NEW APPROACHES TO GATHERING DATA ON STUDENT LEARNING FOR RESEARCH IN STATISTICS EDUCATION[1]

BETH L. CHANCE
Cal Poly – San Luis Obispo
bchance@calpoly.edu

JOAN B. GARFIELD
University of Minnesota
jbg@umn.edu

SUMMARY

Over the last fifteen years there has been a strong emphasis on active learning, use of real data in the classroom, and innovative uses of technology for helping students learn statistics. A recent survey in the United States (Garfield, 2001) documents that many tertiary teachers of statistics courses have made changes toward these recommendations. Now more than ever, more research is needed on the effects of these instructional methods and materials on student learning, retention, and motivation. This research need first requires the determination of effective research methodology in statistics education. In assessing students' conceptual understanding, reasoning abilities, and attitudes, and their development, alternative methods of gathering student data are needed that supplement comparative experiments and improve on traditional assessment items that focus on calculations, definition, and rote manipulations. This article will present and critique additional methods for obtaining research data on how students develop an understanding of statistics, including classroom-based research and videotaped student interviews/observations.

Keywords: Statistics education research; Assessment; Classroom-based research; Clinical interviews

1. INTRODUCTION

During the last fifteen years, educators have witnessed a movement in statistics education aimed at shifting the focus of instruction away from theory and recipes toward statistical thinking, genuine data, conceptual understanding, and active learning. Much of this movement has been motivated by research in educational psychology, psychology, mathematics education, and science education (see Garfield, 1995). However, there is still need for documented evidence of whether such changes enhance student learning, retention, and appreciation of statistics. This evidence is also needed to determine the most effective instructional techniques and to develop models of how students shape their statistical understanding. Without more published research, we will not be able to continue to move these changes forward and expand their impact.

While there is much overlap with research questions in mathematics and science education, statistics education poses a unique set of challenges. For example, problem *context* plays a role in statistics that is not paralleled in mathematics (Cobb and Moore, 1997). Furthermore, probabilistic reasoning and randomness appear to require distinct teaching and learning strategies (see e.g., Falk and Konold, 1992). While there has been a strong increase of activity and publications in statistics education in recent years (e.g., *Journal of Statistics Education*, International Conference on Teaching Statistics (ICOTS), Statistical Education Research Section of the International Association for Statistical Education (IASE)), more discussion and reflection are needed to clarify "what should be considered as research in statistics education, how we establish the validity of our research findings, what priority questions need to be studied, and what theoretical frameworks and research methods might be recommended to carry out this research" (Batanero, et al., 2000). These questions need to be answered

---

[1] *Statistics Education Research Journal, 1(2), 38-41, http:/fehps.une.edu.au/serj*
International Association for Statistical Education

in order to establish visibility, legitimacy, and understanding of research results, as well as to form a basis for on-going research and training of future researchers. In order to advance the field of statistics education and achieve academic recognition, it is essential to have a well-developed research literature and research agenda (Batanero, et al., 2000). Moreover, little is known about, or has been published on, the methodology of statistics education research (Jolliffe, 1998). Statistics education research will not be able to sufficiently impact policy or the practice of teaching until individual studies become grounded in a broader program of study.

2. CURRENT TRENDS IN STATISTICS EDUCATION RESARCH

The teaching of statistics is unique in that instructors typically come from a variety of backgrounds and fields of application. Similarly, statistics education research has been conducted by a variety of individuals who represent different disciplines, educational programs and training in research methods. Research in psychology has revealed ways people reason about statistical or probabilistic information, paying particular attention to faulty reasoning and misconceptions (e.g., Garfield, 1998). However, this research has not been expanded to all age groups or closely tied to teaching practice. Similarly, studies conducted by mathematics education researchers have focused on how children reason about particular topics such as averages or graphs. There has also been work on how K-12 teachers understand and reason about statistics. Recent research efforts in statistics education have focused on comparisons of two types of instruction (e.g., laboratory environment vs. traditional lecture) or prediction of achievement based on mathematical ability, attitudes, and other variables. Attention has also been directed at evaluating student ratings of new implementations. While all these types of research studies offer interesting results, they are often limited in their generalizability and validity. This focus needs to be expanded (to other types of questions, across all age groups, towards research on teachers of statistics) and to tie the research more closely to classroom practice. In particular, little of this research has focused specifically on statistical reasoning or other issues unique to statistics education.

