COLLABORATION IN STATISTICS EDUCATION RESEARCH: STORIES, REFLECTIONS, AND LESSONS LEARNED

Joan Garfield
University of Minnesota, United States
jbg@umn.edu

In this paper I summarize my 25 years of research on teaching and learning statistics, as I participated in the emergence of statistics education as a research discipline. This summary and reflection are presented through stories of research projects I have been involved in, all of which involved collaborations with colleagues who have made important contributions to the research and from whom I have learned many important lessons. I summarize and reflect on three interconnected areas of research: synthesizing and building on research studies across diverse disciplines, developing and using good assessment instruments to evaluate and improve student learning, and studying the role of technological tools in developing students' reasoning about specific concepts.

INTRODUCTION

For the past 25 years I have been involved in collaborative research on teaching and learning statistics. In this paper I reflect on the different types of collaborative projects I have been involved in and by sharing stories of my personal experiences, summarize what I learned regarding the content of these studies and offer some general lessons learned. I have divided the paper into four sections. The first three are each focused on one area of research and the last section includes some general advice to new researchers and final remarks.

FIRST STORY: SYNTHESIZING THE RESEARCH LITERATURE

Beginning with my doctoral dissertation, I learned that an important part of a research study is to try to locate and read as many articles as possible that relate to the topic of interest. In some disciplines this is a routine process: there are well established journals and conference proceedings in which to locate the relevant background readings for a literature review. Twenty-five years ago, this was not true in statistics' education. It was quite a challenge to locate relevant research publications, which at that time appeared in publications from areas such as psychology, educational psychology, statistics, sociology, mathematics education, health sciences, and business. Finding these publications required me to go to different libraries on campus to read and copy articles and in some cases, to request papers from international journals (e.g., *Teaching Statistics*) via interlibrary loan. At that time there were no summaries or syntheses of this research literature or guidelines to follow in trying to write a literature review related to teaching statistics.

Finding sources to write my first comprehensive literature was enormously challenging, but a necessary first step in my doctoral dissertation research. This review later provided the basis for my first major publication: a literature review written with my first mentor, Andrew Ahlgren (Garfield and Ahlgren, 1988). Andrew Ahlgren was not my advisor, but he helped me to synthesize and critique the literature I had located and read. He taught me to look outside the discipline for related work in other areas, such as physics education. He taught me about writing a well crafted paper, and he led me through the process of seeking out people who seemed to be doing interesting work in this area so we could talk to them and learn about their current research. That led to a phone call to a group of psychologists at the University of Massachusetts, Amherst, who had published some interesting studies on the faulty statistical reasoning of college students (e.g., Pollatsek *et al*., 1981) to learn more about their work. The person who answered the phone that day, Cliff Konold, soon became a close friend and colleague.

Our search for information also led us to contact David Green in the UK, who had conducted a large-scale survey of school children's understanding of probability concepts. We learned that he headed the International Study Group for Research on Learning Probability and Statistics, an informal network of researchers in this area, which he later asked me to take over as chair (Garfield and Green, 1988). As a result of our initial contacts and correspondence, David invited Andrew and I to present our work (a summary of difficulties students have learning statistics) at the second International Conference on Teaching Statistics, (Garfield and Ahlgren, 1986). Later, this contact led to an invitation by David, the editor of *Teaching Statistics* to establish a regular research column in that journal. This column was entitled Research Report, and described the work of researchers around the world who were studying the teaching and learning statistics.

The response to my first major publication also led to many requests for talks on the topic of how students learn (and fail to learn) statistics, including a plenary talk to the American Statistical Association's Winter meeting (1992) that led to a second published literature review (Garfield, 1995) and a collaborative review with a master of this genre, Mike Shaughnessy (Shaughnessy, Garfield and Greer, 1997). My initial literature review also led to collaborating on and contributing a chapter to *Chance Encounters* (Ahlgren and Garfield, 1991) an edited volume that combined theory, research, and practical issues in probability education. It seems that one thorough review and synthesis of the literature launched my career and provided valuable contacts as well as led to subsequent publications for many years.

