) quarter of three successive year-long calculus-based introductory physics courses at The Ohio State University.

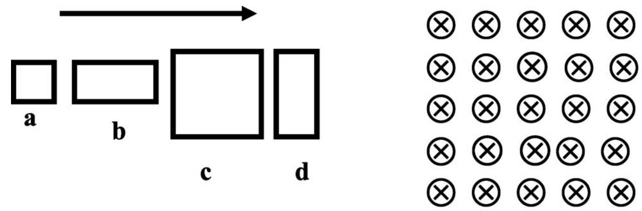
The main goal of the study was to determine if using question sequences with voting machines helps students learn, and if students perceive that this format has a positive effect on their learning. Concept inventory testing in eight previous E & M quarters of large calculus-based introductory physics lecture courses at Ohio State was remarkably stable: pre-scores, post scores, and gains were the same within errors for all quarters, even though the lecture sections were taught by many different instructors. A comparison was then made between lecture sections that used and did not use voting machines. Course content, homework, recitations and labs were kept the same for voting machine and control sections. During three quarters of testing, the voting machine sections consistently scored higher than non-voting machine sections on common examination questions and on post-quarter concept inventories. This difference was observed whether or not lecturers had previous voting machine experience. One instructor lectured in both voting machine and non-voting machine lecture sections. His voting machine section exhibited significant learning gains over his non-voting machine section. It also was determined that female students profited more from the use of voting machines than



The figure shows two square wire loops with edge lengths of L and 2L, respectively. Both loops will move into a region of uniform magnetic field B at the *same constant velocity*. Rank them according to the EMF induced ***just as their front edges enter the B field region***.

Answers: (1) a>b
(2) a=b,
(3) a<b
(4) depends on the magnitude of their common velocity
(5) depends on the B field magnitude

Fig. 1. Faraday's Law, question 1.



The figure shows four wire loops, with edge lengths of either L or 2L. All four loops will move into a region of uniform magnetic field B at the same constant velocity. Rank them according to the EMF induced ***just as they enter the B field region***.

Answers: (1) a<b<d<c
(2) a<b=d<c
(3) a<b<c<d
(4) a=b<c=d
(5) a=b<d<c

Fig. 2. Faraday's Law, question 2.

male students. End-of-quarter surveys indicated that students enjoyed using voting machines, and believed that this tool helped them learn.

II. REVIEW OF THE QUESTION SEQUENCE METHOD

As mentioned in the introduction, several researchers have suggested that learning is context dependent.[12,13] The OSU physics education research group hypothesized that context dependence could be more thoroughly addressed if each concept were presented in a sequence of questions with different context features, in contrast to the usual single question format. The latter approach also permits lecturers and students to assess in real time the level of achieved understanding while learning a single content topic.[10]

Question sequences. The design of each question is based on specific difficulties that students reveal during learning. A total of 45 multiple-choice question sequences containing over 140 individual questions were developed to cover major concepts in the E&M course, each sequence reflecting one or occasionally more of these concepts in a variety of contexts.

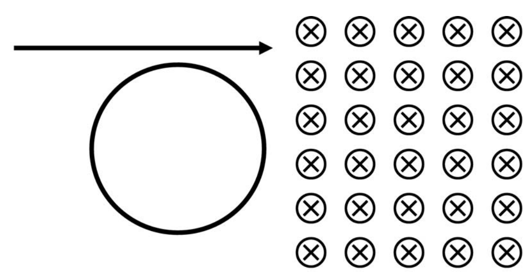
These question sequences can be divided into "easy-difficult-difficult" and "rapid-fire" types. (The difficulty of a question depends on both the population and the content.) Students viewing Faraday's law for the first time have difficulty differentiating between magnetic flux and the rate of change of that flux. As a result they may connect larger induced voltages to larger loops rather than to the rate of change of flux in a loop. The set of three questions in Figs. 1–3 was developed primarily to address this difficulty.

The first question (Fig. 1) is an easy question, especially because the largest loop also has the largest rate of flux change. In a class of 130 students with about 75% lecture attendance, 82% selected answer 3, which is correct.

The second question (Fig. 2) is more difficult. In the same class, 59% of students attending correctly selected 4, but 30% of students selected 2, which connects the emf generated directly to the total area of the loops. Students subsequently discussed why they had selected these answers. A revote was then taken without revealing the correct answer, and 89% of students correctly selected 4.

