A Model of Classroom Assessment in Action: Using Assessment to Improve Student Learning and Statistical Reasoning

Beth Chance
University of the Pacific, Mathematics Department
3601 Pacific Ave.
Stockton, CA 95211
bchance@uop.edu

Joan Garfield
University of Minnesota, Department of Educational Psychology
332 Burton Hall, 178 Pillsbury Drive SE
Minneapolis, MN 55455
jbg@tc.umn.edu

Robert delMas
University of Minnesota, General College
128 Pleasant Street SE
Minneapolis, MN 55455
delma001@tc.umn.edu

1. Introduction

Recent years have seen a shift in the focus of introductory statistics courses, emphasizing skills such as the ability to interpret, evaluate, and apply statistical ideas rather than procedural calculations. Calls for reform also emphasize that instruction should fully incorporate genuine data, technological tools, and active learning (e.g. Cobb, 1992). As instructional goals change, so must the assessment techniques used to evaluate progress toward these goals, both for evaluation of student performance and for research studies on the effectiveness of new instructional techniques. Traditional assessment too often emphasizes the final answer over the process (Garfield, 1993) and may not reflect the desired outcomes of the statistics reform movement. Instead, assessment techniques must also measure how student learning is affected by new approaches. In this paper, we discuss an alternative conceptualization of assessment and its role in statistics education research. This alternative framework is based on a theory of conceptual change. We provide an example of a collaborative classroom study on the effectiveness of computer simulations in guiding student construction and visualization of sampling distribution behavior utilizing this framework.

II. Research Question

Researchers and educators have found that students and professionals often misunderstand foundational statistical ideas. Our students have especially struggled with the concept of sampling distributions, the gateway to statistical inference. Thus, we began to explore the effectiveness of a computer simulation program in teaching sampling distributions. The *Sampling Distributions* program, developed by delMas, allows students to interact with the concept visually, in a dynamic, interactive environment. Similar to other visualization programs (e.g. *ConStatS, Hyperstat, Visual Statistics, ExplorStat*), students change parameters and then run simulations in order to directly see the effects of these changes. To evaluate the impact of such interactions, traditional evaluations, such as summative course evaluations and final exams, are limited. For example, research suggests that students who earn good grades on final exams often still demonstrate poor statistical reasoning skills (Hawkins, Joliffe, & Glickman, 1992). These methods also fail to exploit the diagnostic abilities of these dynamic tools or to inform the instructor of the direct impact of the software interactions or their role in the student's learning process. In particular, traditional assessment strategies do not tell us enough as to <u>why</u> a particular teaching method or activity works, <u>how</u> students' understanding and reasoning are affected or unaffected by the learning experience, or how a teaching practice should be changed.

III. New Assessment Strategies

The University of the Pacific and the University of Minnesota have been incorporating the *Sampling Distributions* program and an activity to guide student interaction with the program into their introductory courses for several years. Much of the initial development was guided by recommendations from the literature on conceptually enhanced simulations (e.g. Nickerson, 1995; Snir, Smith & Grosslight, 1995). As we continually refined the activity and the program, we explored alternative types of assessment to help us better understand what students were learning, and how their conceptions or misconceptions were affected by the activity. Recently we have also been exploring how prior knowledge affects students' experience with the learning activity.

To directly measure student gains from interaction with the sampling distribution activity, we developed graphics-based test items. Our intention was to go beyond traditional questions to see if students could demonstrate a visual understanding of the Central Limit Theorem's implications for sampling distributions. Students were given characteristics for five different populations and asked to choose which picture best represents the resulting sampling distribution for various sample sizes. One version of the questions was given as a pre-test prior to the activity (but after the concepts were introduced in lecture) and another version was given as a post-test after completion of the activity. Students were also asked to provide their own explanation for their choice. Such pre- and post-tests allowed us to directly measure the additional knowledge the program and activity had given the students. Specifically, we saw that while students showed a significant positive change from pre-test to post-test, a substantial number of students still did not appear to understand some of the basic properties of sampling distributions. Well-designed software and guiding directions do not ensure sufficient student engagement or understanding.

These evaluations led to alterations in the software and the accompanying activity. The main adjustment, following a model of conceptual change presented by Posner, Strike, Hewson, and Gertzog, 1982), was to use the pre-test to guide student interaction with the software. Research indicates that people are generally resistant to change and are likely to find ways to either assimilate information or discredit contradictory evidence rather than restructure their thinking in order to accommodate the contradictions (Lord, Ross, & Lepper, 1979; Jennings, Amabile, & Ross, 1982; Ross & Anderson, 1982). Modern information processing theories (e.g. Hollands, Hoyoak, Nisbett, & Thagard, 1987) suggest that it may be necessary to direct attention toward the features of the discrediting experience in order for the contradictory evidence to be encoded. Left to their own devices, people will attend only to those features predicted by their current information structure. Adapting this approach, we had students make predictions on the pre-test, and then use the software to check their answers by embedding the assessment instrument into the activity. When students discover that their prediction is incorrect, this creates cognitive dissonance between the student's current knowledge or expectation and what they are seeing. Students are then able to utilize the software to identify and correct their misconceptions. The improvement in post-test scores was dramatic when this approach was used for 141 students at both schools. Furthermore, the activity allowed us to better track student misconceptions, and what knowledge was lacking in their understanding of sampling distributions. We then altered the activity to better address the most prevalent misconceptions, e.g. the distinction between the sample and sampling distribution.