I continue to find it valuable to read and synthesize research studies that are conducted and published in many different disciplines, but now my focus is to try to relate the reviews to practical suggestions for teaching and learning statistics at the college level. I find that many college statistics teachers are very interested in accessing this information in a way that can help them apply research to improve their teaching. A central purpose of my current book (Garfield and Ben-Zvi, in preparation) is to provide this service to statistics teachers. By reading, discussing, and synthesizing the research related to learning specific concepts in statistics, Dani Ben-Zvi and I have been able to suggest sequences of teaching activities and to adapt and design lesson plans to illustrate these sequences. We hope this material will be useful to teachers and we have found it tremendously challenging and rewarding for our own efforts as teachers and as researchers.

WHAT I LEARNED ABOUT TEACHING AND LEARNING STATISTICS FROM SYNTHSIZING THE LITERAUTRE

I learned (and continue to marvel at) the pervasiveness of the difficulties both students and adults have learning and using statistical ideas. It seems clear that many statistical intuitions are incorrect and misconceptions are stubborn and hard to overcome. It also appears that many seemingly simple concepts (e.g., the mean) are actually quite complex for students to understand beyond performing the calculation. In addition, I have learned that being able to compute a statistic or carry out a procedure does not mean that students understand the underlying concepts. It also seems that learning is fragile, and that concepts that appear to be understood one day may disappear the next, unless they are continually revisited and reinforced in different contexts.

I have learned that it is important for teachers of statistics not to overestimate what students understand or have learned from instruction just as it is important not to underestimate the difficulty of what may appear to be simple concepts. It is also good for teachers to be aware of what some of the difficulties and stubborn misconceptions are, so they can look for and assess their prevalence among students in their courses.

The type of literature review that seems most helpful for teachers of statistics is one that builds on a common set of readings and theory but focuses on a particular concept (e.g., variability), and provides some specific suggestions to teachers about how to help students construct the important, "big" ideas, via sequences of carefully designed activities (Garfield and Ben-Zvi, 2005).

LESSONS LEARNED ABOUT DOING RESEARCH

- When studying a problem in statistics education, it is important to hunt across many disciplines for relevant research. Although articles in different disciplines may look at the problems differently, build on different theories and prior research, use different methods, and focus on different types of subjects, together they can provide a rich background for future research studies.
- Writing a good, comprehensive literature review on your topic of interest in statistics education is an important and valid activity. This review can help identify how a new study

may actually contribute to knowledge in the field and provides a solid base for future research.
- Doing a good literature review in statistics education is like detective work, as you try to find out who has done relevant work and what you can learn from them. This can lead to meeting new people and forming new collaborations. These contacts can be extremely valuable to your learning and development, as well as opening avenues for involvement in the community. Publishing a high quality literature review and becoming visible in the community can lead to many new productive activities and collaborations.

SECOND STORY: STUDYING ASSESSMENT OF STUDENT LEARNING

My early work in assessment developed in three parallel strands and with three sets of collaborators. The first was my work on developing and using assessment items in a research study on misconceptions about probability (with Bob delMas). The second was developing assessment instruments as part of the Chance Plus project (with Cliff Konold). The third was an assessment conference and handbook initiated by Iddo Gal. All three of these projects led to my current work with Bob delMas, Beth Chance, and Ann Ooms on the ARTIST Project: Assessment Resource Tools for Improving Statistical Thinking (see https://app.gen.umn.edu/artist/).

The first project began with a new collaborator: Bob delMas, whom I met soon after I began as an Assistant Professor at the University of Minnesota. Bob was in graduate school, working with the same mentor I had found, Andrew Ahlgren, who introduced us. Bob was interested in misconceptions college students have about probability, and soon dreamed up a software program called Coin Toss, to help students confront their misconceptions about chance events related to coin tossing. I began to collaborate with him, helping to design activities to use with Coin Toss, using the software in my statistics classes, and helping to design assessment items to give to students to evaluate the impact of the software on their learning. We spent a lot time developing and analyzing assessment items, and found that it was important to look at more than responses to individual items, but to also study patterns of responses to items, when determining whether students held certain misconceptions about probability. We presented this work at ICOTS-2 (Garfield and delMas, 1991).

At the same time that I began to work with Bob on assessing probabilistic reasoning, I began to work with Cliff Konold as the evaluator for his new NSF project Chance Plus. My main role in this project was to develop instruments to assess students' statistical reasoning after using the newly developed activities and software (*DataScope* and *ProbSim*). We ended up developing two instruments, one of which eventually became the Statistical Reasoning Assessment (SRA) (Garfield, 1991, 2003). A second instrument, "Statistics in Context," later became part of the ARTIST item data base.