The third question (Fig. 3) assumes that if students really understand Faraday's law, they should be able to apply it to different loop shapes, even though the book and homework problems concentrated on rectangular loops. The students were given two chances to vote on this question, with some peer discussion in between. In the first vote, 62% of students correctly guessed answer 1, and 24% of students selected answer 5. Both answers reasonably characterized the emf changing as a function of time, but 5 introduced a sharp cutoff. More than 90% of students selected answer 1 after peer discussion with neighboring students. The correct answer and underlying physics were then revealed.

"Rapid-fire" question sequences usually contain questions that are at a more modest level of difficulty as judged by experts. Students are given less time to answer these questions than is given for the easy-difficult-difficult type question sequences. An example of a rapid-fire question sequence is shown in Fig. 4. This sequence was developed to give students practice in using the right-hand rule for forces on



A circular wire loop moving at constant velocity enters a long region of uniform magnetic field B. Which one of the graphs describes the emf ε in the loop as a function of time t?

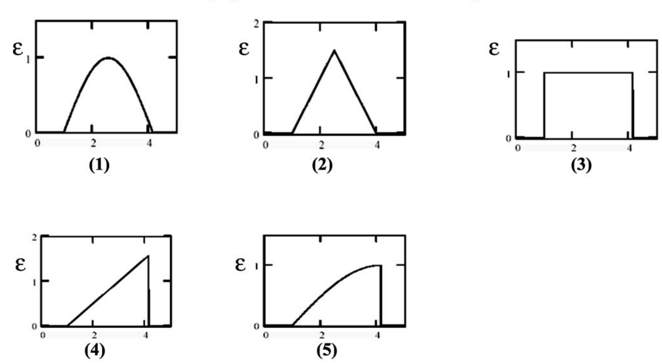


Fig. 3. Faraday's Law, question 3.

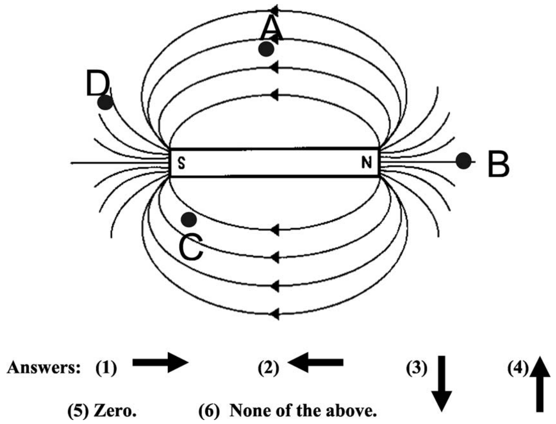


Fig. 4. Forces on charged particles in a magnetic field.

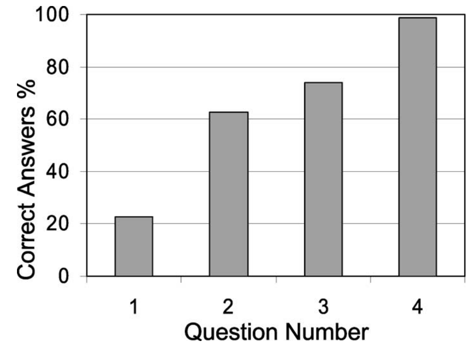


Fig. 5. Percentage of correct answers versus question number for the rapid-fire sequence.

charged particles moving in a magnetic field. All four questions are condensed into a single figure for this paper, but were presented to students in a separate slide for each question. The figure and answers for each question were exactly the same. There is only one particle in each question, and the location of the particle in each question was identified by A, B, C, and D, respectively. Students were given 40 seconds to answer each question, and the discussion that followed was brief. The texts of the four questions and students' responses are given in the following:

Question 1: A permanent magnet has field lines as shown above (see in Fig. 4). An electron moves out of the paper toward you at point A. The magnetic force on the electron is best represented by: (see the choices in Fig. 4)

This question sequence was given just after students had heard the right-hand rule discussed, and was the first time that they practiced it themselves.

On this first question, only 23% of students correctly selected answer 4 (choice 4 in Fig. 4), and an additional 35% of students selected answer 3, which ignored the electron's negative charge.

Question 2. A proton moves to the right at point B. The magnetic force on the proton is best represented by (see the choices in Fig. 4):

On this question, 63% of the students correctly selected answer 5 (see Fig. 4).

Question 3. An electron moves vertically upward at point C. The magnetic force on the electron is best represented by (see choices in Fig. 4):

On this question, 74% correctly selected answer 6, even though students generally hesitate to select "none of the above."

