Direct and Indirect Approaches to Increasing Conceptual Survey Gains

Patrick B. Kohl, Charles Pearl, and H. Vincent Kuo

Colorado School of Mines, Department of Physics, 1523 Illinois Street, Golden, CO 80401

Abstract. Conceptual surveys like the FCI and CSEM are common, and course reforms often have the goal of improving student gains on these surveys. There exist various approaches to improving said gains, and there is occasionally concern that such methods "teach to the test" excessively. To our knowledge, however, there has been little direct experimentation on whether teaching to the test, even intentionally, has the expected result. In this paper, we report on a simple two-semester experiment involving ~900 students where we tried two different approaches to improving CSEM gains in an introductory E&M class. In the first trial, we gave students many of the questions from the CSEM as Peer Instruction-style clicker questions in lecture. In the second, we redeveloped parts of our Studio physics curriculum to target CSEM concepts without replicating CSEM questions. Comparing the CSEM gains in the experimental sections to the previous year's sections, we find that the first trial resulted in significant (~0.20) shifts in normalized gains on the relevant questions, while the second trial resulted in minimal or no shifts.

Keywords: Conceptual Surveys, CSEM.
PACS: 01.40.Fk, 01.40.G-

INTRODUCTION

Research-based conceptual surveys such as the Force Concept Inventory (FCI), Force and Motion Conceptual Evaluation (FMCE), and Conceptual Survey of Electricity and Magnetism (CSEM)[1-3] are in widespread use. These tests provide a way of comparing courses at different institutions or within institutions even if students have differing backgrounds that lead to different pretest scores (but see [4] for a cautionary view).

Instructors naturally want to demonstrate higher gains on these surveys, and numerous papers have been written demonstrating the effect of various techniques on survey scores. There has been discussion on whether "teaching to the test" is a problem; that is, are our course materials similar enough to the conceptual surveys that the performance gains observed aren't meaningful?[5-7] Researchers have argued that student knowledge (or lack thereof) is so robust that this is of little concern. Wieman has a notable example involving a violin[8] where students are given the answer to a question as part of a demo and are asked the question again within the hour. About ten percent of the students answer correctly. Aside from this, however, there is little work in PER formally establishing whether teaching to the test, even intentionally, has a significant impact on conceptual survey results.

At the Colorado School of Mines, the introductory physics with calculus sequence is taken by every undergraduate student (900-1000 students per year). These courses are taught in a hybrid lecture/Studio mode, described elsewhere.[9] The first semester mechanics course (Physics 100) is evaluated in part using the FMCE, and the second semester E&M course (Physics 200) uses the Conceptual Survey of Electricity and Magnetism (CSEM). In this paper, we report on two different approaches to improving scores on the CSEM in Physics 200. In the first experimental semester, some of the in-lecture clicker questions we use are replaced by select questions from the CSEM. In the second experimental semester, the original set of clicker questions is restored, but two of the two-hour studio sessions are revised to better address the concepts on the CSEM. The study has two major goals: First, we hope to establish the magnitude of the effect when we intentionally teach to the CSEM by way of clicker questions. Second, we hope to compare this effect to a more indirect approach in which we change the curriculum without making direct reference to CSEM questions.

METHODS

A typical Physics 200 course at Mines has an enrollment between 400 and 500 students, distributed amongst several lecture (2 hours/wk) and Studio (4 hours/wk) sections. The CSEM is given in lecture on the first day of class and on the last day of class before the final exam. Students receive a small amount of extra credit for taking the CSEM post-test. Normalized gains of 0.35-0.40 have been typical for several semesters.

As preparation for the study, we grouped the CSEM questions into the eleven categories originally used by Maloney[3] (Table I; note that questions can

	Category	Questions		Category	Questions
I.	**Conductors & insulators**	***1, 2, 13***	VII.	**Magnetic force**	21, 22, 25, ***27, 31***
II.	**Coloumb's law**	***3, 4, 5***	VIII.	**B-field from currents**	23, 24, 26, ***28***
III.	**E-field superposition**	***6, 8, 9***	IX.	**B-field superposition**	23, ***28***
IV.	**Force from E-field**	***10, 11, 12, 15,*** 19, 20	X.	**Faraday's law**	***29, 30, 31,*** 32
V.	**Work, potential & field**	***11,*** 16, 17, 18, 19, 20	XI.	**Newton's 3rd law**	***4, 5, 7,*** 24
VI.	**Induced charge & field**	***13, 14***			