My work on the SRA continued beyond Cliff's grant, and led to a small internally funded study to translate that instrument into French and Spanish, and to administer it to students in other countries. That project led me to establish new collaborations with Marie-Paule Lecoutre (in France) and Carmen Batanero (in Spain), both of whom administered translated versions of the instrument in their countries (Batanero *et al*., 1996). I developed a rubric to relate the items to specific misconceptions and types of correct reasoning. I also worked with a graduate student studying educational measurement in my department, H.J. Liu, who translated the SRA into Chinese and conducted a comparative study in Taiwan and the US (Liu and Garfield, 2002), and who gathered data to determine the reliability and validity of the SRA. We found that Statistical Reasoning is very complex, not consistent from topic to topic, and not related to traditional assessments in a statistics course (e.g., exams).

The third area of my early assessment work was with Iddo Gal, who saw the importance of this work and the need to assemble, develop, and disseminate information and resources to the statistics education community. We first met in 1992 while he was working in a postdoctoral position, and discussed our common interest in assessment. He then became the driving force behind an NSF-funded assessment conference held in 1994, and a co-edited book (Gal and Garfield, 1997). As part of our work together, I tried to find out what we could learn about assessment from other areas, in particular, from Mathematics Education. I became connected with the assessment community in Mathematics Education as part of the NCTM Research Catalyst project and tried to bring many of their ideas to statistics education (Garfield, 1993, 2000).

I learned a tremendous amount form my collaboration with Iddo. I learned what it really means to edit a book and how a good editor can greatly improve a chapter and create a high quality, unified book. It is no surprise that Iddo now provides his top-notch editing skills to the IASE Statistics Education Research Journal, where I am sure many authors and associate editors continue to learn from him, and where I continued to benefit from his advice when serving as guest editor and author for two special issues of that journal (Ben-Zvi and Garfield, 2004; Garfield and Ben-Zvi, 2005).

My experience with the SRA and with the assessment book soon led me to an important new collaborator: Beth Chance, who had recently completed her PhD. She was a new college teacher of statistics and extremely interested in student learning and assessment. We began to collaborate on many projects (some described in the next story) and I enjoyed watching her develop into an important contributor and then respected leader in Statistics Education. I especially enjoyed writing papers about assessment with Beth, and found that she offered a critical perspective that challenged me to tighten my writing, provide more examples, and pay more attention to details. She also brought her expertise as a grader of AP Statistics exams and later as an AP test developer to our assessment work and writing (Chance and Garfield, 2002; Garfield and Chance, 2000).

In my work on assessment over the years, I found that some instructors and researchers did not pay much attention to the assessment instruments they used. For example, in many empirical studies based in statistics classes the outcome measures used were (often poorly defined) final exams or end of course grades. There instruments were not valid and reliable measures of important learning outcomes, and detracted from the studies' results and implications. It also seemed that many instructors based their assessment primarily on computations and definitional knowledge, rather than on conceptual understanding or statistical thinking. However, I also found that many instructors and researchers were often seeking high quality assessment materials and methods to evaluate student learning of statistics. Despite the widespread interest in the SRA I felt that the instrument had several flaws and that a better instrument needed to be developed.

In 2002, Bob, Beth and I were funded by the National Science Foundation to develop online resources for teachers of statistics (the ARTIST project). One component of this project was to develop a new instrument, the *C*omprehensive *A*ssessment of *O*utcomes in a First *S*tatistics course (CAOS), to provide a high quality, valid, reliable test for use by teachers and researchers in statistics education. This 40-item, multiple-choice test has been carefully developed and validated, and used in a large scale class testing in an online format. We have learned an enormous amount through this process of developing, validating and testing this instrument. Some of it is about assessment and some of it is about student understanding of basic statistical concepts. We have learned how an item that seems clear and well written can mean something different to a colleague, or is missing something that is important. This has lead to at least 40 revisions of the current CAOS test. We have seen items that seem quite easy and straightforward elicit incorrect answers and reveal errors in student's reasoning that we had not anticipated.

In addition, we have collected, revised, and coded thousands of items for our searchable online data base and test builder (The Assessment Builder). Without working closely with my three ARTIST collaborators and our nine advisors, we would never have been able to develop such a good website, test bank, and instruments. I have found that even with all the work, advice, and revision, there are still ways assessment items and instruments can all be improved, and that the work is never done. Nevertheless, we are pleased with the positive responses we have received about ARTIST products and hope to see many people use our resources and tests in future projects or as part of their student or course assessments.

