

# SITUATING QUALITATIVE MODES OF INQUIRY WITHIN THE DISCIPLINE OF STATISTICS EDUCATION RESEARCH

RANDALL E. GROTH  
*Salisbury University*  
*regroth@salisbury.edu*

## ABSTRACT

*Qualitative methods have become common in statistics education research, but questions linger about their role in scholarship. Currently, influential policy documents lend credence to the notion that qualitative methods are inherently inferior to quantitative ones. In this paper, several of the questions about qualitative research raised in recent policy documents in the U.S. are examined. Each question is addressed by drawing upon examples from existing statistics education research. The examples illustrate that qualitative methods can be used profitably to study statistical teaching and learning, and that in some cases qualitative methods are preferable to quantitative ones. By using the examples presented, qualitative researchers in statistics education can begin to more strongly situate their work within scholarly discourse about empirical research.*

**Keywords:** *Qualitative research; Rigor; Scientific research; Generalizability; Exploratory studies; Subjectivity*

## 1. QUALITATIVE RESEARCH “THEN” AND “NOW”

Qualitative research gained a great deal of momentum in mathematics education during the 1980s and 1990s, and that momentum carried over to the field of statistics education. During those two decades, mathematics education researchers sought to provide an epistemological basis for qualitative research and refine methods for gathering qualitative data through strategies like interviews and classroom observations (Teppo, 1998). The two decades before the beginning of the twenty-first century could reasonably be called a “golden age” of qualitative research in mathematics education. In characterizing the influence that qualitative research gained, Silver (2004) went so far as to state, “One might argue that researchers in mathematics education have in recent decades erected a monument to qualitative research methods and nonexperimental modes of inquiry” (p. 154). As one browses through past issues of *SERJ*, the influence of the qualitative approaches employed and developed by mathematics educators is apparent in the use of data gathering methods such as clinical interviews (Canada, 2006) and modes of inquiry such as design experiments (Watson, 2007).

Although qualitative methods have helped the fields of mathematics and statistics education move forward, and are still employed in current studies, the usefulness of qualitative studies is still frequently debated in political and scholarly discourse. Questions about qualitative research that were addressed during its rise to prominence among mathematics education researchers have re-surfaced in recent policy documents

written by governmental agencies and scholars in the field. Four influential recent policy documents raising such questions are

- The *Procedures and Standards Handbook* of the What Works Clearinghouse (What Works Clearinghouse [WWC], 2008). WWC was established by the U.S. Department of Education to review and synthesize research and offer recommendations for practice.
- The National Mathematics Advisory Panel's *Foundations for Success* report (National Mathematics Advisory Panel, 2008). The Panel was established by presidential order, and was asked to investigate questions such as, "What is known from research about how children learn mathematics?" and "What is known about the effectiveness of instructional practices and materials?" (p. 1-1).
- The No Child Left Behind Act (2002). This piece of legislation called for "scientifically based research" to inform educational practices, motivating debate about the meaning of the term.
- *Using Statistics Effectively in Mathematics Education Research (SMER)*, Scheaffer & Smith, (2007). *SMER* was a joint effort among statisticians, statistics educators, and mathematics educators. It set forth a specific view about the elements an effective research program should contain.

The purpose of this paper is to highlight some of the questions raised about qualitative research in the four documents above and to address them by drawing upon examples from existing empirical literature in statistics education.

Although the four documents motivating this commentary are from the United States researchers in other countries will likely recognize similar issues within their own political or scholarly contexts. Questions raised (and often answered) by the four documents include the following:

- Is qualitative research scientific?
- Is qualitative research rigorous?
- Is qualitative research generalizable?
- Is qualitative research useful beyond exploratory studies?
- Is qualitative research objective?

Such questions are often raised, either explicitly or implicitly, in discourse that elevates quantitative methods for research above qualitative ones. Hence, formulating responses to each question is an important part of situating qualitative studies within any given area of educational research.

## **2. PHILOSOPHICAL ORIENTATIONS**

The manner in which an individual responds to questions about the usefulness of qualitative studies depends heavily upon his or her philosophical orientation toward research. Lankshear and Knobel (2004) identified two different approaches to demonstrating the soundness of qualitative research. The traditional approach is to try to show that qualitative studies can satisfy the quality control criteria to which quantitative studies are traditionally held. The second, more recent approach is to use a different, unique set of criteria. Elements of both approaches will be considered in the discussion of persistent questions about qualitative research, so each one is briefly summarized below.

The traditional approach to assessing qualitative research is often associated with *positivism*. Positivists assume that "physical and social worlds exist in a given way, and that through the rigorous and expert application of research procedures we would come to know these worlds more or less as they are" (Lankshear & Knobel, 2004, p. 361). Qualitative researchers embracing positivism generally seek to demonstrate the quality of

their studies by showing that they satisfy standards for various types of validity and reliability by which quantitative studies are judged. Traditional standards for judging quantitative studies include correct execution of prescribed data analysis techniques, production of statistically generalizable results, formulation of instruments that contain a minimum amount of ambiguity, production, and use of a replicable study design, and similarity in how members of a research team view a set of data. Guidebooks exist with suggestions on how to adapt these elements of quality control for use in qualitative studies (e.g., Miles & Huberman, 1994).