More recently, we utilized a pre-test of basic skills to clarify misconceptions in prerequisite knowledge before students interact with the program. We also embedded the activity into a contextual example, hoping to help students learn to apply the implications of the Central Limit Theorem. We again administered post-tests in our different institutional settings and compared post-test scores (55 students) on the graphic based questions for two population shapes to scores from previous versions of the activity. However, these results were not as impressive. There are several possible explanations that we are now investigating:

- Insufficient development and definition of sampling distributions in lecture prior to use of the computer program: How the concept was introduced and the amount of time spent varied. What do students need to understand prior to their interaction with the activity?
- Student engagement in "prediction questions": Previously, pre-test predictions were turned in to the instructor before students used the program to check their knowledge. The recent activity relied on the student to invest sufficiently in the activity to create significant dissonance.
- Length of the activity: Incorporating the contextual example significantly lengthened the activity. The current activity may be trying to convey more information than the students can attend to in one interaction with the simulation program.
- Number of predictions: To compensate for the additional length of the activity, fewer "prediction" questions were asked. Before, students worked through 5 distinct populations, the latest activity involved 3. Perhaps this prevented students from receiving sufficient practice or opportunity to refine and reinforce their conceptual framework.

We continue to investigate these questions using a collaborative assessment framework which allows us to share more detailed and informative results with the research community. Our framework now consists of: pre-test measures of prerequisite knowledge and intuitions that may interfere with the learning activity; listings of assessment goals used to develop the learning activity; assessment items embedded in the learning activity to promote conceptual change; post-test measures of desired outcomes assessing correct and incorrect types of reasoning. We have also developed post-test questions that are being used on final exams at our institutions and in classrooms that do not use the same simulation program. This will help us identify long-term gains from use of the program. These instruments are freely available at our website: http://www.gen.umn.edu/faculty_staff/delmas/stat_tools/index.htm

IV. Conclusion

While we have not identified a definitive approach to teaching sampling distributions, our assessment framework has provided substantial insight into students' misconceptions and their sources. For example, the establishment of cognitive dissonance appears to be a crucial component to effective interaction with technology, providing students with the opportunity to immediately test and reflect on their knowledge in an interactive environment. By incorporating the assessment into the activity, both students and faculty have learned much more about the processes involved, receiving immediate feedback on the students' level of understanding. Working together in different colleges and educational settings over several years, we've learned a great deal about why an activity works, how students' understanding and reasoning are affected, and how prior knowledge affects their experience with the activity, all of which suggest changes for improved teaching practice.

REFERENCES

Cobb, G. (1992). Teaching Statistics. In Heeding the Call for Change: Suggestions for Curricular Action (ed. L. Steen). Mathematical Association of America Notes, No. 22.

Garfield, J. (1993). An authentic assessment of students' statistical knowledge. In National Council of Teachers of Mathematics 1993 Yearbook: Assessment in the Mathematics Classroom (ed. N. Webb), 187-196. Reston, VA: NCTM.

Hawkins, A., Joliffe, F., and Glickman, L. (1992). Teaching Statistical Concepts. London: Longman Publishers.

Holland, J. H., Holyoak, K.J., Nisbett, R.E., and Thagard, P.R. (1987). Induction: Processes of Inference, Learning, and Discovery. Cambridge, Massachusetts: The MIT Press.

Jennings, D., Amabile, T., and Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In Judgment Under Uncertainty: Heuristics and Biases (eds. D. Kahneman, P. Slovic, and A. Tversky), Cambridge: Cambridge University Press.

Lord, C., Ross, L., and Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology 37, 2098-2109.

Nickerson, R. S. (1995). Can technology help teach for understanding? In Software Goes to School: Teaching for Understanding with New Technologies (eds. D. N. Perkins, J. L. Schwartz, M. M. West, and M. S. Wiske. New York: Oxford University Press.

Posner, G.J., Strike, K.A., Hewson, P.W., and Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education, 66(2), 211-227.

Ross, L. and Anderson, C. (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In Judgment Under Uncertainty: Heuristics and Biases (eds D. Kahneman, P. Slovic, and A. Tversky). Cambridge: Cambridge University Press.

Snir, J., Smith, C., and Grosslight, L. (1995). Conceptually enhanced simulations: A computer tool for science teaching. In Software Goes to School: Teaching for Understanding with New Technologies (eds. D. N. Perkins, J. L. Schwartz, M. M. West, and M. S. Wiske). New York: Oxford University Press.

FRENCH RÉSUMÉ

Cette communication presente le compte rendu d'une etude de la recherche en collaboration en classe afin d'etudier l'impact de la technologie sur la comprehension de la statistique chez les etudiants. De telles etudes exigent une conceptualisation autre de l'evaluation et son role dans l'enseignement de la recherche en statistique.

This paper summarises a collaborative classroom research study to investigate the impact of technology on student understanding of sampling distributions. Such investigations require an alternative conceptualisation of assessment and its role in statistics education research.