Question 4. A proton is at rest at point D. The magnetic force on the proton is best represented by (see the choices in Fig. 4):

On this question, almost 99% of students answered correctly.

The voting pattern on successive questions, as shown in Fig. 5, is typical for rapid-fire question sequences and is a desired effect. The goal is to have students improve by practicing skills with slightly changing context variables. Of course, such understanding in part may have been just learning how to answer questions, which is difficult to disentangle in practice and is beyond the scope of this paper.

III. A CONTROLLED QUANTITATIVE STUDY: RESEARCH DESIGN

The primary research question is whether using voting machines with the new methodology during a small percentage[16] of otherwise traditional lecture class time improves student conceptual performance. Answering this question was accomplished using a variety of tests that compared the voting machine lecture section to a lecture section with a similar population of students taught in a traditional manner without voting machines. The course content, homework, and labs were the same for both sections. Question sequences usually were shown once and only once during voting machine-section lectures, and after being used were not made available for reviewing by students. Students in the voting machine class were encouraged to discuss questions with each other while voting. Occasionally, students held additional discussions with each other and with the lecturer after viewing a voting summary, and then a re-vote was taken before revealing the correct answer. The latter approach is quite similar to Peer Instruction.[4–6]

A secondary goal of our study was to answer affective questions such as whether students enjoyed using voting machines, and whether they perceived that using voting machines helped them learn.

The primary independent variable in the research design was the use or not-use of voting machines in a given quarter. There are several contextual variables that were not controlled, such as population variations and gender issues. Other variables in the test environment were controlled as much as possible.

A three-quarter test (fall, winter, and spring) was conducted in the year-long calculus-based introductory physics course at OSU. All three quarters of calculus-based introductory physics are taught in several lecture sections during each academic-year quarter. Thus far, a complete set of voting machine question sequences has been developed only for the E & M quarter, although sequences for the other two quarters are accumulating. There were two E & M lecture sections each in the fall, winter, and spring quarters. Based on years of historical pre-post testing data with the Conceptual Survey of Electricity and Magnetism (CSEM),[17] the average pre- and post-test scores of students in the two E & M sections showed no statistically significant variation. Therefore, the two lecture sections in a given quarter were treated as equivalent random samples from a single population.

During each quarter voting machines were implemented in one of the two lecture sections, using a total of approxi-

Table I. CSEM pre and post testing during the year-long voting machine test. The numbers listed are the average number of correctly answered questions on the 32-question concept inventory. For comparison, scores for pre/post CSEM tests that were given in the second week/last week laboratories for two previous non-clicker years (16 lecture sections) of the same course averaged 11.4/15.2.

Quarter	Voting machine section			Non-voting machine section		
	Pre test	Post test	Gain	Pre test	Post test	Gain
Fall	12.1 (±0.36)	17.9 (±0.57)	5.8 (±0.52)	11.2 (±0.43)	15.6 (±0.55)	4.5 (±0.41)
Winter	9.3 (±0.25)	15.8 (±0.35)	6.5 (±0.34)	9.3 (±0.29)	15.0 (±0.47)	5.7 (±0.42)
Spring	10.9 (±0.44)	19.7 (±0.64)	8.9 (±0.69)	8.9 (±0.34)	17.6 (±0.50)	8.7 (±0.51)

mately 40 question sequences (about 130 individual questions) that covered all major E & M topics. In the fall and winter quarters, voting and subsequent discussions occupied less than 20% of the total lecture time. Because material covered using voting machines also was presented in non-voting machines classes, the additional time required by voting machines was due to the mechanics of taking votes and presenting voting summaries. The fall and winter lecturers agreed that this process took less than 10% of lecture time. In contrast, almost 50% of the total lecture time was used for voting machines in the spring quarter, because the lecturer added a considerable number of his own questions in addition to question sequences designed for this study. Non-voting machine lecturers had access to question sequence material, but otherwise taught in a traditional manner.

Ensuring that voting machine and non-voting machine lecture sections devoted the same time to each concept was difficult to manage. In the fall quarter it was controlled both by class observation and by weekly discussions between the two lecturers. The lecturers in the winter and spring quarters used the same book and essentially the same syllabus, but were less tightly controlled. All lecture sections were graded separately on a curve with mean grades ranging from C+ to B−. Recitation teaching methodologies were the same for all lecture sections, the homework was identical, and students from the two lecture sections were in common laboratories.