The second approach to assessing qualitative research rejects positivistic criteria as being overly simplistic and restrictive. Some who take this approach question the claim that “there is some fixed reality or truth that can be ‘discovered’ by means of investigation or measurement” (Lankshear & Knobel, 2004, p. 362). From this perspective, research ought to present coherent interpretations and claims about a given situation that are supported by a range of appropriate data. The resulting vivid descriptions that are produced are seen to represent a given researcher’s well-formulated perspective rather than an objective reality that cuts across a tightly specified range of contexts. Those embracing the newer approach often advocate the view that traditional standards for assessing quantitative research are insufficient for qualitative research because of its fluid and evolving nature. Simon (2004), for example, stated, “Analyzing qualitative data is much like solving a mathematics problem: each step leads to greater insight as to what might be tried subsequently” (p. 161). He also pointed out that initial research questions often necessarily change as interesting findings with the potential to move the field forward emerge from the data.

Because the two approaches to assessing qualitative research outlined above rest upon different ontological assumptions, it is not possible to identify a universally accepted framework for responding to charges that qualitative research is inherently inferior to quantitative. Some strategies for doing so reflect positivist assumptions, and others reject them. Qualitative researchers are invited to consider the examples and arguments below in tandem with their personal philosophical assumptions in order to form their own well-grounded approaches to communicating the nature and importance of their work.

### **3. ADDRESSING QUESTIONS ABOUT QUALITATIVE RESEARCH**

Questions about qualitative research have been raised both implicitly and explicitly in policy documents. Implicit attacks are perhaps more detrimental, as they can easily flow undetected into various discourse arenas. Some of the questions raised below are not explicitly stated in the policy documents to be examined, but are implied by the positions advocated in the documents. Left unexposed, such implicit assumptions can become accepted parts of doing research without being critically analyzed. This section begins by addressing the question of whether or not qualitative research is scientific, and then goes on to examine a variety of additional related questions implicit in current policy documents.

#### **3.1. IS QUALITATIVE RESEARCH SCIENTIFIC?**

Policy documents sometimes implicitly de-legitimize qualitative research by offering definitions for “scientific research” that automatically exclude it. One such definition comes from the WWC (2008):

The focus of the WWC is on scientifically-based evidence. Therefore, to be included in the WWC review, a study must use one of the following designs (described in the

later section on evidence standards): randomized controlled trial, quasi-experimental, regression discontinuity, or single subject. (not paginated)

Under the WWC definition, there is no room for discussion about whether or not qualitative research is scientific. Qualitative studies are excluded from the outset and the door for further dialogue is closed. The definition holds up randomized controlled trials as the gold standard in educational research.

There are some deep-seated problems, however, with using randomized controlled trials for educational research. Holcomb (2002) identified two major ethical concerns from a statistician's perspective: (i) evaluating (for grading purposes) a control group of students on skills or tasks that instruction had not prepared them to do; and (ii) teaching the control group in a manner the experimenter believes, from experience, is not as effective as the approach used with the treatment group (or vice versa). Schoenfeld (2007) raised methodological concerns. He warned that rushing toward randomized controlled trials in a research program is unlikely to produce information valuable for informing instruction. Simply knowing that an instructional intervention performed well in one study usually says little about whether or not it should be adopted in another particular setting. Contextual characteristics of the students, instructors, and broader communities in each setting need to be taken into account in order to make a reasonable judgment. Given these difficulties with using randomized controlled trials in education, those espousing a WWC-like definition of scientific research predictably draw the conclusion that much of educational research simply is not very scientific.

Before simply dismissing much of educational research as unscientific, however, it makes sense to re-examine the WWC definition and other definitions that automatically exclude qualitative research. The National Research Council (NRC, 2002), for instance, provided a much different description of the characteristics of scientific research:

A wide variety of legitimate scientific designs are available for education research. They range from randomized experiments of voucher programs to in-depth ethnographic case studies of teachers to neurocognitive investigations of number learning using positive emission tomography brain imaging. To be scientific, the design must allow direct, empirical investigation of an important question, account for the context in which the study is carried out, align with a conceptual framework, reflect careful and thorough reasoning, and disclose results to encourage debate in the scientific community. (p. 6)

The NRC definition is based upon examination of the common characteristics that exist among researchers using different methodologies and having different philosophical orientations. It is worth noting that rather than automatically categorizing qualitative research as "unscientific," the NRC actually included a predominantly qualitative method, ethnographic case study, as an example of a legitimate design for scientific research.