TABLE 1. Categorization of the questions on the CSEM, from [3]. Bolded, italicized questions are those given as in-lecture clicker questions in the current study.

appear in more than one category) and searched for CSEM categories on which Mines students have struggled. For this we reviewed previous years' exam scores on a problem-by-problem basis, previous years' CSEM normalized gains on a category-by-category basis, and a modified parameter we refer to as the C-factor. The C-factor is the normalized gain on a question (or collection of questions) multiplied by the average of the pre- and post-test scores on that question:

$$C = \left(\frac{post - pre}{100 - pre}\right)\left(\frac{post + pre}{2}\right)$$

Being a composite of normalized gains and absolute pre/post-test scores, the C-factor discriminates strongly between questions that students do well on and questions that students do poorly on, with very general standards for "well" and "poorly". More detailed use of the C-factor will be reserved for a longer paper. In short, analysis suggested that of the CSEM categories, students were having the most trouble with V (work and potential) and VII (magnetic force).

The experiment took place in two semesters. In the fall of 2009, we performed what we will refer to as the *direct trial* with 397 students completing both the pre- and post-CSEM. In the lecture portion of Physics 200, students engage in clicker questions on a regular basis. Twenty of these questions were replaced with questions from the CSEM. Some of the CSEM categories from Table I (I, II, and III, for example) were covered in their entirety. Others were covered partially (IV, VII) or hardly at all (V, VIII).

All of the selected CSEM questions were presented in lecture according to the Peer Instruction methodology,[10] with multiple votes if necessary to reach a consensus (defined here as >70% correct). CSEM questions (and clicker questions in general) do <u>not</u> show up later in the course on exams or homeworks; in other words, students had no special incentives to study the clicker questions (beyond simply learning the material). The course was otherwise extremely similar to that of fall 2008, allowing for a control during data analysis.

In the spring of 2010, we performed what we will refer to as the *indirect trial* with 335 students providing complete pre/post CSEM data. We restored the original set of clicker questions, and re-wrote the curriculum for two two-hour Studios to better target the concepts in CSEM categories V and VII. These new Studios featured a blend of problem solving and conceptual questions that supported each other, presented online via the LON-CAPA system. The first portions of the studios featured significant scaffolding for both numerical and conceptual questions, while the latter portions required that students work through them with very little support. This is an adaptation of the cognitive apprenticeship model[11, 12] used previously at Mines.[9] This new material replaced material that largely predated conversion of the course to Studio. The spring 2009 course was used as a control.

In a study like this, test security is an issue. Widespread public dissemination of CSEM questions could easily render the results invalid either now or in the future. Every semester (experimental or not), we check the final CSEM data for anomalies that could indicate that the exam is being circulated out of class, such as C students getting 25+ out of 32 questions correct. We have not seen anything suspect, even in the semesters of this experiment.

DATA

For brevity, we present only normalized gains on the CSEM (no pre- or post-test scores). In Figure 1, we see normalized gain <g> as a function of CSEM category for the direct trial and for the previous fall's class. Error bars are omitted for clarity; all are equal to or less than +/- 0.05. Categories I, II, III, IV, VI, X, and XI are those for which more than half of the questions in the category were given in lecture. To test the differences between the control and experimental groups for statistical significance, we applied two-tailed z-tests with a Bonferroni correction (11 categories, reducing the threshold p-value to 0.0045 from 0.05). Most of the categories covered substantially in lecture showed statistically

