Classroom Response Systems in Statistics Courses

Michael B. Richman[1]

Teri J. Murphy[2]

Curtis C. McKnight[2]

Robert A. Terry[3]

[1]School of Meteorology, University of Oklahoma

[2]Department of Mathematics, University of Oklahoma

[3]Department of Psychology, University of Oklahoma

J4.4 Tuesday Jan. 22, 2008

Joint Session 4 Using technology to enhance learning: innovations and evidence (Part 2) (Joint between the 17th Symposium on Education and the 24th Conference on IIPS)

1. Introduction

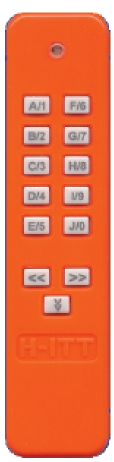
On a test of conceptual understanding, called the Statistics Concept Inventory (SCI), post-test averages have consistently been about 50% for undergraduate and graduate students (Stone, 2005; Stone et al., 2003 – a disappointing result for instructors. This result is not isolated: for some time, statistics educators have been aware that students lack data intuition and understanding of basic statistics concepts (e.g., Cobb, 1992).

In 2003, the American Statistical Association (ASA) funded the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project College Report* (Garfield et al., 2005). This report makes recommendations for the teaching of statistics. Examining the evolution of enrollment in statistics courses, the report notes that statistics courses now serve much larger numbers of students with a more diverse set of backgrounds, goals, interests, and attitudes. Courses are now offered in a wide variety of departments including business, economics, educational psychology, engineering, mathematics, psychology, sociology, and statistics. The content of statistics courses has also evolved in response to the availability of technology as well as advancements in statistics as a field of study. In turn, the teaching of statistics courses has been evolving. Building on recommendations put forth in Cobb (1992), the GAISE report recommends that statistics education (verbatim from p. 1 of Garfield et al., 2005):
1. Emphasize statistical literacy and develop statistical thinking;
2. Use real data;
3. Stress conceptual understanding rather than mere knowledge of procedures;
4. Foster active learning in the classroom;
5. Use technology for developing conceptual understanding and analyzing data;
6. Use assessments to improve and evaluate student learning.

Our courses at the University of Oklahoma already meet goals 2, 5, and 6; these aspects (e.g., other technologies, other assessment techniques) will continue to be an integral part of the courses, but are not the focus of this project. Toward *creating learning materials and teaching strategies*, we have used our local courses and our educational research expertise to develop annotations for a set of multiple choice questions (a.k.a., *items*) that develop statistical thinking (goal 1) and emphasize conceptual understanding (goal 3); these items will be for use with classroom response systems (explicitly mentioned with goal 5) to promote active learning (goal 4) and to provide frequent and immediate feedback to instructors and students about understanding. This project adapted work in physics education. Physics educators note that even students who succeeded on conventional problems do not necessarily understand the underlying concepts (e.g., Hestenes, Wells, & Swackhamer, 1992; Mazur, 1996). Several recent books describe instructional strategies that put this education research into teaching practice, advocating the use of instructional technologies, including classroom response systems, to incite active learning and to provide immediate and frequent feedback to students and instructors (e.g., Knight, 2002; Mazur, 1996; Novak et al., 1999). Compelling physics education research has provided discipline-specific evidence that the use of active learning methods has had positive impact on student understanding of basic physics concepts (e.g., Hake, 1998, 2002).

2. Our strategy for teaching statistics

In the influential physics education book, *Peer Instruction,* Mazur (1996) describes the use of conceptual questions to generate student-student discussion, which enables students to struggle actively with ideas. To facilitate the process, and to acquire immediate feedback about student understanding, instructors can use technology called *classroom response systems*. The technology consists of a few pieces of equipment: transmitters, a receiver, an LCD projector, and a computer. Each student in a class obtains a transmitter that resembles a TV remote control, colloquially referred to as a *clicker* (Figure 1). Each transmitter includes a set of buttons labeled, for example, A-E, and is tagged with a unique identifier that links to the course records of the student owner. The classroom itself is equipped with the receiver, LCD projector, and computer. To use the system during class, an instructor presents a multiple choice (MC) question. Each student indicates his/her answer by pressing the corresponding button on the clicker and the receiver acknowledges the signal. The answers are recorded on the computer. Publicly, responses are anonymous; however, the software stores (among other data) both the response and the particular clicker making that response. Once all answers have been received, the instructor has the option to display a bar graph indicating the set of student responses. This bar graph can inform the instructor's decisions about how to proceed with the course content. The instructor also has the option to award credit for responses. Examples of a clicker, a MC question for a statistics course, and a possible bar graph from the question are shown in Figure 1.