Although some qualitative researchers are uneasy with the NRC (2002) definition of scientific research because they perceive positivist leanings within the overall NRC framework (Ryan & Hood, 2004), the NRC definition at least opens the door for discussions about what qualitative research might contribute to the science of statistics education. Ethnographic case study, because it is explicitly acknowledged as a scientific design by the NRC, provides a good starting point for discussion. An ethnographic case study involves spending extended amounts of time at a research site, being personally involved in the activities of the case, and constructing descriptions and explanations of phenomena observed through reflection on empirical observations (Stake, 2005). Elements of ethnographic case study can be seen in Pfannkuch and Horring's (2004) research. In the study, the researchers sought to understand the existing practices for

teaching statistics that were in place at a school. The purpose for doing so was to help the school implement a new curriculum based on the notion of statistics as an investigative process (Wild & Pfannkuch, 1999). The researchers gathered data from a variety of sources, including existing curricular materials, interviews with teachers, observations of classrooms, and students' responses to assessment items. They found that the existing curriculum did not help students understand how the parts of the statistical investigation cycle relate to one another. Having identified problematic aspects of practice, they worked with teachers to develop and implement the new curriculum. When the new curriculum was put in place, the researchers continued to gather qualitative data from classroom observations and teacher interviews to identify and resolve problematic areas with implementation for two successive years. The research report Pfannkuch and Horning ultimately produced provided a portrait of the complexity of helping teachers put a new statistics curriculum in place and working with them to implement it successfully.

Pfannkuch and Horning's (2004) extended case study, which was primarily qualitative in nature, reflects a number of characteristics of scientific research as it is conceptualized by the NRC (2002). First, it involved direct, empirical observation of an important question. Many statistics curricula are not structured in a manner that helps students understand the investigative process of statistics (Shaughnessy, 2007). Hence, identifying ways to change existing statistics curricula to attain this ideal is paramount. Simply handing teachers a set of curriculum materials is not sufficient, as extensive support is needed if curricula are to be implemented in the manner intended by the writers (Lloyd & Behm, 2005). Such support was available for teachers as they participated in the study.

The Pfannkuch and Horning (2004) study also meets the criteria for scientific research on the grounds that it includes a rich description of the setting in which the study was carried out. Characteristics of the school were documented through a variety of qualitative data gathering techniques like interviews and classroom observations. This contextual information became important in helping explain how the new curriculum was designed and implemented. Other researchers and curriculum developers seeking to help improve statistics instruction can then benefit from comparing the details of the setting to those of the settings in which they work.

The Pfannkuch and Horning (2004) study also meets the criteria of being scientific in that a conceptual framework was employed. The Wild and Pfannkuch (1999) framework was used to characterize the type of statistical thinking the researchers wanted students and teachers to exhibit. The framework itself was formed by interviewing individuals engaged in statistical investigations, including professional statisticians. Hence, the framework had the potential to help draw the statistics discussed in the classroom closer to normative practices of statistical data analysis. It served as a guide to identify instructional practices that could be modified so the process of doing statistics would be more effectively portrayed for students in the classroom.

Finally, Pfannkuch and Horning's (2004) study meets the characteristics of scientific research in that it consists of a careful and thorough chain of reasoning that was presented to the larger statistics education community for discussion and debate. In addition to the report described above, smaller pieces of the research were presented in journal articles (e.g., Pfannkuch, 2006). Given all of the valuable characteristics of the research, it would be counterproductive and to call it "unscientific" simply because qualitative techniques were used extensively. Consider an alternative: To meet the WWC (2008) definition of scientific research, the researchers could have simply designed a curriculum that reflected the process of statistical investigation. They then could have set up control and experimental groups to see how well the curriculum "worked" along some desired dimension. Such a process would satisfy the WWC definition of "scientific" research.

However, in such a study, we would likely learn less about the school context, the reasons for the success or failure of the curriculum, and the dynamics of helping teachers implement unfamiliar, innovative curricula. Hence, such a study would ultimately be of less value to the study participants and to the statistics education research community. This suggests that narrow, restrictive WWC-like definitions of scientific research that exclude qualitative designs can actually lead to the production of lower-quality studies.

### **3.2. IS QUALITATIVE RESEARCH RIGOROUS?**

The word “rigorous” has been used in some policy documents to help define the phrase “scientific research.” For instance, The No Child Left Behind Act (2002) defined scientifically-based research as “research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs.” WWC (2008) later expanded upon this idea to formulate a definition of scientific research that excluded qualitative studies, as discussed earlier. Because “rigor” is considered to be a hallmark of scientific research, it is worthwhile to examine what might be meant by the term. The word “rigor” is used frequently in discussions of standards for research, but seldom defined formally (Davies & Dodd, 2002).

When definitions for rigor are given, they inevitably spark controversy. From the positivist perspective, it is often asserted that there are universal standards for reliability and validity that ought to apply to all research, but many qualitative researchers believe that such a universal set of standards is not feasible (Golafshani, 2003). The very nature of qualitative inquiry gives rigor a different feel in qualitative studies, even if one characterizes rigor in terms of reliability and validity. Morse, Barrett, Mayan, Olson, and Spiers (2002) described the process of establishing rigor in qualitative research in the following manner:

Qualitative research is iterative rather than linear, so that a good qualitative researcher moves back and forth between design and implementation to ensure congruence among question formulation, literature, recruitment, data collection strategies, and analysis. Data are systematically checked, focus is maintained, and the fit of data and the conceptual work of analysis and interpretation are monitored and confirmed constantly. Verification strategies help the researcher identify when to continue, stop, or modify the research process in order to achieve reliability and validity and ensure rigor. (p. 17)

The Morse et al. perspective on rigor holds elements appealing to those holding a positivist paradigm as well as those rejecting it: The perspective acknowledges the fluid, contextually situated nature of qualitative inquiry while still asserting the importance of validity and reliability in some form.