Measures used to evaluate the impact of the treatment with voting machines include student performance on pre-post CSEM testing and common conceptual multiple-choice questions on midterms and final examinations in both classes. Students in the voting machine lecture sections also took an end-of-quarter voting machine survey addressing affective issues regarding the second research question.

The timing of the exams and incentives offered for taking them varied between quarters. The manner of administering conceptual tests can sometimes impact test results, depending on the content of the test and the course structure. There are no documented results showing how varying the process of administering the tests might impact CSEM pre-post test results. Evidence of such impact will be discussed in later sections.

IV. DATA ANALYSIS AND DISCUSSION

Students in the E&M quarter of calculus-based introductory physics were given pre and post tests using the CSEM concept inventory. It was found that on average, students viewing and discussing voting machine question sequences once and usually only once during lectures achieved higher post-test scores than student in non-voting machine lecture sections. The timing of the pre test and the incentives offered for taking the post test appeared to affect the results. In the winter quarter, both sets of students scored equally on pre tests given prior to instruction. In the fall and spring quarters, students in voting machine classes scored significantly higher than students in non-voting machine classes on pre tests given after the start of instruction. The higher scores were consistent with better performance on questions containing material that had already been presented in lectures. Voting machine students scored higher than the other set of students in all three quarters on post tests. However, the percentage of students taking the post test and the average class scores varied significantly in both types of classes, which might have been due to differing incentives offered for taking the post test. The differences between post and pre test were the same within our statistics for male and female students in voting machine classes. In contrast, gains for female students were significantly lower than male students in non-voting machine classes. Students in voting machine classes scored more highly than those in non-voting machine classes on common conceptual questions given on two midterm examinations and the final examination.

Post-quarter surveys indicated that students in all three quarters enjoyed using voting machines and believed that their use had a positive effect on their learning process. The responses in the spring were less positive than in the fall and winter quarters. This difference might have been due to the fact that voting machines were used less than 20% of the time in the first two quarters and responses were not graded. In contrast, voting machines were used for almost half of the class time in the spring and the responses were weakly graded.

A. Results for CSEM pre/post testing

Results for the CSEM pre and post tests and the fraction of students taking both tests are shown in Table I. The results are raw and show the correctly answered questions out of the 32 possible.

The interpretation of the data must be done carefully, because there have been non-trivial variations among individual implementations of voting machine methods by non-physics education research instructors, as well as variations of incentives and timing of delivering pre and post tests due to uncontrollable constraints common to real classroom settings. For example, delaying the pre test until just a few days after the first lecture can significantly impact pre-test results. These items will be discussed in Sec. IV B.

Another issue is low lecture attendance, which is typical at OSU and at other large universities. Low attendance introduces another uncertainty, as our treatment is solely in the lectures. Voting machine responses were not graded in the fall and winter quarters, so there were no score-based incentives for students to attend the lectures. In the spring quarter the voting machine class was at 8:30 am, so voting machine responses were weakly graded to enhance attendance. Based on sampling in several lectures, the average voting machine lecture section attendance was approximately 75%, 50%, and 70% in the fall, winter and spring quarters, respectively. Based on the same sampling, attendance in non-voting machine sections was 10 to 15% lower. The impact on student learning due to attending lectures is not addressed and was not controlled. However, studies have shown that students attending traditional lectures exhibit smaller conceptual gains than students attending classes with interactive engagement.[18] The concern in this study is that a significant percentage of subjects in the voting machine group didn't receive the intended treatment.

There have been some issues in controlling variables in this test. Such constraints are often inevitable in field testing under real education settings. The results from our study are presented to help researchers and instructors understand how voting machine based methods perform in real teaching environments.

B. Pre-post testing data analysis

CSEM pre testing was performed in the fall, winter, spring, and summer during the second-week laboratories for two years prior to the voting machine test. Post testing for the same classes was performed in labs during the final week of the same classes. All the students used the same book, the same recitation style, did the same laboratories, and had the same homework delivery system. The 32-question CSEM pre and post tests averaged 11.4 and 15.2 questions answered correctly. The variance on pre tests, post tests, and gains for eight quarters of data was statistically insignificant. Performing an analysis of variance test gave $p=0.697$. Demonstrating a statistical difference in the means of the groups requires a p value that is less than 0.05.