A concern regarding many of these studies is that they lack visibility as well as cross-disciplinary or cross-institutional collaboration. While there are some large conferences, such as the ICOTS series, there appears to be little connection among researchers in the years between these conferences. In an effort to coordinate these research efforts, there are several research study groups aimed at statistics education including the Statistical Reasoning, Thinking and Literacy International Research Forum, and the Statistics Education Research Group based at the University of Massachusetts at Amherst, as well as the RSS Centre for Statistical Education. While these organizations have been conducting and supporting statistics education research, they are relatively new (so support is not extensive) and they have not been as focused on establishing standards of research, or exploring effective methods. It can also be difficult for a new researcher to find an obvious focal point of these efforts.

Consequently, the need to explore the role and future of statistics education research has recently become a theme of sessions at international conferences. For example, a session of the Sixth ICOTS was held in July 2002 was entitled “Research in Statistics Education” and included talks on developing statistics education research; theoretical models of statistical knowledge, thinking, reasoning, and learning; and a roundtable discussion of major problems and directions in statistics education research. Thus, discussions are beginning on how to legitimize statistics education as a research domain and how to train future researchers in statistics education.

We see the pressing needs as falling into two categories. One is documenting evidence of the effects on students of these instructional changes, trying to identify the most effective instructional techniques, while also developing models of how students come to understand statistics which will help foster additional reform and be closely tied to practice. The other is generating more discussion and reflection on acceptable research methods and a research agenda. If we hope to establish the validity and legitimacy of statistics education as an area of research then we need a well-developed research literature that we can point people to, as well as accountability in our methods. These discussions will also help inform the training of future researchers.

3. LIMITATIONS OF CURRENT RESEARCH ACTIVITIES

As statistics education research becomes a more visible discipline, it is vital to consider appropriate research techniques. When formulating a research methodology, we must meet several criteria. The research must be valid so that the data correspond to what the research purports to measure and so that the information can be generalized beyond the study at hand. The knowledge gained must be reliable, consistent, and replicable, while providing sufficient documentation of the activities and observations involved. There must be objectivity and belief that the research process is not affecting the outcomes being measured and that the evaluation is fair and equitable.

In designing studies to meet these criteria, the obvious place to start is with the gold standard, the randomized comparative experiment. For example in clinical trials, we can impose a treatment, control the environment, and draw clear cause and effect conclusions. However, when trying to apply this to the education world, things become more complicated. For example, the first step – randomization – is not really possible in the educational setting, especially if we are looking for long-term, even semester long, effects. It is not feasible for us to manipulate which courses students will take and self-selection into sections is not a sufficient substitute for randomization. Student mobility between sections and drop rates are also serious difficulties. Thus, we will not be able to maintain the independence of observations that is required by our traditional statistical techniques. Variables interacting with the instructional environment such as instructor attitude, time of day, resources available, and classroom culture may have dramatic effects on student achievement and attitude and cannot be controlled or even measured as in a laboratory setting.

It also may not be feasible to create the classroom implementation that is of interest. For example, in deciding what factors most directly affect student achievement, to be an authentic reflection of student performance, the assessment most likely needs to be tied to a grade and there are ethical issues involved in manipulating variables that could adversely affect student grades. In the standard research model, we can try to design a more controlled environment but this leads to separation of the research from classroom practice and to significant time delays before the research can be applied to the classroom environment. Often the results from a controlled laboratory setting are not immediately relevant.

Furthermore, this type of research too often ignores the integral role played by the classroom instructor. Trying to achieve objectivity through an outside observer ignores that the presence and demeanour of the observer can also affect the results. An external researcher may not sufficiently understand the details of the classroom environment and culture, or even have the subject matter knowledge to adequately appreciate and document students' experiences. Directly involving the classroom participants (teachers, curriculum designers, students) in the research and immersing the researcher into the classroom environment allows for more in-depth study, more reflection, and better interplay between theory and practice.

The above is not an exhaustive discussion of the issues involved but does begin to support the view that the classroom environment is a sufficiently complex and dynamic world that is not always well described by traditional research techniques. Similarly, many traditional measurement techniques, such as standardized exams, final course grades, and student ratings, are not sufficient, especially when trying to measure student reasoning. Studies have shown that students can do well on final exams but then still demonstrate poor statistical reasoning on other tasks. As Lesh and Lovitts (2000) caution, "Most existing high-impact standardized tests are poorly aligned with national standards for instruction and assessment." Current assessment methods typically are not dynamic in nature and fail to inform the researcher of the learning processes involved. One-time measures of achievement also fail to explore the developmental nature of learning or provide concurrent feedback to the study, nor allow an iterative research approach. In particular, traditional assessment strategies do not tell us enough as to why a particular teaching method or activity works, how students' understanding and reasoning are affected or unaffected by the learning experience, nor provide direction for how teaching practice should be changed. Advances in technology also enable new methods of measurement and the collection of different types of student data, from tracking student focus on a computer screen to extensive videotaping and internet-based records. These new techniques allow movement beyond multiple choice, standardized paper-and-pencil exams. This movement is crucial in order to expand the types of information we obtain and the types of research questions we can ask.