Current efforts to establish rigor in statistics and mathematics education research studies resonate strongly with the Morse et al. (2002) perspective on rigor. As an illustration, consider the “teaching experiment” methodology (Steffe & Thompson, 2000) that has become a frequently used mode of inquiry. Despite the use of the word “experiment,” this methodology does not involve statistical comparison of randomly formed control and treatment groups. Instead, it is intended to be a tool for describing the intricacies and challenges inherent in teaching concepts in a classroom setting. At the outset of a teaching experiment, researchers form a tentative plan, or “hypothetical learning trajectory” (Simon, 1995) describing the path classroom instruction should take. Then, the original plan is implemented in a classroom setting over a period of several days or weeks. Careful observations are made about how effective the original plan is in

helping students attain learning goals each day. These observations are used to fine-tune the original plan to be more responsive to the circumstances of the classroom. At the conclusion of a teaching experiment, written accounts describe critical decisions made by researchers during the study, and how their choices influenced students' learning.

Cobb (1999) illustrated how the teaching experiment methodology can be applied to statistics education research. He described a teaching experiment that had the overarching goal of developing middle school students' abilities to "view data sets as entities that are distributed within a possible space of values" (p. 10). On the basis of existing literature, it was conjectured that helping students reason about parts of a data set as proportions of the whole would be key to the endeavor. Before the experiment began, pre-assessments indicated that the students to be involved in the study considered data analysis simply to be "doing something with numbers" (p. 12) rather than an investigative process. Hence, engaging students in the sort of argumentation prevalent in the field of statistics was another desired outcome of the experiment.

At the outset of the teaching experiment described by Cobb (1999), students generally did not listen to or challenge their classmates' analyses of data. Instead, they performed exercises they had previously done in statistics, like computing the mean for every data set, regardless of the question at hand. Students began to engage in argumentation when a computer "mini-tool" was introduced that facilitated organizing data with informal representations. Once conversations among students opened up, the researchers acted upon opportunities that occurred for steering discourse toward normative modes of argumentation. Students' comments during the conversations revealed that many did not use proportional reasoning in analyzing data sets. This motivated the creation of a second computer mini-tool and selection of data sets that would provide opportunities to move students toward developing strategies for comparing unequal-sized groups. Throughout the teaching experiment, instructional strategies were adjusted as more was learned about students' statistical understanding and modes of statistical argumentation.

Given the constant changes in the research procedures in the teaching experiment described by Cobb (1999), should the study be called "rigorous?" After all, if similar shifts in instruction had occurred in the course of an experiment (in the statistical sense) measuring the effectiveness of a given teaching intervention, we might say that the study design had been corrupted. To make a judgment about whether or not the "rigor" label applies, it must be recognized that the purposes of teaching experiments and statistical experiments differ. The randomized controlled trials advocated in some policy documents (e.g., National Mathematics Advisory Panel, 2008; WWC, 2008) are generally done for the purpose of comparing the effectiveness of one curriculum against another. Steffe and Thompson (2000) described a much different purpose for conducting teaching experiments:

A primary purpose for using teaching experiment methodology is for researchers to experience, firsthand, students' mathematical learning and reasoning. Without the experiences afforded by teaching, there would be no basis for coming to understand the powerful mathematical concepts and operations students construct or even for suspecting that these concepts and operations may be distinctly different from those of the researchers. (pp. 267–268)

The models of students' mathematical learning and reasoning that are constructed from teaching experiments may ultimately be used as the basis for curriculum design, which one may desire to test later on using a statistical experiment, but the models themselves are necessarily in flux throughout the teaching experiment. Indeed, the fundamental tenet that teaching experiments rigorously adhere to is that the model *must* remain in flux throughout the study so that a more accurate account of students' difficulties in learning a

given content area is constructed as qualitative data on students' responses to learning interventions are obtained. The original model is regarded as being continually revisable as more data describing students' learning during the experiments are gathered and analyzed. In Cobb's case, adjusting teaching methods throughout the course of the study produced a model of the students' progression toward reasoning proportionally with data sets and engaging in normative modes of statistical argumentation. Adhering rigorously to the original plan for instruction would not have allowed teaching episodes to be tailored to meet the student learning needs that emerged during data analysis, and subsequently would not have allowed for the production of a model of students' cognitive progression for use as a starting point in future research and curriculum development.