Table I gives the pre and post tests for the year of voting machine testing. The fall pre (post) testing was performed during the second-week (final week) laboratories, which is the same time frame as for the historical data. The pre (post) test scores of 11.2 (15.6) for the non-voting machine section were consistent with historical results, but the voting machine section scored slightly higher on the pretest. By doing a detailed item analysis of the pre test, it was found that the increase of 0.9 questions, though statistically insignificant, was consistent with being entirely on material that had been presented during the lectures in the first week before the pre test was given. The voting machine lecture section post-test score was 2.3 questions higher than that of the non-voting machine class. A two-tailed t test gave $p=0.005$, indicating that it is highly unlikely that the post tests came from distributions with the same mean. Approximately 76% of the students took both pre and post tests and attempted to answer at least 85% of all questions.

During the winter quarter the book was changed for both lecture sections, and homework was switched to an online delivery system. All other aspects of the course remained unchanged. Pre tests were administered in the first recitation, which by accident occurred before the first lecture and hence before the presentation of any course material. Both types of classes scored 9.3 on the pre test, a score that was approximately 2 questions lower than in the two previous years of testing. A two-tailed t test gave $p=5\times10^{-13}$, indicating that the two previous years of pre tests did not come from distributions with the same mean as a pre test given before any instruction takes place. The voting machine section scored 0.8 questions higher on the post test than the non-voting machine section on an examination administered in the last recitation section. The corresponding p-value was 0.131, so this difference is not statistically significant. The score may have been impacted by low lecture attendance, which averaged 50%.

In the spring the pre test was given the day after the first lecture. The non-voting machine section pre-test score of 8.9 was consistent with the winter-quarter pre tests, but the voting machine section score of 10.9 was 2.0 questions higher. This significant increase was mostly on material presented in the first voting machine lecture that, unknown to the lecturer, was strikingly similar to many questions on the CSEM inventory. The lecture content was determined from the lecturer's notes and from notes taken by project researchers, who observed the first few lectures. To give a numerical basis for the statement that the increase was mostly on presented material we divided the CSEM into the first 14 questions, which covered material presented in the first 25% of the quarter, and the remaining 18 questions, which covered material presented in the remaining 75% of the quarter. On the pre test, the voting machine class scored 1.6 questions higher than the non-voting machine class for the first 14 CSEM questions, but scored only 0.4 questions higher for the final 18 questions.

Our results indicate that pre tests should be given before any instruction takes place in order to avoid fluctuations due to the timing of the pre test.

The voting machine section scored 19.7 on the post test administered in the final recitation section, 2.1 questions higher than for the non-voting machine section. The corresponding two-tailed t test p value is 0.009, indicating that it is highly unlikely that the voting machine and non-voting machine post tests come from distributions with the same mean. Using the same division as for the pre test, the voting machine class scored only 0.4 questions better than the non-voting machine class over the first 14 CSEM questions, but 1.7 questions better for the remaining 18 questions. Seeing voting machine questions once and only once during lectures may not be sufficient to lock them in students' memories. Such a possibility will be investigated in future research.

Incentives played a major role in post testing. In two years of post testing prior to using voting machines, individual lecture section scores varied only by a standard deviation of 0.35 questions, which is not statistically significant. The two-year average post-test score also agreed with the non-voting machine class score during the fall quarter, when post tests were given in the historical manner. Even when the voting machine and non-voting machine sections were averaged, post-test scores still fluctuated with a standard deviation of 1.5 questions during the fall, winter, and spring quarters. This fluctuation may be due in part to the fact that incentives for taking the pre and post tests impact the percentage of students taking these tests. The different incentives were largely due to the departmental course structure change and instructor preferences.

Table II. The ratio of the number of students who took both pre- and post-CSEM tests and the total number of students who took the final exam.

Quarter	(Pre+post)/total Voting machines	(Pre+post)/total Non-voting machines
Fall	98/130 (75%)	98/126 (78%)
Winter	162/184 (88%)	174/193 (90%)
Spring	57/76 (75%)	95/157 (61%)

No incentives were offered in the fall quarter. Approximately 76% of the students took both the pre and post tests in labs, and answered at least 27 questions on the post test. In the winter, students were offered a small number of points for taking the post test without regard to their scores. The percentage of students taking both tests rose to an average of 89%, and CSEM post-test scores dropped for both sections. In the spring, students were told that their CSEM score, if sufficiently high, would be appropriately scaled and would replace their lowest quiz score. The number of students taking both pre and post tests dropped below 70%, but average scores increased significantly when compared to the fall quarter for both sections. The average total course score in the spring voting machine class was 576 out of 720, a grade of B, for students who took both pre and post tests, as compared to 448, a grade of D+, for students who missed at least one of these tests. There was a similar distribution for the non-voting machine class. Such a large grade shift is a strong indicator that conceptual post-test scores rose in the spring because a significant number of lesser-achieving students were not included.