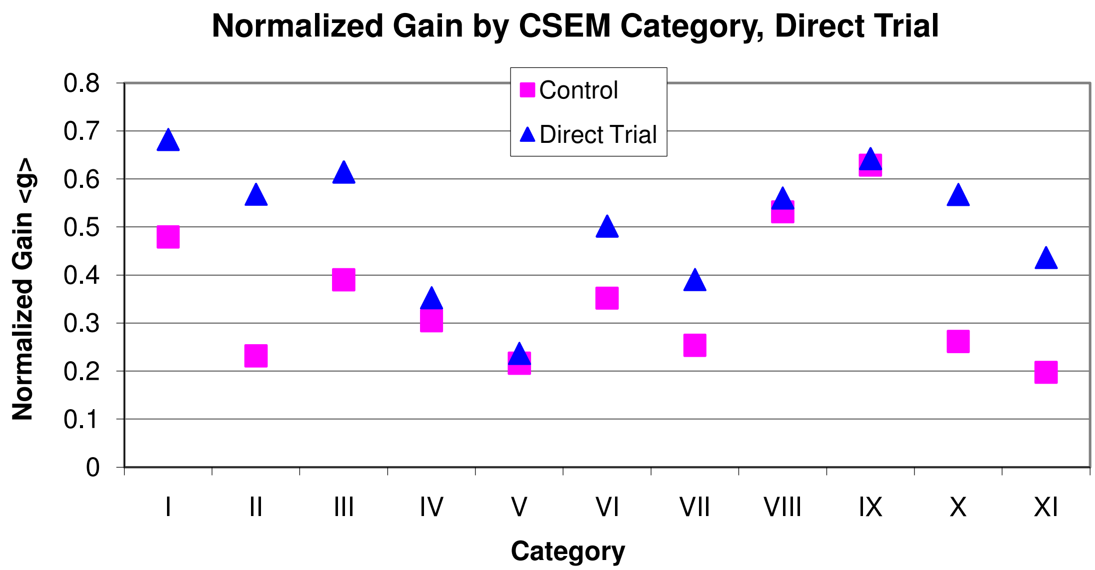

FIGURE 1. CSEM normalized gain scores separated by category, direct trial (CSEM questions in lecture). Categories V, VIII, and IX are the categories with the fewest questions used in lecture. Categories I, II, III and VI were entirely covered in lecture. Error bars are omitted for clarity; all are equal to or less than +/- 0.05. Categories I-III, X, and XI showed scores higher than the same semester in the previous year with a $p < 0.001$ (total N=691).

significant differences (I, II, III, X, and XI), with the experimental group performance higher ($p<0.001$). Category VI's difference was marginal ($p=0.046$). Categories (II and X) showed normalized gain increases that were especially high, in excess of 0.3. Of the remaining four categories (V, VII, VIII, IX) none showed a statistically significant difference between control and experiment after the Bonferroni correction. The overall CSEM <g> for this semester was 0.48, the highest recorded at Mines.

Figure 2 shows the same data from the indirect trial and the previous spring's class. Categories V and VII were those targeted by the new Studio activities. None of the eleven categories showed statistically significant differences after correction.

DISCUSSION AND CONCLUSIONS

We have presented two primary results: First, using CSEM items as clicker questions in lecture resulted in significant gains in student performance on those questions (and no others) as compared to previous semesters, with shifts sometimes in excess of 0.3. Second, revising Studio activities covering particular topics on the CSEM had little or no significant effect on student scores on those portions of the CSEM. The second result is somewhat disappointing, but perhaps understandable given the limited course changes made. The first result is more surprising. It has been established[8,13] that in some cases student performances do not change even if the exact question under consideration is given to the students. In this experiment, student CSEM scores responded significantly after a single in-lecture exposure to selected CSEM questions.

There is one important caveat: Without interview data, we cannot be sure *how* students used the available clicker questions. Lecture notes were posted online and were accessible to students in the course up until the final exam; it is possible that students studied clicker questions in preparation for the final. Anecdotally, the instructors involved have seen no indication of students studying clicker questions in past courses, but it is a possibility. These results may also be taken rather simply as further evidence of the effectiveness of the Peer Instruction methodology.

Regardless of mechanism, in-lecture exposure to CSEM questions had a very large impact on eventual CSEM gains. This is a noteworthy result for anyone interested in using research-based surveys to generate data that can be compared to data from other institutions. Some courses may well use clicker questions (or homework questions) based on FCI, FMCE, CSEM, or other questions under the assumption that such limited exposure to the questions won't affect final gains. The results of this study raise the possibility that this assumption might be erroneous. Using material directly from the