My collaboration with Bob, which began with my first assessment project, has continued thought the rest of my career. We shared common interests but different strengths, and together we were creative and productive. When I had an idea, Bob would always find a way to make it happen either by designing a software tool or developing an activity. He was the driving force behind our ARTIST Website and Assessment Builder (online data base to create tests and produce reports). He constantly devises new ways to meet the needs of ARTIST users. I might take the lead in writing up our work, while Bob takes the lead in analyzing data and thinking of new ways to examine our results. It is now hard to imagine working on any project without bringing Bob on board to provide his thoughts, skills, and partnership.

WHAT I LEARNED ABOUT ASSESSMENT IN STATISTICS EDUCATION

Assessment is an extremely important part of a good research study and good assessments are needed to conduct research studies that can be extended, replicated, or used for comparisons. There are no perfect instruments, and in fact, many are flawed. But using assessments that have been carefully developed and validated is important and can provide valuable information and insight. Using an instructor designed instrument such as a final exam makes it difficult for readers or researchers to tell if the instrument is measuring something of importance. For example, many final exams include more computational and definitional items than items that assess statistical reasoning and thinking, and are often not in alignment with stated goals of the course

Our work with the SRA indicated that while it was reasonably reliable in a test-retest analysis, student responses were not internally consistent, suggesting that students reason differently about different concepts, and that there is not one construct of statistical reasoning. This inconsistency in reasoning supports earlier findings by Konold *et al*. (1993).The comparative studies using the SRA revealed strikingly similar patterns of common difficulties students have reasoning about statistical concepts, across all countries and setting.

I have learned that it is important to study student responses to more than one item at a time, to see if correct answers are stable or if response patterns reveal interesting misconceptions (e.g., Konold, 1995) and that it is possible to develop good forced choice items that probe and assess students' statistical reasoning and thinking (see Cobb, 1997).

It is very difficult to write a good assessment item or task that is valid for a particular purpose. There is a great need for more work in developing good instruments to use for research and evaluation purposes. In the process of collecting, writing, and continually modifying assessment items, we have been struck by how one or two good items can reveal some important misunderstandings in student learning (delMas *et al*., 2005). For example, an item from the CAOS test, adapted from an item provided by Marsha Lovett, compares two boxplots which represent final exam scores for students in different sections of a course, showing the same median score of 80. Students are asked "Which section has a greater percent of scores at or above 80?"Although students may understand how to find the median, or can report the definition, only 27% of over a thousand students, tested at schools across the country, selected the correct answer to this problem. Most students erroneously picked the box plot that showed a whisker that extended farther to the right.

LESSONS LEARNED ABOUT DOING RESEARCH

- There seem to be three types of approaches that researchers can take regarding the use of assessment in their studies. The first (measurement optimist) is to develop a new instrument or use an existing test (e.g., a final exam) without carefully examining the qualities of the test or determining its validity. This is a type of optimism, believing that the test measures what you think it should and what you value. The second approach (measurement adopter) is to use an existing instrument in a study because it has already been developed, has been written about, and may relate to the intended outcomes of a particular study. The third approach (measurement developer) is to carefully define what the desired student outcomes are and to develop a new instrument, which may include items or modifications of previous instruments. The goal is to develop a high quality, valid and reliable instrument that will be useful beyond one particular study.
- There is great interest in the results of the third approach. If you take time to develop an instrument, and write about it and disseminate it, others will want to use it. I have also learned that if you develop and disseminate an instrument, that may have some flaws and limitations, it may continue to be used and used until something new and better comes along.

- Developing a good assessment for a research study takes time, much revision, and much feedback from others. And despite all one's best efforts, it will never be a perfect instrument. The collaboration (and resulting revision) processes are especially crucial in making sure you are seeing the whole picture and getting lots of different perspectives, including students'. It can seem deceptively easy when considering what the question is asking and anticipating student responses.
- While statisticians know a great deal about data and data analysis, they often have not studied principles of educational or psychological measurement, yet another reason to collaborate with a colleague who has this expertise.