Based on the fact that 17 traditional lecture sections tested under identical conditions were consistent with coming from the same sample, it seems reasonable that voting machine versus non-voting machine comparisons can be made on the basis of post-test scores alone. Averaged over a year, the voting machine classes scored approximately 1.3 questions higher on the CSEM post test. The corresponding two-tailed t test gave $p=0.001$. It is statistically likely that giving students the opportunity to see and vote on each voting machine sequence once and only once in a quarter had a positive impact on their level of understanding. (Table II shows the ratio of the number of students who took both pre- and post-CSEM tests and total number of students who took the final exam.)

C. Results for common examination questions

In addition to the pre-post testing method, which was biased by several constraints, we also used common exam questions in the research design as an alternative measure of the performance difference between the two classes. In a given quarter, an identical set of 16–20 multiple choice questions were given on the midterm and final examinations in both classes. The questions were similar and occasionally isomorphic to voting machine questions used in the class, and covered most topics in the course. Multiple choice questions of this kind are a regular part of examinations traditionally given in these courses, and students from both classes are familiar with this type of question. The results from these common exam questions are shown in Table III.

As shown in Table III, voting machine lecture sections scored higher than non-voting machine sections on relevant multiple-choice conceptual examination questions for all three quarters. To compare the difference between the two classes, we use a scaled difference, which is based on the concept of normalized gain as used by Hake to evaluate pre/post testing results.[18] The difference refers to the difference of average scores between the two classes. The scaled difference is the ratio between the raw score difference and the possible maximum score difference based on the score of the non-voting machine class. Scaled differences are used under the condition that the average score of a voting machine class is higher than that of the corresponding non-voting machine class. Our result is that using voting machines in a lecture-based class increases performance on related conceptual questions by 22%–26%.

D. Effect of gender

Gender difference in science learning has long been studied in physics and science education research (see, for example, a long set of references compiled by Mallow and Hake[19]). It is generally recognized that women tend to be less interactive and are more intimidated by scientific and mathematical topics. In this study we also collected gender information to explore how male and female students differ in reacting to the use of voting machines. Pre-post CSEM average score gains for male and female students are shown in Table IV along with the number of students and calculated standard errors.

For voting machine sections, men had an average gain that was only 0.5 questions more than women, which is not significant. In non-voting machine sections, the gain for men was 2.3 questions larger than for women, which is signifi-

Table III. Performance on questions related to major concepts that were used on tests in the voting machine and non-voting machine lecture sections. The scaled difference is given by $(100\%)(\text{raw difference})/[1-(\text{non}VM\text{ score})]$.

Common exams	Voting machines	Non-voting machines	Raw difference	Scaled difference
Fall quarter (20 questions)	$72\% \pm 1.5\%$	$64\% \pm 1.5\%$	8%	22%
Winter quarter (16 questions)	$68\% \pm 1.1\%$	$56\% \pm 0.9\%$	11%	26%
Spring quarter (19 questions)	$63\% \pm 2.4\%$	$52\% \pm 1.4\%$	11%	23%

Table IV. Pre/post CSEM score gains for women and men in both types of sections for each of the three quarters of the test. The number of male and female students participating is shown in parentheses for each lecture section. Average results weighted by the number of participating students are given in the bottom row. The standard error of the mean and the sample size is also given. The fact that male gains appear comparable in the fall and spring sections is due to the fact that pre-test scores were significantly higher in voting machine sections when pre tests were given after the start of instruction.

Quarter	Voting machines section		Non-voting machines section	
	Female	Male	Female	Male
Fall	5.9±1.4 (14)	5.8±0.6 (84)	3.5±0.9 (18)	4.5±0.4 (80)
Winter	5.3±0.8 (23)	6.6±0.4 (139)	3.6±1.0 (34)	6.2±0.5 (140)
Spring	9.1±1.2 (9)	8.8±0.8 (48)	6.7±1.2 (15)	9.1±0.6 (80)
Average	6.2±0.9 (46)	6.7±0.4 (271)	4.3±0.7 (67)	6.6±0.3 (300)

cant. This result is consistent with anecdotal evidence shared from other groups[20] and with reported results.[21] It has been suggested that women may feel more comfortable participating anonymously with voting machines. Research to identify more solid evidence testing this and other hypotheses is beyond the scope of this paper and will be pursued in future studies.