Thus, effective research requires a careful combination of appropriate research methods and assessment techniques. Alternative approaches include classroom-based research, teaching experiments, naturalistic observation, and videotaped interviews. These techniques have been used in other disciplines such as social science and anthropology, and a recent handbook (Kelly & Lesh, 2000) acquaints mathematics and science education researchers with these approaches. However, many statistics education researchers do not have background in these different areas. A statistician in particular is typically schooled in controlled experiments and may not be familiar with theories of learning, assessment, or education.

4. NEW DIRECTIONS

While the standards of research are constant, how they are met is not. While traditional research methods also aim for "hard data" that allow cause-and-effect conclusions, alternative methods of research aim for systematic, scientific investigations from which inferences can be drawn. More and more investigators are replacing purely statistical procedures with the collection of rich, diverse data from multiple sources that document the situation being investigated and provide a scholarly account of the situation and/or the intervention.

For example, Moschkovich and Brenner (2000) outline how the above research standards are met in naturalistic research: Validity can be obtained through prolonged investigation, immersion into the environment, triangulation of multiple data sources, and frequent member checks of interpretation. Generalizability can be achieved through extensive description of the classroom situation being investigated and multi-site designs. Such documentation allows the reader to determine if the results are relevant to their environment, shifting the responsibility from the researcher to the reader. Reliability is enhanced by combining efforts and perspectives of multiple researchers and by direct involvement of participants in the research program. Objectivity can be assessed through extensive documentation and dissemination. However, naturalistic researchers also admit that pure objectivity may not be feasible or even desirable and instead aim for acknowledged and controlled subjectivity through researcher-participant immersion, while being explicit about how prior assumptions and beliefs could be influencing the research observations.

Classroom-based research, or Action research, has been defined as "ongoing and cumulative intellectual inquiry by classroom teachers into the nature of teaching and learning in their own classrooms" (Cross and Steadman, 1996; Feldman & Minstrell, 2000). Thus, instead of ignoring the integral role the teacher plays in the learning process, the classroom teacher becomes a key partner in the research team, helping to develop the questions to be investigated and assisting in data collection. By directly involving the teachers in the research process, classroom research aims to incorporate their perspectives, insight, and understanding of the classroom culture into the analysis. This approach also allows for further probing into the student and instructor experience, and adjustments in the evaluation process can be immediately implemented. The gap between theory and practice is narrowed and the evaluation becomes a dynamic process that changes in response to results and feedback, while simultaneously focusing on curricular development, instruction, and assessment. Below we offer an example of a classroom based research project we have been conducting to investigate how interaction with a conceptually based interactive software program helps students develop statistical reasoning skills.

Videotaped clinical interviews build on techniques used by cognitive psychologists. These methods are used to study the form of knowledge structures and reasoning processes (Clement, 2000). Researchers in statistics education are utilizing these techniques in studies that explore student understanding of data, relationships, tendency, and inference. These studies have been helpful in generating models, and allow for independent viewers/ coders to compare their interpretations and see if there is convergent validity in their findings. For example, participants at the International Research Forum on Statistical Reasoning, Thinking, and Literacy bring videotapes with them to share with colleagues in small working groups to discuss and validate their results. (See http://www.beeri.org.il/srtl/.)

5. EXAMPLE

We now provide an example of a classroom research project used to investigate how student interaction with a simulation program affects their statistical reasoning. Our goals were to understand student thinking and to inform other instructors about the use of such a simulation program to teach the topic of sampling distributions. This is a notoriously difficult topic for statistics students to understand, but is also the gateway to understanding statistical inference. We hoped to provide insight into *how* to best integrate technology into instruction, *why* particular implementations appeared to be more effective, and *how* student understanding evolved through use of the program. Thus we were more focused on understanding students' knowledge structure and reasoning process than on establishing a simple cause and effect relationship.