THIRD STORY: STUDYING THE ROLE OF TECHNOLOGICAL TOOLS IN DEVELOPING STUDENTS' REASONING ABOUT SPECIFIC CONCEPTS

My first collaboration with Bob delMas, described in Story 2, involved an extension of his doctoral dissertation, which examined students' misconceptions about probability. He developed the Coin Toss program, which presented visual images of a coin being tossed and generated simulated data for different sequences of coin tosses. With a small internal grant, we developed an activity to guide students to make and test predictions about different coin toss results, and gathered student data using assessment items described above. This work was presented at conference (e.g., Garfield and delMas, 1991) but never written up for a formal publication because we were constantly making changes to the program and activities, and soon branched off in a new and related area.

Soon after that time, I was teaching an introductory course to graduate students and was perplexed because despite my explanations and lengthy class discussion, the students did not understand the idea of statistical power. Bob and I decided to study the idea of statistical power and any research that would help us to understand students' difficultly with this concept. We formed a "Power Study Group" and found articles such as "Belief in the Law of Small Numbers" (Tversky and Kahneman, 1971) to be very illuminating. Bob developed a new simulation tool called the Power Simulator, and we developed an activity to guide students from informal conceptions of statistical power to formal ones (Garfield and delMas, 1994), but once again, we went on to related work rather than polishing and publishing this paper.

Bob next developed simulation tools for the normal and other statistical distributions, and then a sampling simulation program (now known as Sampling Sim), in response to our feeling that students needed to see and interact with some good tools to help them visualize abstract statistical concepts. We narrowed our focus to exploring ways to help students visualize and reason about sampling distributions and the Central Limit Theorem. This work established a second connection with Beth Chance who soon became a collaborator on a new NSF Grant we received to extend this work (Tools for Teaching and Assessing Statistical Inference, at http://www.gen.umn.edu/research/stat_tools/). This project produced software, activities, and assessments designed to help students develop an understanding of sampling, confidence intervals, and p-values. Although the grant was to develop materials, we began a collaborative classroom-based research project to study the effectiveness of the tools and activities in our own, very different, introductory statistics courses (delMas *et al.*, 1999). These studies included analyses of assessment data and videotaped student interviews, and led to a tentative model of statistical reasoning about sampling distributions (see Chance *et al.*, 2004). It also led us to explore foundational concepts that students need to understand before they encounter sampling distributions such as distribution, center, and variability. Soon we were involved in a second, larger collaboration that examined students' informal and formal conceptions of variability (Garfield *et al.*, in press). We used Japanese Lesson Study (JLS) as our research method. This worked well as we developed, taught, observed, and evaluated a set of collaboratively designed lessons that utilized technology to help students develop their reasoning about variability. This collaboration involved both experienced and novice teachers (graduate teaching assistants) and led to the formation of a second JLS group at a nearby college.

My earlier work with technology also led to an important new collaborator: Dani Ben-Zvi, whom I met at the 1996 IASE Roundtable Conference on Research on the Role of Technology in Teaching and Learning Statistics (Garfield and Burrill, 1997). Dani and I soon found that we had many interests in common, and both wanted to find a better mechanism for researchers to share their work in a way that would allow richer discussion and feedback from other researchers. This led to the first International Research Forum on Statistical Reasoning, Thinking and Literacy in 1999, a successful forum that brought together a small group of researchers to share and discuss their work.

This unique research forum has been held every two years since that time, in a different part of the world (http://www.stat.auckland.ac.nz/srtl4/index.html) and led to the publication of an edited book (Ben-Zvi and Garfield, 2004) and two special issue of *SERJ* (see Ben-Zvi and Garfield, 2004; Garfield and Ben-Zvi, 2005). A third special issue is currently in the works. In addition, this forum has connected researchers, both new and experienced, to form new collaborations and to focus new research around particular issues of importance (e.g., reasoning about variability). What I learned from Dani was that it is possible to make a dream come true if you can find someone to help put plans into actions. Without Dani's willingness to organize and handle details, SRTL would never have happened nor would our book have been published. I have also learned from my collaboration with Dani that it is better to take your time and do a good, careful job than to rush through a project and have it be full of mistakes.

WHAT I LEARNED ABOUT TEACHING AND LEARNING STATISTICS FROM OUR RESEARCH ON TECHNOLOGY AND STATISTICAL REASONING

I learned that even with the most clever and carefully designed technological tool, and a good activity, that students may still fail to correctly understand and reason about an abstract statistical concept. After 10 years of research on sampling distributions, using new and improved versions of the *Sampling Sim* software along with revised activities, several students still failed to understand the process behind a sampling distribution or did not retain what they had learned.