E. Attitude survey results

Students' self-reporting of preferences and attitudes has been used for many years as supplemental information for evaluating education innovations.[22] In this study students using voting machines were given an end-of-quarter attitude survey soliciting their views about using them. Twelve questions in the fall and 15 questions in the winter and spring were answered using a five-point Likert scale ranging from −2 (totally disagree) to +2 (totally agree). Results for four of these questions are shown in Table V. Each type of preference or attitude includes several similar questions worded differently and sometimes in both positive and negative tones. Results for the other questions are similar to the ones shown.

The results show significant variance of students' rating over different quarters, which suggests that the attitude survey did measure something that was varying systematically. Investigating the actual causes of such a variance is a project in itself. Here, we will only present a preliminary analysis of the possibilities. The fall voting machine section lecturer (Reay) was an experienced voting machine user, but the voting machine section lecturers in the winter and spring had no prior experience with voting machines. Inexperienced lecturers were observed by Reay for their first two lectures, but were observed only sporadically throughout the remainder of the quarter. Negative survey statements were added in the winter to address possible bias due to the positive question style.[23,24] The concern that students might rate voting machines more highly when all questions were positive did not show up in the winter and spring results.

Voting machines were used for approximately 20% of each lecture period in the fall and winter quarters; 9% of students in the fall and 5% of students in the winter answered that voting machines took too much time away from the lecture. In contrast, 0% of students in the fall and 9% of students in the winter thought that the discussion of the questions was inadequate. Positive responses in the fall and winter quarters indicate that students enjoy voting machines and believe that they profit from their use. The winter and spring results also indicate that lecturers with little or no previous experience can use voting machines with minimal training.

However, it is clear that students were less positive toward the use of voting machines in the spring than in the fall and winter quarters. The spring voting machine section started at 8:30 am. As mentioned, voting machine participation was lightly graded based on attendance. Also, the spring lecturer added his own questions, doubling both the number of questions and the percentage of lecture time used for voting machines. Typically, the fall and winter voting machine classes used 3–7 voting machine questions per class. In the spring the typical number of voting machine questions ranged from 5 to 13. In the end-of-quarter anonymous survey, there was

Table V. Sampling of voting machine attitude survey results.

Statement	Fall average	Winter average	Spring average
"I like using voting machines."	+1.79	+1.59	+0.83
"Voting machines helped me understand lectures better."	+1.72	+1.46	+0.64
"I would recommend using voting machines in all future introductory physics courses."	+1.77	+1.43	+0.52
"I will avoid classes using voting machines in the future."	Not used	−1.54	−0.73

no indication that students felt handicapped by the limited discussion of questions. However, 21% of students felt that voting machines took too much time, and an additional 13% of students complained about being graded. The lecturer also made his own survey, which students took using voting machines despite the fact that individual students could be identified by the lecturer. Students responded that voting machines were being used about twice as much as they would have preferred. Discussions with students during the fall quarter suggested that a few students were concerned that using voting machines would take time away from problem solving. This minor concern might have become a major one for the spring quarter. This evidence suggests that grading voting machines results and/or using them for almost 50% of lecture time may have contributed to a less positive student attitude.

V. SUMMARY AND CONCLUSIONS

Data from pre/post CSEM tests and common exam questions indicate that students using voting machines and discussing solutions with each other during voting achieved a small but significant gain in conceptual learning. Gender-specific results showed that using voting machines reduces the gap between male and female student performances on tests: female gains were similar to those of males in voting machine lectures but were significantly smaller than those of males in the non-voting machine lectures.

Surveys indicated that students are strongly positive regarding the use of voting machines and believe that they help them learn. Grading voting machine responses and/or overusing voting machines may lower student enthusiasm.

This study has also created a new set of research questions including the issue of gender difference and the optimization of question structures, numbers, and the optimal time to use voting machines in class. This study also is part of a longer term program, which will develop and test a complete course package of question sequences for an entire year of algebra and calculus-based introductory physics.

ACKNOWLEDGMENTS

The authors wish to thank The National Science Foundation under Grant No. DUE 0618128, and also The Ohio State University for providing funding in support of this research.