Which measure is most robust and best to use as summary measure if the data have outliers?

A. mean
B. median
C. mode
D. range
E. standard deviation

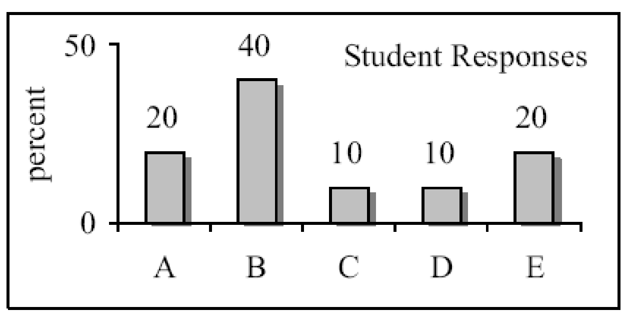


Figure 1. Example of a clicker, a MC question, and a bar graph of student responses to the item.

Based on a decade of experience in physics education, Crouch and Mazur (2001) assert that appropriate [clicker questions] are essential for success. They should be designed to give students a chance to explore important concepts, rather than testing cleverness or memory, and to expose common difficulties with the material. For this reason, incorrect answer choices should be plausible, and, when possible, based on typical student misunderstandings. A good way to write questions is by looking at students' exam or homework solutions from previous years to identify common misunderstandings, or by examining the literature on student difficulties (p. 974). Mazur (1996) notes that "composing these questions from scratch constitutes perhaps the largest effort required to convert from a conventional lecture presentation to a *Peer Instruction* format" (p. 26). This project will produce an annotated set of 250 MC questions for use in statistics courses. Consider for example the item in Figure 1; the annotation for this item might resemble Figure 2. (This figure is a condensed annotation with invented data and evidence, although the references are real. Actual annotations will also include the results of item characteristic curves, concept maps, and other results where appropriate and informative.)

In the test classes, when this question was used as a novice item, the distribution of responses tended to be 20-40-10-10-20. When it followed [another item about robustness], it generated productive student-student conversation. When it was used as an advanced item, about 90% answered correctly.
A. mean: Students may be attracted to this answer because mean is the most commonly used statistic. Students' tendency to confuse the mean and the median has been documented (Mokros & Russell, 1995; Watson & Moritz, 1999; Zawojewski & Shaughnessy, 2000a, 2000b). Also, this choice may reflect a lack of understanding that the mean depends on the magnitude of the actual numbers and is thus influenced by any large or small outlier.
B. median: correct answer
C. mode: Some students guess this answer. Others may reason that mode, which is the most common number, cannot be the outlier, which is unique (so students think), so the mode is robust. Guessing and this reasoning were identified in student focus groups.
D. range: Students do not understand that range depends only on two numbers – the extremes of the tails. This confusion was detected in task-based interviews.
E. standard deviation: Students do not understand that the standard deviation depends on the mean and on the size of the individual numbers and thus, like the mean, is influenced by any large or small outlier.

Possible instructional decisions: If students chose "C. mode", use a polymodal data set that has pairs of numbers plus a single outlier; compare that to the related data set that has an outlier as a pair. For the other responses, use a data set, first without then with the outlier, that illustrates the effect of one number whose magnitude differs considerably from the remaining data.
Figure 2: An example annotation for the MC question in Figure 1.

3. What classes we teach for this project

Since the project spans three departments, a variety of classes are taught using clickers. This gives us a wide variety of topics. Table 1 outlines the specific details.