### **3.3. IS QUALITATIVE RESEARCH GENERALIZABLE?**

Even if one agrees that the rich models and descriptions provided by qualitative studies are valuable to the statistics education community in some manner, some maintain that such models and descriptions are not generalizable. This appears to be a primary motivation for the positivist-oriented push to conduct randomized, controlled trials in education (as advocated by the What Works Clearinghouse, 2008). For instance, the National Mathematics Advisory Panel (2008) listed "Random assignment to conditions; low attrition; valid and reliable measures" (p. 2–3) as a summary of its criteria for "high quality" studies of the effects of interventions. The push for randomized, controlled trials, however, overlooks the fragility underlying the potential applicability of the findings from such studies. Rigorously applying experimental methods does not dismiss the researcher from providing information about contextual variables that exert influence on the results. Schools are replete with variables that differ greatly from one classroom to the next, including teacher quality, students' background knowledge, demographics, gender distribution, time of day a subject is taught, curriculum sequence, students' behavior, and additional influential variables that are very difficult to anticipate and control. Substantial differences along any combination of these variables can greatly influence the applicability of findings from a statistical experiment to any particular classroom setting. Cronbach (1975) summed up the circumstances of the quantitative social sciences researcher by stating, "When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion" (p. 125).

If one accepts Cronbach's (1975) "working hypothesis" characterization of quantitative findings, it follows that qualitative and quantitative social science researchers are on similar footing when it comes to the generalizability issue. This is not to say that qualitative research is in some sense statistically generalizable, as positivists may demand before considering it scientific and useful. Rather, it means that much of the burden of determining the usefulness of findings, qualitative or quantitative, lies with the individual trying to apply them to his or her own particular situation. Stake and Trumbull (1982) called the process of reconciling one's particular setting with vicarious experiences of other settings provided by reading research one of "naturalistic generalization." When it comes to making naturalistic generalizations, qualitative studies that include rich descriptions of the context in which a study occurred and its findings can be quite helpful. Such descriptions allow the reader to do detailed comparisons between his or her own particular situation and the setting in which the study occurred. Stake and Trumbull attributed much of the failure of educational research to influence practice to lack of recognition of the importance naturalistic generalization plays in practitioner's day-to-day decision-making processes.

As an example of the importance of naturalistic generalization to statistics education, consider the exploration of the nature of statistical thinking undertaken by Wild and Pfannkuch (1999). By drawing upon existing literature and the results of qualitative studies that involved interviews with small samples of students and practicing statisticians, they proposed a four-dimensional framework for statistical thinking in empirical enquiry. The four dimensions of the framework included an investigative cycle, types of thinking, an interrogative cycle, and dispositions. The framework resonated with many in the statistics education community who used it to inform their own research on statistical thinking (Ben-Zvi & Garfield, 2004).

What is one to make of the research community's acceptance of Wild and Pfannkuch's (1999) characterization of statistical thinking? Is the statistics education community misguided in utilizing a framework that is based on qualitative interviews from a small sample of individuals? Should researchers demand a quantitative study with a larger sample size before using the findings to frame their own research? The preceding questions will seem somewhat odd to those who recognize Wild and Pfannkuch's work as an elegant expression of the nature of statistical thinking that had been largely tacit up until its publication. Those who have engaged in the process of statistical enquiry described by the framework simply recognize high degrees of resonance with their own experiences. In other words, the Wild and Pfannkuch model appears to have catalyzed the process of naturalistic generalization among the statistics education community. Although some may question aspects of the model (as readers often do with quantitative models), it is generally recognized that it provides a means to help move discourse forward about what statistical thinking is and how it can be developed. This example from the statistics education research community illustrates that sometimes naturalistic generalizations, rather than statistical generalizations, are instrumental in the advancement of a field. Hence, rather than dismissing qualitative studies on the charge that they are not statistically generalizable, it is usually more productive to consider their potential to foster naturalistic generalizations that can inform research and practice.

### **3.4. IS QUALITATIVE RESEARCH USEFUL BEYOND EXPLORATORY STUDIES?**

Even some of those who staunchly maintain that qualitative studies are not in any sense generalizable can endorse the idea that insights gained from qualitative studies can be useful in framing larger studies. Campbell (as cited in Miles & Huberman, 1994) remarked that all research has some sort of qualitative grounding. Some quantitative researchers recommend using findings from qualitative interviews to refine written questions on surveys to be given to larger samples that are statistically representative of a given population (Rossi, Wright, & Anderson, 1983). Because of their flexible nature, qualitative studies can help researchers identify variables and critical issues and test instruments and research strategies in the field before undertaking a larger study. In these ways, qualitative, exploratory studies can serve the function of informing the construction of larger quantitative ones.

The *SMER* document (Scheaffer & Smith, 2007) acknowledged the importance of qualitative studies in setting up larger quantitative studies in giving recommendations for the components of a strong research program. It stated that, "good quantitative studies generally require a qualitative rationale" (p. 43). According to *SMER*, early stages in a research program were to include generating ideas and questions, framing the research, and examining it. Qualitative studies such as teaching experiments were portrayed as valuable for informing these early stages. Later stages, however, in the development of a

mature research program were to involve generalizing (in the statistical sense) and extending the program. Setting statistical generalizability to be the ultimate goal of a research program essentially relegates qualitative research to the role of being a servant for quantitative studies.