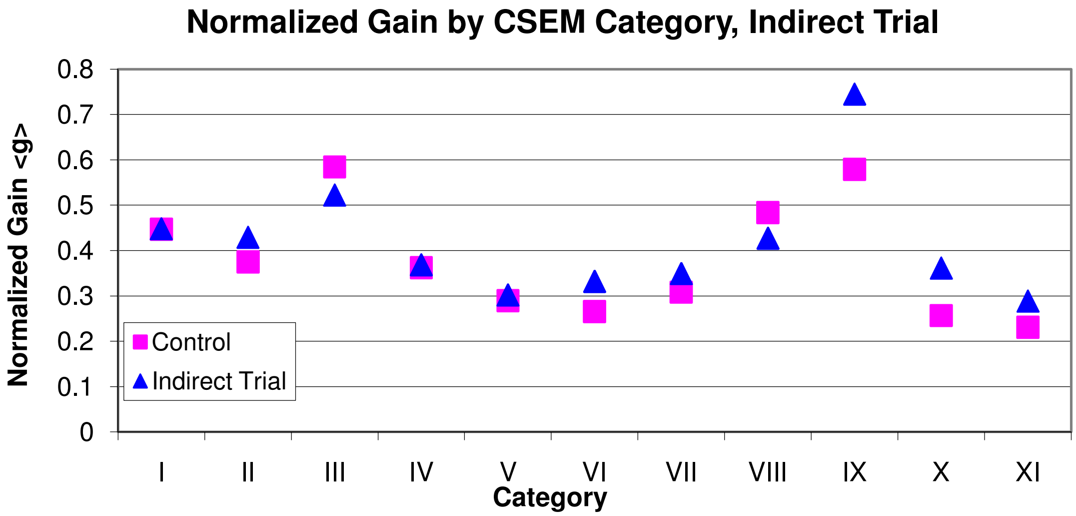

FIGURE 2. CSEM normalized gain scores separated by category, indirect trial (revised Studio sessions with conceptual scaffolding). All standard errors are equal to or less than +/- 0.05. No categories showed a statistically significant difference between the experimental and control years (total N=565) after applying a Bonferroni correction for multiple comparisons.

conceptual survey that assesses the course may well render the results less comparable to results from courses that don't use such material.

This experiment bears repeating, but repeating it at Mines in the near future may establish a pattern or norm that students would perceive, biasing the results. Thus we have no plans to repeat the direct trial anytime soon, and we encourage interested parties from other institutions to replicate or expand on this study. Curriculum reform in the style of the indirect trial continues at Mines. The conceptual scaffolding scheme was well-received by students, and we anticipate that with further development we will be able to increase CSEM scores meaningfully without drawing on CSEM questions directly.

ACKNOWLEDGEMENTS

Special thanks to Tom Furtak and Matt Young for their strong support of the undergraduate Senior Design program. Thanks also to E. Weisgerber and T. Lim for helpful discussions.

REFERENCES

1. D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, The Physics Teacher, **30**(3), 141-151, 1992.
2. R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The FMCE and the evaluation of active learning laboratory and lecture curricula, Am. J. Phys. **66** (4), 338-353, 1998.
3. D.P. Maloney, T.L. O'Kuma, C.J. Hieggelke, & A. Van Heuvelen, Surveying students' conceptual understanding of electricity and magnetism, Am. J. Phys., **69**(S1), S12-S23, 2001.
4. V. P. Coletta and J. A. Phillips, Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, Am. J. Phys., **73**(12), 1172-1182, 2005.
5. R. R. Hake, Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**(1), 66-74, 1998.
6. R. Ehrlich. Am. J. Phys. **70**(1), 24-29, 2002. See also the response by R.R. Hake, Am. J. Phys. **70**(10), 1058-1059, 2002.
7. A. Bell, in "Teaching Mathematics", M. Selinger ed. 41-46, 1994. Routledge, NY, NY.
8. C. Wieman and K. Perkins, Tranforming Physics Education, Physics Today **58**(11), 36-41.
9. P. Kohl, V. Kuo, and T. Ruskell, Documenting the conversion from traditional to Studio Physics formats at CSM: Process and early results, Proceedings of the 2008 PERC, 135-138, 2008.
10. E. Mazur, Peer Instruction: A Users Manual. Prentice-Hall, Upper Saddle River, NJ, 1997.
11. J. Lave and E. Wenger, Situated learning: Legitimate peripheral participation. 1991, New York, NY: Cambridge University Press.
12. J.S. Brown, A. Collins, and P. Duguid, Situated Cognition and the culture of learning, Educational Researcher **18**(1), 32-41, 1989.
13. Z. Hrepic, D. Zollman, and N. Rebello, Students' understanding and perceptions of the content of a lecture. Proceedings of the 2003 PERC. 189, 2004.