I learned that good tools need to be used and reused with carefully designed activities, giving the students enough time and repetition, to help build the concepts. I learned that assessment plays a crucial role in helping reveal to teachers and researchers what students actually learn and retain from these instructional activities. I also learned from our collaborative classroom research that students' difficulties in reasoning about statistical concepts appear across different courses and types of students, and are resistant to change.

I learned that studying student' understanding of difficult concepts, and discussing these concepts with colleagues as we plan and evaluate lessons and the use of technology tools, is a process that broadens and deepens our own knowledge of the concept. In this way, our own understanding of sampling distributions and variability is much deeper and complex and we have a new appreciation for why these concepts are so difficult for student to fully understand.

LESSONS LEARNED ABOUT DOING RESEARCH

- Collaborative classroom research is a helpful way to study a problem in different classroom settings.
- Japanese Lesson study works well in a college statistics setting. Collaboratively designing a lesson, testing it, observing, and revising it, can produce high quality lessons that are better than lessons any of the individuals could produce alone.
- It is important to help make sure new researchers are not working in isolation and to provide more venues for providing constructive feedback to researchers during the research process.
- Creating new research forums can benefit both novice and experienced researchers, can help move the field by promoting research focused on a particular field, and can help form productive new collaborations.
- It is worthwhile to reflect on our own strengths and weaknesses as a researcher and to find individuals who can help complement our skills.

SUMMARY

In each of the collaborations described in this paper I have experienced the thrill of creativity, energy, synergy, and collegiality that became a valued friendship. Whether studying the research literature, assessment, or technological tools and activities, the collaborative work has been fulfilling and I have learned many useful things from my collaborators and from our work together. I feel strongly that all the work I have done with others is far better than any work I have done on my own. I will conclude this paper with advice to new researchers who will continue to grown and improve our field and provide new knowledge to improve the teaching and learning of statistics.

ADVICE TO NEW RESEARCHERS

General Advice

- Do not work alone. Look for collaborators who share your research interests but who may bring a different background or discipline, as well as different strengths to a new collaboration. Try to allow your research group to be open to adding new and interested colleagues.
- When you are just starting out, actively seek a mentor who is as interested and passionate about the subject as you are, but is savvier about writing, grant getting, and publishing. Team up with the mentor and learn all you can from him or her. In statistics education this may mean someone other than your advisor, who may be in an area like mathematics or psychology and who may know little about teaching and learning statistics.
- As you develop your own research agenda, watch for bright young beginning researchers to work with, mentor, and collaborate with.
- Look for interesting questions that arise in teaching. If students have trouble with a topic, explore why.
- Keep a narrow focus. For example, rather than studying the broad question: “Does technology improve student learning” consider focusing on a narrow question, such as “How can a particular use of a specific technology tool help students understand the meaning of confidence interval?”
- Do not select a research method first (e.g., compare a new teaching method to a control group) but let the research question suggest a method. Often it is important to conduct one or more small, descriptive, exploratory studies to begin to understand the problem (e.g., how students understand and misunderstand sampling distributions) before designing a more complex study that tests the effectiveness of a particular intervention. Give yourself time to study, discuss, and informally explore the problem before designing a comparative study.
- Use a variety of methods to collect data. For example, gather quantitative assessment data along with qualitative data from student interviews or writing tasks. Use the qualitative data to help interpret the quantitative data. If your study includes qualitative data, use appropriate and rigorous methods to analyze these data (e.g., Miles and Huberman, 1994).
- When doing classroom research that involves a new intervention, implement and study it in one class, then replicate the intervention in other classes (form a collaborative classroom research group). Understand the limitations in your study design and consider methods of strengthening the results (e.g., triangulation and extensive documentation) before attempting to generalize the results.
- Form a study group to read and discuss the literature on a particular topic of interest.
- Find creative ways to fund research (e.g., a detailed evaluation component of a materials development grant).
- Borrow good ideas! Adapt from existing projects (and always credit the original developers) rather than trying to start from scratch in developing a new activity or technology tool to use in a new study.
- Publish your work while it is still fresh, and do not keep going on to extended or new projects before you have published some of your previous work.