Given the acknowledged utility of exploratory qualitative studies for improving the quality of larger quantitative studies, it is important not to overlook that quantitative research at times actually catalyzes qualitative investigations. For example, Batanero, Estepa, Godino, and Green (1996) conducted a study with both quantitative and qualitative aspects investigating students' understanding of contingency tables. The quantitative portion of the study revealed that students had substantial difficulty interpreting some types of tasks involving contingency tables. The qualitative portion of the study provided an in-depth explanation of the causes of the difficulties by examining students' reasoning strategies. One of the two-way tables students were asked to interpret suggested that there was no relationship between bronchial disease and smoking. When presented with the bronchial disease and smoking task data, 17–18 year old students gave several different types of responses. Some completely disregarded the data and opted instead to rely upon previous theories about the dependence of illnesses on smoking. Other students based their judgment on the fact that there were more smokers in the sample with bronchial disease than non-smokers. Although this was true, it neglected the fact that the percentage of smokers in the sample with bronchial disease was the same as the percentage of non-smokers with the illness. Still others based their judgment on examining just one cell of the table (e.g., observing a high number of smokers with the disease). The various different responses to the task illustrated how contextual knowledge and proportional reasoning (or lack thereof) can interact to influence interpretations of data. This insight, gained qualitatively, helped explain patterns of incorrect reasoning underlying the quantitative results.

In addition to being a pre-cursor or follow-up to a quantitative study, there are also times when qualitative studies are valuable when not linked in any fashion to quantitative data. In some cases, qualitative studies, on their own, help challenge or modify existing theory. One developmental theory often used in statistics education research is the Structure of the Observed Learning Outcome (SOLO) Taxonomy (Biggs & Collis, 1982, 1991). Early formulations of SOLO posited that students' responses to academic tasks can be characterized in terms of three levels: unistructural, multistructural, and relational. The three levels were believed to be nested within modes of development (e.g., iconic, concrete symbolic, formal). Early theorists held that each mode of development contained one unistructural-multistructural-relational (UMR) cycle. Watson, Collis, Callingham, and Moritz (1995) applied the SOLO model to the analysis of students' responses to a task that involved examining and organizing a set of data cards showing characteristics of a hypothetical group of students. Their revised model, positing the possible existence of multiple UMR cycles within a mode of development (one for development of concepts and another for application), helped broaden the scope of subsequent qualitative analyses of students' responses to statistical thinking tasks (e.g., Groth, 2003; Reading & Reid, 2006).

### **3.5. IS QUALITATIVE RESEARCH OBJECTIVE?**

Even those who acknowledge that there is some value in qualitative research may still favor quantitative research on the grounds that it is more objective. The *SMER Report*, for example, emphasized the importance of “measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple

measurements and observations, and across studies by the same or different investigators” (Scheaffer & Smith, 2007, p. 4). Presumably, this point of view, based on the positivist assumption that it is possible to access and measure an objective reality, fed into *SMER*’s recommendation to relegate qualitative studies mainly to the preliminary stages in the development of a research program.

From a qualitative researcher’s point of view, the idea that qualitative research is more subjective than quantitative research can be a little difficult to understand, as all forms of research involve subjectivity. When designing a quantitative intervention study, for example, researchers must make decisions on a number of measurement issues, such as the nature of the treatments, the context for the study, and the particular assessment measures to be used (Scheaffer & Smith, 2007). These subjective decisions, made at the outset of a study, largely determine the types of conclusions and interpretations that can ultimately be drawn. Parallel decisions need to be made during a qualitative study, although the decisions are likely to be re-visited as a qualitative study progresses. Re-visiting framing decisions and revising them certainly is a subjective process. However, this sort of decision making is often vital in maintaining the scientific integrity of a qualitative study, as described earlier for teaching experiment and ethnographic case study methodologies.

Some of the uneasiness with subjectivity in qualitative research comes from the methods of data analysis it employs. Whereas methods for quantitative data analysis are highly codified and simple to state explicitly, qualitative data analysis techniques can vary greatly. For instance, the phrase “grounded theory” has a variety of shades of meaning for different qualitative researchers (Corbin & Strauss, 2008), but there is no ambiguity about what a researcher means when he or she states that a two-sample *t* test was used to analyze data. Qualitative researchers, recognizing this stark contrast between quantitative and qualitative methods of data analysis, have addressed it in different ways. The different ways of addressing the subjectivity issue are influenced strongly by whether or not one accepts positivist ontological assumptions.

One response to the perceived ambiguity in qualitative data analysis has been to work on making qualitative techniques more uniform, accessible to scrutiny, and objective in appearance. Miles and Huberman (1994), for instance, published a sourcebook dedicated to the cause of describing qualitative methods of analysis that are “practical, communicable, and non-self-deluding” (p. 1). Among the techniques they described were constructing specific types of matrices and causal networks to display findings. They also discussed methods for testing or confirming interpretations, such as examining rival explanations and getting feedback from study participants. Such methods align closely with the ontology of positivism, as an implicit assumption is that by considering several different opinions, researchers can eventually converge on the most “accurate” description of a given situation.