Table 1: List of topics covered by target statistics courses at OU.

topics identified by ARTIST (condensed to conserve space)	**MATH 4753**	**MATH 4773**	**MATH 4793**	**METR 4303**	**PSY 2003**	**PSY 4023**
Data Types				x	x	x
Producing and Collecting Data				x	x	x
Representing Data	x	x	x	x	x	x
Measures of Center; Spread	x			x	x	x
Comparing Groups	x			x	x	x
Measures of Position	x			x	x	x
Normal Distribution	x	x	x	x	x	x
Bivariate Data: Quantitative	x			x	x	x
Linear Regression	x	x		x	x	x
Categorical (incl. Chi-Square)	x			x	x	x
Probability	x			x	x	x
Samples and Sampling	x			x	x	x
Tests of Significance: GENERAL	x			x	x	
1-Sample Means; Proportions	x			x	x	
2-Sample Means; Proportions	x			x		
Distributions for Test Statistics	x			x	x	x
Confidence Intervals: GENERAL	x			x	x	x
1-Sample Means; Proportions	x			x	x	x
2-Sample Means; Proportions	x			x		x
One-Way Analysis of Variance	x					
additional topics	**MATH 4753**	**MATH 4773**	**MATH 4793**	**METR 4303**	**PSY 2003**	**PSY 4023**
Experimental Design				x	x	x
Bayesian Interpretation of Probability				x	x	x
Forecast Evaluation Statistics				x	x	x
Random Number Generation				x	x	x
Central Limit Theorem	x			x	x	x
Graphical Summary Techniques	x			x	x	x
Pearson's Correlation	x			x	x	x
Cross Validation; Resampling Statistics	x			x	x	x
Distributions: Uniform; Binomial	x			x	x	x
Reexpression	x	x		x	x	x
Regression: Non-Linear		x			x	x
Multiple	x	x			x	x
Pitfalls of	x	x		x		
Outliers	x	x		x	x	x
ANCOVA			x		x	x

4. Writing the questions

Our group convenes a minimum of once a week to construct questions. The most challenging aspect is to conceive distractors that test student misconceptions. A crucial step in developing items for assessing misconceptions is the construction and subsequent evaluation of meaningful distractors. The process begins by applying theoretical and empirically substantiated findings from the literature to the initial item development phase. Once the initial phase (identifying and generating items) is completed and empirical data collected, the item response theory model can be used to exploit the identification of relevant concept information in incorrect choices to MC items, particularly when specific misconceptions are embedded within the choice set and can be used to infer how a student might be approaching a question. Moreover, for relevant distractors written in accordance with varying depths of understanding, distractor analysis can be used to guide instruction. For example, if only a few students select a specific distractor, we can then assume that it is of little pedagogical value given the current knowledge state in the classroom. To illustrate, we consider the item in Figure 1 and its invented annotation in Figure 2. We might hypothesize that the choice of *mode* as the answer indicates little to no knowledge of outlier effects on statistics, especially if an earlier novice item suggested that the student knew the meaning of mode. Since the mode is primarily used with discrete data, it would be a poor choice out of those listed if the data were continuous. Alternatively, a choice of *mean* or *standard deviation* (SD) as the answer would further indicate the student does not know that both the mean and SD depend on the actual magnitude of the numbers and therefore are influenced by any large or small outlier that differs in relative magnitude from the remaining data. Choice of *range*, although not involving the mean in any way, also depends solely upon the magnitude of two data points – the extremes. The *median* is the most robust of all. This requires that the student understand that the tails of a distribution most affect the lack of robustness of a statistical estimate to an outlier, regardless of whether the statistic is an estimate of location or spread. Any statistic that disregards the tails of a distribution is likely to be the most robust.

5. Unexpected outcome of weekly meetings

The interdisciplinary aspect of the project had an unexpected, but positive, outcome. Each of the instructors altered her or his teaching strategies based on the in-depth discussions we had. For instance, an interview with one of the mathematics professors elicited the following thoughts:

"The statistics class I teach is housed in a mathematics department. Most mathematics classes include a high level of absolute truth (one right answer even if multiple solutions, a proof or a counterexample, etc.). Working in an interdisciplinary team of faculty teaching statistics courses helped me to adjust to a number of aspects about statistics:

-- Whereas in mathematics the level of absolute truth is high, in statistics the level of judgment required is high.

-- Many mathematics problems can be applied to specific context, but can also be context-free (mathematics can be powerful specifically because of the level of abstraction), whereas in statistics the context for the data is critical for producing a valid, meaningful conclusion.

-- Different disciplines emphasize different aspects of statistics (e.g., an introductory course in psychology may emphasize sampling ideas, such as random assignment to treatment groups, that are not necessarily appropriate in science, such as meteorology, but a course in meteorology may emphasize a number of probability distributions that are not needed at the introductory level in psychology).

-- I was surprised to learn that there is not always agreement among statistics practitioners about the meaning of certain terms or the importance of certain topics or the, as evidenced by the number of student questions I brought to the group that generated discussion or the number of instances of questions that one person would write that another person would claim not to understand.