We began gathering data in three diverse college settings: an introductory statistics course for non-traditional students, an introductory statistics course for business and science majors, and a graduate level course in education. This allowed us to work with students from a wide variety of programs of study, educational backgrounds, and ages. As researchers, we also had a wide variety of backgrounds from cognitive science to educational psychology to statistics. We were able to create the learning environment that we wanted to investigate and fully integrate this into the existing course (students used a dynamic, interactive computer program developed by delMas, see delMas, 2001). This allowed us to continually relate what we were observing in the classroom to existing theory while also generating new models of student learning.

We have used a wide variety of measurement techniques to try to capture student reasoning (Chance, Garfield, & delMas, in press). In the first stage of the research, we utilized graphics-based test items to determine whether students could demonstrate a visual understanding of the implications of the Central Limit Theorem by choosing the appropriate pictures that corresponded to an empirical sampling distribution for different sample sizes. Initially, students were asked to justify their choice of graphs and explain their reasoning. These responses were then categorized so that future instruments asked students to select which statement best represented their own reasoning. Students were given these test instruments before using the program and after using the program the next day in order to isolate the change in understanding from interacting with the program. We also developed some open-ended questions where students had to provide their own justification and reasoning for their answers and some post-test application problems that could be given later in the course to see if they could still apply and use their knowledge.

In an effort to improve student performance further, we incorporated a model of conceptual change. Students were asked to select their responses and then use the program to check their predictions. This forced students to more directly confront the misconceptions in their understanding and improvements in their performance on the post-test were significant (delMas, Garfield, and Chance, 1999). This led to additional investigations in the role of pre-requisite knowledge, development of tools for identifying prevalent misconceptions, and further refinements to the activity.

Currently, our research is focused on formulating a model of student development of statistical reasoning. Through videotape analysis we are documenting student explanations of their reasoning using students at several levels of development to validate this model. For example, this has allowed us to further document student choices, and how correct choices are often still accompanied by faulty reasoning. These methods have allowed us to gain much more in depth understanding of individual students and what they know.

This brief summary showcases how collaborative, classroom-based research can be used in statistics education. Does this type of research meet the standards we discussed before? We feel that through prolonged investigation and immersion into the learning environment and by gathering multiple sources of data and looking at the consistency in information from these multiple sources, then we do have validity in what we are measuring. We don't claim to have used all instructional settings but that they were diverse enough that something that works in these three situations has a good chance of working in other settings as well. By extensively describing our situation then, the readers can decide if the results are relevant to their environment. We also feel it is important to have our multiple perspectives to be able to check either other's observations in what we are seeing in these data as well as to directly involve the teacher in the study (instead of the top-down model where the teacher is told to try something based on theory where they may not have the belief or skills). Still, it is important to have the non-participant viewpoint as well and through documentation and public dissemination we allow others to evaluate whether we were able to achieve that outsider perspective.

We also see additional benefits. This approach allows the classroom results to immediately provide feedback to the theory. It also provides direct access to the students and instructor, focusing more on the process than the end result, which allows for more probing and follow-up. It is also a dynamic process that allows for immediate adjustment in the research process and creates a much more iterative approach. This allows the student response to drive the investigation more than our prior beliefs. We feel we have learned a lot more about students' understanding than we would have ever anticipated.

Using our own students as "subjects" allowed direct access to the students and the ability to specifically create the desired learning environment, as well as additional insight into the students' experiences. We have used our diverse perspectives in the development of the learning environments and the cross-checking of interpretations and evaluations. The investigations were tied to existing theory and are helping to generate new theory. The project has led to the development of new assessment instruments that are now available to other instructors. While no part of our research could be considered a traditional experiment, we feel we have contributed insight into why an activity works, have demonstrated transferability in the learning gains to other instructional settings, and are continuing to employ a variety of research methods to evaluate the progress of this project.

6. SUMMARY

This paper has examined new approaches to gathering data on student learning. Many of these techniques can be applied equally well to statistics education research. Perhaps the two most important lessons to take from this discussion are that the research methodology must match the research question and that a variety of tools should be employed: “Different techniques generate different types of information, and it is often the case that a single technique will not provide the breadth of information necessary to answer unequivocally the research questions under investigation” (Mestre, 2000). These techniques should combine qualitative and quantitative data: “There is no single correct approach to evaluation problems. The message is this: some will need a quantitative approach; some will need a qualitative approach; probably most will benefit from a combination of the two” (Herman, Morris & Fitz-Gibbon, 1987). We are not trying to say one should never do experiments in an educational setting. There are certainly (very focused) questions and situations where such methods are appropriate. Instead we want to advocate using a variety of tools, using different techniques to answer different kinds of questions, gathering both qualitative and quantitative data. Even with clinical trials, pharmaceutical companies spend millions of dollars on research before the clinical trial stage. Relaxing the strict adherence to classical experimental methods will allow richer sources of information through complementary techniques and new research questions. Statistics education research in particular is still developing a set of coherent research questions. By gathering information from a variety of sources we will be able to develop more informed research questions. This development will be greatly aided through additional naturalistic observation and documentation of students prior to more systematic investigations.