The influence of Miles and Huberman’s (1994) work can be seen in statistics education research studies. Watson and Moritz (2001), for example, used coding procedures described by Miles and Huberman to identify levels of thinking involved in understanding pictographs. They began their analysis by grouping students’ responses into clusters corresponding to levels in the SOLO Taxonomy. The two authors then compared the clusters they formed independently with one another. In their final report, they documented the percentage of agreement in categorization, as well as reasons for disagreement in coding responses. Disagreements between the researchers on how to code some segments of qualitative data led to critical conversations and refinements of the categorization scheme ultimately presented in the final manuscript. Several other research studies in statistics education have incorporated similar procedures to attempt to

attain a greater degree of objectivity in categorization schemes based on qualitative data (e.g., Jones et al., 2000; Mooney, 2002).

Although some researchers have sought to counteract the subjectivity of the interpretation of qualitative data by using procedures like those described above, others, rejecting positivist ontology, have embraced subjectivity as an essential aspect of qualitative research. Denzin and Lincoln (2005) are perhaps the strongest proponents of the idea that “Objective reality can never be captured. We know a thing only through its representations” (p. 5). Denzin and Lincoln portrayed the qualitative researcher as a bricoleur. A bricoleur can be conceptualized as one who pieces images on film into montages. A connection to research is that a qualitative study, like a montage, represents an individual’s unique way of representing phenomena of interest. Under this metaphor, no montage is necessarily more accurate than another. Rather, each one is a different attempt to convey the aspects of a situation that are perceived to be of importance. Researchers working independently with the same body of data may notice some of the same things about a situation, but their past experiences will lead to them to emphasize some aspects of the situation more strongly than others. Giving voice to multiple different interpretations of a given body of data thus becomes a way to learn about it in more depth. Subjectivity thus becomes a strong point, rather than a weakness, of qualitative research by providing several different perspectives about the data.

Studies that provide multiple interpretations of sets of data are not yet prevalent in statistics education research. There are, however, models of this mode of qualitative inquiry in mathematics education that could be appropriated and extended by the statistics education research community. One pertinent example is the 14<sup>th</sup> monograph of the *Journal for Research in Mathematics Education (JRME)* (Schoenfeld, 2008). In the monograph, several researchers provided analyses of a single six-minute video clip from a third-grade class taught by Deborah Ball. Each analysis drew upon the unique research background of the commentator writing it, and therefore emphasized different aspects of the situation, such as teacher decision-making, classroom argumentation, and equity concerns. At times, interpretations in the monograph conflicted with one another. For instance, an uncomfortable exchange between two students during the lesson was interpreted by one researcher as stemming from the teacher’s lack of attention to equity issues, and it was interpreted by others as stemming from irreconcilable mathematical definitions held by the students (Lewis, 2008). The actual video clip under analysis was included on a CD with the monograph so readers could compare the analyses against the original empirical qualitative data. Ultimately, such innovative uses of video technologies may be the most powerful means for showing that the subjectivity involved in qualitative analyses can be a boon to the field of statistics education.

#### 4. CONCLUSION

Electronic journals like *SERJ* may be particularly well-suited to make qualitative research analyses like those in *JRME* Monograph 14 prevalent in statistics education. An online format offers a means for displaying qualitative data stored in video files, audio files, and screen capture movies of students’ interaction with dynamic statistics software. Posting such files online and downloading them has become much less technically prohibitive within the past few years. The online format also offers a means for readers to compare and contrast their own interpretations with one another, because published articles can be accompanied by discussion boards and blogs inviting readers to offer their own comments. This sort of reader commentary feature is already incorporated in other online journals such as *Contemporary Issues in Technology and Teacher Education*. The

parallel development of qualitative research techniques and online technologies thus has the potential to advance qualitative research in statistics education in powerful ways.

The value of qualitative research in advancing the field of statistics education should not be underestimated. It provides a means for conducting scientific studies that contribute to our understanding of statistics teaching and learning. The rich descriptions provided by qualitative research are often instrumental in helping readers understand the relationship between the context of a given study and their own circumstances. Insights gained from qualitative research may be helpful in framing quantitative studies, but their potential value extends further. They can also help explain the causes of patterns observed in quantitative findings, as well as modify and extend existing theories. In recent years, some researchers have sought to reduce the ambiguity associated with qualitative techniques by developing and applying more highly codified methods of analysis. Others, rejecting positivist assumptions, have embraced the subjectivity inherent in qualitative analysis as a means to generate multiple representations of a given phenomenon, and thus come to a more nuanced understanding of it. Given the potential of qualitative research to contribute to the field, future discussions would do well to avoid elevating either qualitative or quantitative research above the other in favor of simply focusing upon what each approach can contribute to our collective understanding about the teaching and learning of statistics.