[a]Electronic mail: reay@mps.ohio-state.edu

[1]R. J. Dufresne, W. J. Gerace, W. J. Leonard, J. P. Mestre, and L. Wenk, "Classtalk: A classroom communication system for active learning," J. Comput. Higher Educ. **7**, 3–47 (1996).

[2]See ⟨www.telr.osu.edu/clickers/⟩ and ⟨www.vanderbilt.edu/cft/resources/teaching-resources/technology/crs-biblio.htm⟩ for discussion of voting machine use.

[3]See ⟨www.turningtechnologies.com⟩, ⟨www.iclickers.com⟩, and ⟨www.quizdom.com⟩ as examples of voting machine websites.

[4]E. Mazur, "Linking teaching with learning," in *Science Teaching Reconsidered: A Handbook, Committee on Undergraduate Science Education*, Committee on Undergraduate Science Education, National Research Council (National Academy Press, Washington DC, 1997), pp. 21–26.

[5]C. H. Crouch and E. Mazur, "Peer instruction: Ten years of experience and results," Am. J. Phys. **69**, 970–977 (2001).

[6]E. Mazur, *Peer Instruction: A User's Manual* (Prentice–Hall, Upper Saddle River, NJ, 1997).

[7]I. D. Beatty, W. J. Gerace, W. J. Leonard, and R. J. Dufresne, "Designing effective questions for classroom response system teaching," Am. J. Phys. **74**, 31–39 (2006).

[8]D. E. Meltzer and K. Manivannan, "Transforming the lecture-hall environment: The fully interactive physics lecture," Am. J. Phys. **70**, 639–654 (2002).

[9]S. W. Draper and M. I. Brown, "Increasing interactivity in lectures using an electronic voting system," J. Comput. Assisted Learning **20**, 81–94 (2004); S. A. J. Stuart, M. I. Brown, and S. W. Draper, "Using an electronic voting system in logic lectures: One practitioner's application, J. Comput. Assisted Learning **20**, 95–102 (2004).

[10]N. W. Reay, L. Bao, P. Li, R. Warnakulasooriya, and G. Baugh, "Toward the effective use of voting machines in physics lectures," Am. J. Phys. **73**, 554–558 (2005).

[11]A. Van Heuvelen, *ALPS Kit: Active Learning Problem Sheets, Mechanics* (Hayden-McNeil, Plymouth, MI, 1990).

[12]L. Bao and E. F. Redish, "Model analysis: Assessing the dynamics of student learning," Phys. Rev. ST Phys. Educ. Res. **2**, 010103-1–16 (2006).

[13]*Transfer of Learning: From a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age, Greenwich, CT, 2005).

[14]R. E. Yager, "The constructivist learning model," Sci. Teach. **58**, 52–57 (1991).

[15]D. I. Dykstra, Jr., C. F. Boyle, and I. A. Monarch, "Studying conceptual change in learning physics," Sci. Educ. **76**, 615–652 (1992).

[16]Most of the material covered in voting machine questioning was covered by lectures in the non-voting machine classroom. This statement was validated by observing both voting machine and non-voting machine lecture sections during the fall quarter. Because most of the same material was covered for an approximately equivalent amount of time, "small percentage" specifically relates to the mechanics of taking a vote, typically 10% of lecture time.

[17]D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, "Surveying students' conceptual knowledge of electricity and magnetism," Am. J. Phys. **69**, S12–S23 (2001).

[18]R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. **66**(1), 64–74 (1998).

[19]J. V. Mallow and R. R. Hake, "Gender issues in physics/science education (GIPSE)–Some annotated references," ⟨physics.indiana.edu/~hake/GIPSE-4b.pdf⟩.

[20]C. E. Wieman, private communication.

[21]⟨www.ncsu.edu/PER/SCALEUP/FailureRates.html⟩ studies failure rates of men, women and minorities in traditional versus SCALEUP classes.

[22]J. Armarego, "Deconstructing students' attitude to learning: A case study in IT education," in *Proceedings of the 2007 Computer Science and IT Education Conference*, the Republic of Mauritius, November 16–18, 2007, ⟨csited.org/2007/38ArmaCSITEd.pdf⟩.

[23]J. J. Ray, "Reviving the problem of acquiescent response bias," J. Social Psych. **121**, 81–96 (1983).

[24]D. Watson, "Correcting for acquiescent response bias in the absence of a balanced scale," J. Theor. Probab. **21**(1), 52–88 (1992).