The Importance of Good, Thorough Literature Review

- When beginning a research study related to teaching and learning statistics, look at the literature beyond your discipline. Today this is much easier, as one can take advantage of

current online resources (e.g., http://CAUSEweb.org research page), online journals, and the *Google Scholar* search option. There are also good, comprehensive summaries you may refer to as a starting point, such as those by Shaughnessy (1992, 2003) and Konold and Higgins (2003).
- Don't assume you have seen all the literature that is out there. Search all areas. Find someone who is doing good work in this area and contact them for additional advice and to help you find the important articles on this topic that you might have missed.
- When you are interested in a particular problem, first read as much as you can to learn what is already known, rather than add the literature review later on when you want to publish your results.
- If you write a thorough and high quality literature review, it may turn out to be a contribution in its own right, and may be published in a good journal.

Assessment
- Think carefully about the role of assessment in a study. Do not solely use a final exam or instructor-developed test as an outcome measure that will be analyzed and reported. Instead, try to use an existing instrument that has demonstrated quality and is aligned with important student learning outcomes.
- If an instrument does not exist that meets your need, then make the development of such an instrument part of your study, and do it right, which may initial consulting with a measurement expert and going through the necessary steps of instrument design and validation.
- Make sure that the instrument(s) you use are aligned with important learning outcomes, so that they assess what you and the statistics education community value about student learning.

Technology
- Do not overestimate the power of a good, clever tool. Pay careful attention to how it is used and how it affects student learning.
- Do not focus on whether one software tool is better than another, but rather, how effectively software tools may be used to develop students' statistical reasoning. Look out for features of the software that may interfere or confuse students.
- Do not feel that you have to build a new tool to use in a research study before you make sure that one does not already exist. A good place to search for current tools is http://CAUSEweb.org. The same is true for inventing a new activity. Look at and build on the best ones currently available.

FINAL REMARKS

I feel very fortunate to have worked with some bright, enthusiastic, and skilled collaborators. I have learned a tremendous amount from each of the people I have worked with and have benefited greatly from our collaborative projects. While working on the various projects described above, I also came in contact with graduate students working on dissertations in statistics education at other universities but with advisors in other disciplines (e.g., mathematics education). Many of these collaborations have turned into friendships and I enjoy watching these students develop into contributing members of the research community after they completed their doctoral degrees. Four years ago I was fortunate to finally work with my own doctoral students in statistics education, by starting a new graduate program in Statistics Education within my Department (Educational Psychology). These new collaborations have been rich and exciting, as I have encouraged my students to pursue studies in their own areas of interest and as a result, I have learned from them while helping them turn their questions into research studies. As part of this program I have developed a research seminar where I meet with my own students and other interested graduate students to read, discuss, and even attempt to synthesize the research literature in a particular area. In the Spring 2006 version of this seminar we are reviewing and synthesizing research studies of college students' statistical understanding, and plan to write a collaborative paper to share our work with the broader community. I look forward to seeing other Statistics Education graduate programs develop in coming years, and a new breed of researchers in statistics education develop who will enrich this field with their findings and whose insights will improve our knowledge of teaching and learning statistics.

REFERENCES

Ahlgren, A. and Garfield, J. (1991). Analysis of the Probability Curriculum. In R. Kapadia and M. Borovcnik (Eds.), *Chance Encounters: Probability in Education*, (pp. 107-134). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Batanero, C., Serrano, L. and Garfield, J. (1996). Heuristics and biases in secondary school students' reasoning about probability. In the *Proceedings of the International Meeting of the Psychology of Mathematics Education Meeting*, Valencia, Spain.

Ben-Zvi, D. and Garfield, J. (2004). Research about reasoning about variability: A forward. *Statistics Education Research Journal* 3(2), 4-6.

Ben-Zvi, D. and Garfield, J. (Eds.) (2004) *Challenges in Developing Statistical Reasoning, Thinking and Literacy.* Dordrecht, The Netherlands: Kluwer Publishers.

Chance, B., delMas, R. and Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, (pp. 295-324). Dordrecht, The Netherlands: Kluwer Publishers.

Chance, B. and Garfield, J. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal*, 1(2).