## REFERENCES

- Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151–169.
- Ben-Zvi, D., & Garfield, J. (Eds.) (2004). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, The Netherlands: Kluwer.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic.
- Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and quality of intelligent behavior. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57–66). Hillsdale, NJ: Erlbaum.
- Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal*, 5(1), 36–63.  
[Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ\\_5%281%29\\_Canada.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ_5%281%29_Canada.pdf) ]
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30(2), 116–127.
- Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health Research*, 12(2), 279–289.
- Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3<sup>rd</sup> ed.) (pp. 1–32). Thousand Oaks, CA: Sage.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597–607.
- Groth, R. E. (2003). High school students' levels of thinking in regard to statistical study design. *Mathematics Education Research Journal*, 15(3), 252–269.

- Holcomb, J. (2002). The ethics of comparison: A statistician wrestles with the orthodoxy of a control group. In P. Hutchings (Ed.), *Ethics of inquiry: Issues in the scholarship of teaching and learning* (pp. 19–26). Menlo Park, CA: Carnegie Foundation for the Advancement of Teaching.
- Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning*, 2(4), 269–307.
- Lankshear, C., & Knobel, M. (2004). *A handbook for teacher research: From design to implementation*. Berkshire: Open University Press.
- Lewis, J. (2008). Through the looking glass: A study of teaching. In A. H. Schoenfeld (Ed.), *A study of teaching: Multiple lenses, multiple views. Journal for Research in Mathematics Education, Monograph Number 14* (pp. 1–12). Reston, VA: National Council of Teachers of Mathematics.
- Lloyd, G. M., & Behm, S. L. (2005). Preservice elementary teachers' analysis of mathematics instructional materials. *Action in Teacher Education*, 26(4), 48–62.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning*, 4(1), 23–63.
- Morse, J. M., Barrett, J. M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13–22.
- National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.  
[Online: <http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf> ]
- National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- No Child Left Behind Act of 2001. (2002). Pub. L, No. 107–110, 115 Stat. 1425.
- Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27–45.  
[Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ5%282%29\\_Pfannkuch.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ5%282%29_Pfannkuch.pdf) ]
- Pfannkuch, M., & Horring, J. (2004). Developing statistical thinking in secondary school: A collaborative curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistics Education 2004 Roundtable*, (pp. 163–173). Voorburg, The Netherlands: International Statistical Institute.  
[Online: [www.stat.auckland.ac.nz/~iase/publications/rt04/5.1\\_Pfannkuch&Horring.pdf](http://www.stat.auckland.ac.nz/~iase/publications/rt04/5.1_Pfannkuch&Horring.pdf) ]
- Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal*, 5(2), 46–68.  
[Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ5%282%29\\_Reading\\_Reid.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ5%282%29_Reading_Reid.pdf) ]
- Rossi, P. H., Wright, J. D., & Anderson, A. B. (Eds.). (1983). *Handbook of survey research*. New York: Academic Press.
- Ryan, K. E., & Hood, L. K. (2004). Guarding the castle and opening the gates. *Qualitative Inquiry*, 10(1), 79–95.
- Scheaffer, R., & Smith, W. B. (2007). *Using statistics effectively in mathematics education research: A report from a series of workshops organized by the American Statistical Association with funding from the National Science Foundation*. Alexandria, VA: American Statistical Association.  
[Online: <http://www.amstat.org/education/pdfs/UsingStatisticsEffectivelyinMathEdResearch.pdf> ]

- Schoenfeld, A. H. (2007). Method. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 69–107). Charlotte, NC: Information Age Publishing.
- Schoenfeld, A. H. (Ed.) (2008). *A study of teaching: Multiple lenses, multiple views. Journal for Research in Mathematics Education, Monograph 14*. Reston, VA: National Council of Teachers of Mathematics.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Charlotte, NC: Information Age Publishing.
- Silver, E. S. (2004) Ella minnow pea: An allegory for our times? *Journal for Research in Mathematics Education, 35*(3), 154–156.
- Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145.
- Simon, M. A. (2004). Raising issues of quality in mathematics education research. *Journal for Research in Mathematics Education, 35*(3), 157–163.
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3<sup>rd</sup> ed.), pp. 443–466. Thousand Oaks, CA: Sage.
- Stake, R. E., & Trumbull, D. J. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science, 7*(1/2), 1–12.
- Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267306). Mahwah, NJ: Erlbaum.
- Teppo, A. R. (Ed.). (1998). Qualitative research methods in mathematics education. *Journal for Research in Mathematics Education, Monograph 9*. Reston, VA: National Council of Teachers of Mathematics.
- Watson, J. M. (2007). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal, 7*(2), 59–82.  
[Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ7%282%29\\_Watson.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ7%282%29_Watson.pdf) ]
- Watson, J. M., Collis, K. F., Callingham, R. A., & Moritz, J. B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation, 1*(3), 247–275.
- Watson, J. M., & Moritz, J. B. (2001). Development of reasoning associated with pictographs: Representing, interpreting, and predicting. *Educational Studies in Mathematics, 48*(1), 47–81.
- What Works Clearinghouse. (2008). *Procedures and standards handbook*. Washington, DC: U.S. Department of Education Institute of Education Sciences.  
[Online: <http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocId=3> ]
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265.

RANDALL E. GROTH  
Salisbury University  
Department of Education Specialties  
1101 Camden Ave.  
Salisbury, MD 21801