The gathering of data is closely connected to theoretical frameworks, prior research, and type of methods and design used. In considering ways to improve the information gathered in educational research studies in statistics, we need to keep in mind the context of the growing and developing discipline of statistics education. Toward this end, statistics education needs to establish standards for preparing researchers in statistics education. This includes recommended coursework in statistics, education, learning theory, measurement, and qualitative and quantitative research methods. Clearly, future researchers in statistics education need to have training and cross-disciplinary collaboration in psychology, education, math and science education, and alternative research methodologies. The body of statistics education research also needs to become more visible and accessible across disciplines. This should include publication of examples of high quality research, literature reviews, bibliographies on certain topics, and a research handbook specific to statistics education research. Such tools will enable researchers to combine traditional methods with alternative approaches in order to best answer a wider array of research questions.

ACKNOWLEDGEMENTS

This discussion was originally presented at the 53rd Session of the International Statistical Institute, Seoul Korea, August, 2001. A shorter version (Chance and Garfield 2001) appears in the *Proceedings of the 53rd Session of the International Statistical Institute*, 2001.

REFERENCES

Batanero, C., Garfield, J., Ottaviani, M. G., & Truran, J. (2000). Research in statistical education: Some priority questions. *Statistical Education Research Newsletter*, 1(2), 2-6.

Chance, B. & Garfield, J. (2001). New approaches to gathering data on student learning for research in statistics education. *Proceedings of the 53rd Session of the International Statistical Institute. Bulletin of the International Statistical Institute* (Tome LIX, Book 2, pp. 205-208). Seoul, Korea: International Statistical Institute.

Chance, B., Garfield, J., & delMas, R. (In press). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.) *The challenge of developing statistical literacy, reasoning, and thinking*, The Netherlands: Kluwer Publishers.

Clement, J. (2000). Analysis of clinical interviews. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 547-589). Mahwah, NJ: Lawrence Erlbaum Associates.

Cobb, G. & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104(9), 801-824.

Cross, K. P. & Steadman, M.H. (1996). *Classroom research: Implementing the scholarship of teaching*. San Francisco: Jossey-Bass Publishers.

delMas, R. (2001). *Sampling SIM* (version 5). http://www.gen.umn.edu/faculty_staff/delmas/stat_tools/

delMas, R., Garfield, J. & Chance, B. (1999). Assessing the effects of a computer microworld on statistical reasoning, *Journal of Statistical Education*, 7(3).

Falk, R. & Konold, C. (1992). The psychology of learning probability. In *Statistics for the twenty-first century* (pp. 151-164). MAA Notes, Number 26, The Mathematical Association of America.

Feldman, A. & Minstrell, J. (2000). Action research as a research methodology for the study of the teaching and learning of science. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 429-456). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63, 25-34.

Garfield, J. (1998). The challenge of assessing statistical reasoning. Paper presented at *American Educational Research Association Annual Meeting,* San Diego.

Garfield, J. (2001). *Evaluating the impact of educational reform in statistics: A survey of introductory statistics courses.* Final Report for NSF Grant REC-9732404. Minneapolis, MN.

Herman, J.L., Morris, L.L., & Fitz-Gibbon, C.T. (1987). *Evaluator's handbook, second edition*, Newbury Park, CA: Sage Publications.

Jolliffe, F. (1998). What is research in statistical education? In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 801-806). Singapore: International Statistical Institute.

Kelly, A. & Lesh, R. (Eds.), (2000). *Handbook of research design in mathematics and science education* (pp. 361-512). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lesh, R., & Lovitts, B. (2000). Research agendas: Identifying priority problems and developing useful theoretical perspectives. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 45-71). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mestre, J. (2000). Progress in research: The interplay among theory, research questions, and measurement techniques. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 151-168). Mahwah, NJ: Lawrence Erlbaum Associates.

Moschkovich, J. N. & Brenner, M. (2000). Integrating a naturalistic paradigm into research on mathematics and science cognition and learning. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 361-512). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

BETH L. CHANCE
Department of Statistics
California Polytechnic State University
San Luis Obispo, CA 93407