-- I would definitely have relied more on the textbook (which I don't like) if I had not been able to meet regularly with experienced statistics instructors. I learned more about the subtleties of statistics more quickly that I could possibly have if my primary resource had been the textbook."

6. How well did clickers work to enhance statistics education?

Working as a cohesive group has improved the delivery of material among each member. The SCI pre and post test is administered at the beginning and end of each class. These are a set of 38 multiple choice questions. The details are available at **https://engineering.purdue.edu/SCI**

Gains maybe underestimated because instructors don't cover every topic on the SCI and disagree about correct answers in some cases (therefore expect students to answer as instructor taught). The gain for fall 2007 class using clickers was: pre: $\bar{x}$ = 16.7, s = 4.4, post: $\bar{x}$ = 19.5, s = 5.5, t = 3.485, p = .002, (df = 26), five number summaries: pre: 8, 14, 16, 20, 27; post: 8, 16, 19, 23, 33. Similarly, the gain for a separate fall 2007 class using clickers was: pre: $\bar{x}$ = 18.6, s = 5.7, post: $\bar{x}$ = 23.9, s = 6.9, t = 4.002, p = .001, (df = 15), five number summaries: pre: 9, 16, 18, 23, 29; post: 12, 18, 25, 30, 32. While these results look very promising, a few caveats need to be considered. Confounded causal effect cannot attribute statistically significant gains to clicker use. Confounding variables include: instructor effect, use of small group work on non-clicker questions, use of real data and computer software (see GAISE 2005 recommendations).

Acknowledgements.

This material is based upon work supported by the National Science Foundation's Course, Curriculum and Laboratory Improvement (NSF CCLI) program in the Division of Undergraduate Education, under Grant No. 053589. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors also wish to acknowledge the extensive contributions by other members of the research team: Krista Hands, Matthew Schuelke, and the student participants.

References

ARTIST. Assessment Resource Tools for Improving Statistical Thinking. Accessed April 22, 2005, http://data.gen.umn.edu/artist/.

CAUSE. Consortium for the Advancement of Undergraduate Statistics Education. Accessed May 1, 2005, http://www.causeweb.org.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 295-323). New York: Springer.

Cobb, G. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the Call for Change: Suggestions for Curricular Action (MAA Notes No. 22)* (pp. 3-43). Washington, DC: The Mathematical Association of America.

Crouch, C. H., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics, 69*(9), 970-977.

Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., et al. (2005). Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project College Report. Accessed April 3, 2005, http://it.stlawu.edu/~rlock/gaise/.

Hake, R. R. (1998). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*, 64-74.

Hake, R. R. (2002). Lessons from the physics education reform effort. *Conservation Ecology, 5*(2). Accessed May 1, 2005, http://www.ecologyandsociety.org/vol5/iss2/art28/.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teachers, 30*, 141-158.

Knight, R. D. (2002). *Five Easy Lessons: Strategies for Successful Physics Teaching*. San Francisco, CA: Addison Wesley.

Mazur, E. (1996). *Peer Instruction: A User's Manual*. Upper Saddle River, NJ: Prentice Hall.

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20-39.

Novak, G., Gavrin, A., Christian, W., & Patterson, E. (1999). *Just-In-Time Teaching: Blending Active Learning with Web Technology*. Upper Saddle River, NJ: Prentice Hall.

Stone, A. (2005). *An Analysis of the Statistics Concept Inventory.* Unpublished Dissertation, The University of Oklahoma, Norman.

Stone, A., Allen, K., Rhoads, T. R., Murphy, T. J., Shehab, R., L., & Saha, C. (2003). *The Statistics Concept Inventory: A pilot study.* Paper presented at the Frontiers in Education 2003 Conference, Boulder, CO.

Watson, J. M., & Moritz, J. B. (1999). The development of concepts of average. *Focus on Learning Problems in Mathematics, 21*(4), 15-39.

Zawojewski, J. S., & Shaughnessy, J. M. (2000a). Data and chance. In E. A. Silver & P. A. Kenney (Eds.), *Results from the Seventh Mathematics Assessment of the National Assessment of Education Progress* (pp. 235-268). Reston, VA: National Council Teachers of Mathematics.

Zawojewski, J. S., & Shaughnessy, J. M. (2000b). Mean and median: Are they really so easy? *Mathematics Teaching in the Middle School, 5*(7), 436-440.