Cobb, G. W. (1998) The objective-format question in statistics: Dead horse, old bath water, or overlooked baby? Paper presented at the annual AERA meeting. Copy of paper may be found at: https://app.gen.umn.edu/artist/articlesauthor.htm.

delMas, R., Garfield, J., and Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3).

delMas, R., Garfield, J., and Ooms, A. (2005). Using assessment items to study students' difficulty with reading and interpreting graphical representations of distributions. Presented at the Fourth Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-4), July 6. Auckland, New Zealand.

Gal, I. and Garfield, J. (Eds.) (1997). *The Assessment Challenge in Statistics Education*. Amsterdam: IOS Press and the International Statistical Institute.

Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1).

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

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

Garfield, J. (1991). Evaluating students' understanding of statistics: Development of the statistical reasoning assessment. In *Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, Blacksburg, VA, Volume 2, (pp. 1-7).

Garfield, J. and Ahlgren, A. (1988). Difficulties in learning basic concepts in statistics: Implications for research. *Journal for Research in Mathematics Education*, 19, 44-63.

Garfield, J. and Ahlgren, A. (1986). A study of difficulties in learning probability and statistics. In R. Davidson and J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*, Victoria, B.C.: University of Victoria, (p. 270-274). Voorburg, The Netherlands: International Statistical Institute.

Garfield, J. and Ben-Zvi, D. (in preparation). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice.* Emeryville, CA: Key College Press.

Garfield, J. and Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92-99.

Garfield, J., Chance, B., and Snell, J. L. (2001). Technology in college statistics courses. In D. Holton *et al.* (Eds.), *The Teaching and Learning of Mathematics at University Level: An ICMI study*, (pp. 357-370). Dordrecht, The Netherlands: Kluwer Publishers.

Garfield, J. and Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematics Thinking and Learning, 2,* 99-125.

Garfield, J. and delMas, R. (1994). Informal and formal conceptions of statistical power. Paper presented at Fourth International Conference on Teaching Statistics, Marrakech, Morocco, July.

Garfield, J. and delMas, R. (1991). Students' conceptions of probability. In D. Vere-Jones, S. Carlyle and B. P. Dawkins (Eds.), *Proceedings of the Third International Conference on Teaching Statistics*, Dunedin, New Zealand, Vol. 1, (pp. 338-339). Voorburg, The Netherlands: International Statistical Institute.

Garfield, J. and delMas, R. (1989). Reasoning about chance events: Assessing and changing student's conceptions of probability. *Proceedings of the 11th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education,* Rutgers University, Volume 2, (pp. 189-195).

Garfield, J. and delMas, R. (1994). Students' informal and formal understanding of statistical power. Paper presented at the Fourth International Conference on Teaching Statistics, Marrakech, Morocco, July.

Garfield, J., delMas, R., and Chance, B. (in press). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. Lovett and P. Shah (Eds.), *Thinking about Data: 33rd Carnegie Symposium on Cognition*. Mahwah, NJ: Earlbaum.

Garfield, J. and Gal, I. (1999). Teaching and assessing statistical reasoning. In L. Stiff (Ed.), *Developing Mathematical Reasoning in Grades K-12: National Council Teachers of Mathematics 1999 Yearbook*, 207-219. Reston, VA: NCTM.

Garfield, J. and Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, 67, 1-12.

Garfield, J. and Green, D. (1988). Probability and statistics study group: Looking back and looking forward. *Teaching Statistics*, 10(2), 55-58.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1), http://www.amstat.org/publications/jse/v3n1/konold.html

Konold, C., and Higgins, T. L. (2003). Reasoning about data. In Jeremy Kilpatrick, W. Gary Martin, and Deborah Schifter (Eds.), *A Research Companion to Principles and Standards For School Mathematics*, (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Pollatsek, A., Well, A. D., Lohmeier, J., and Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, 24, 392-414.

Liu, H. J. and Garfield, J. (2002) Sex differences in statistical reasoning. *Bulletin of Educational Psychology*, 32(1), 123-138.

Miles, M. and Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* **.** Newbury, CA: Sage Publications Inc

Pollatsek, A., Lima, S., and Well, A. (1981). Concept or computation: Students' misconceptions of the mean. *Educational Studies in Mathematics*, 12, 191-204.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning*, (pp. 465-494). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, and D. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics*, (pp. 216-226). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M., Garfield, J. B., and Greer, B. (1997) Data handling. In A. Bishop *et al.* (Eds.), *International Handbook on Mathematics Education,* (pp. 205-237). Dordrecht, The Netherlands: Kluwer Publishers.

Tversky, A and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105-110.