

## Standards of Evidence in Qualitative Research: An Incitement to Discourse

by Melissa Freeman, Kathleen deMarrais, Judith Preissle, Kathryn Roulston, and Elizabeth A. St. Pierre

In a climate of increased accountability, standardization, federal control, and politicization of education research and scholarship, this article briefly reviews various positions outlined by qualitative researchers about quality in qualitative inquiry, showing how these are implicated in the acquisition, conceptualization, and use of qualitative evidence. It concludes by identifying issues in and challenges to setting standards of evidence for qualitative researchers in education.

**Keywords:** qualitative research; quality; standards of evidence; validity

**T**his article addresses standards of evidence in qualitative research in education. Our premise is that it is neither desirable nor possible to reach consensus about or prescribe standards of evidence in this diverse field. Such prescriptions, we believe, amount to disciplinary action (Foucault, 1975/1979) that constrain the generation of knowledge rather than improve it. We do argue, however, that qualitative researchers both accomplish research of high quality and have a long tradition of demonstrating quality in reports of their investigations.

We begin by discussing the importance of this conversation at this historical and political moment in the United States. We then review how validity, a preferred term for the overall merit of a study, has been discussed by qualitative researchers. In this discussion, we consider commonalities in practice across qualitative research communities, describe how qualitative researchers have treated validity in relation to data and evidence, and explain how they have justified their claims. We conclude our review by emphasizing the heterogeneity of qualitative research and cautioning against recent calls for restrictive and disciplinary standards of evidence.

### The Context of the Discussion

The culture of science and academic scholarship is heterogeneous,<sup>1</sup> with multiple, overlapping communities of practice and knowledge that split into specialties or combine into interdisciplinary and cross-disciplinary areas of study. Within disciplinary communities such as education or sociology and across communities such as the anthropology of education, scholars have a long history of disagreeing and challenging one another about how to

ensure the quality of qualitative work and even how to catalog and categorize the various kinds of qualitative research.

Qualitative research is open and supple, and one of its strengths is that it incorporates philosophies, theories, and research designs and methods as diverse as postpositivist multi-methods approaches and postmodernist social critiques.<sup>2</sup> Rather than being prescriptive about what qualitative research is and what makes it good (true, valid) across all instances of research, qualitative researchers have tended to (a) study what researchers who say they are doing qualitative research are doing and then (b) encourage those who do that work to develop those practices considered excellent. Although discussions of quality in qualitative work have always been intense, they have usually been civil. We believe that this respect for one another's conventions is the result of the heterogeneity in qualitative design and the absence of an enduring hegemonic presence.

We situate our response to the call for standards of evidence in qualitative research in a pervasive discourse that describes educational research in general as historically and presently broken and in need of repair (e.g., Lagemann, 2000; Kaestle, 1993). Whether the situation is as dreary as it is made out to be is beside the point; today's fix is to make educational research scientific, and the federal government has taken the lead in this project by mandating scientific method into law (for scientism, see Lather, 2006; Ruccio & Amariglio, 2003). The fundamental idea is that rigorous science will make better schools, that quality science will enable us to finally reengineer schools so they work.

The National Research Council (NRC) has taken a leadership role in this conversation. Having defined the nature of science in its 2002 and 2005 reports, the NRC has received funding to begin a new "broad, long-term initiative related to the quality of evidence" (The National Academies, Division of Behavioral and Social Sciences and Education, 2006) in the social and behavioral sciences. With the NRC prepared to define evidence and the American Educational Research Association (AERA; 2006) imposing standards for reporting on research methods in its publications, qualitative researchers may feel under siege. Top-down efforts such as these to legislate scientific practice and mandate research design threaten to harden the boundaries of what counts as science, to devalue many qualitative research endeavors, and to limit creative research practice of all kinds.

## Commonalities in Standards of Practice Across Qualitative Research Communities

All sciences emphasize innovation and boundary breaking as well as standardization and systemization. Conventions themselves evolve and develop through trial and error and other kinds of assessment. Self-correction among communities of scientific practitioners has been the strength of science across alternative knowledge systems, and attempts to legislate standards have sometimes shut down scientific activity; the Soviet suppression of genetics research in the 20th century is a case in point (Soyfer, 1994). The tension between innovation and conventionalizing practice is, we believe, a sign of healthy inquiry. Overconventionalizing can lead to mindless recipes for research and limiting scholarly orthodoxies. Underconventionalizing has its own pitfalls: intellectual fads, a lack of continuity, fragmentation, and intellectual alienation. One way to achieve a balance here is to assess the status of conventions such as standards of evidence against what researchers actually do.

A key source, then, of standards of evidence and quality throughout the history of the scientific method and its application in qualitative inquiries has been the systematic and careful documentation of all procedures—an *account of practice*—to provide a record for a researcher's ongoing contemplation as well as for peer review. This is *descriptive* work. The documentation of procedure is crucial if we want to know what exemplary researchers whose methodology is innovative and effective do and how they make their work convincing. Examining expert researchers' practices, then, contributes to establishing agreed-on, albeit ever changing, standards of quality and rigor. Of course, working with such experts is much to be desired (e.g., Wolcott's [1994] version of how Louis Agassiz taught careful observation), and such conventionalizing not only lends confidence to results but also provides models for novices. As novices acquire experience and develop refined judgment, however, some come to prefer methodological uncertainty (e.g., Lather, 2004), and most come to rely less frequently on routine protocols than they did as newcomers (e.g., Benner, 2001; Flyvbjerg, 2001).

Completely entangled with practice is a second source of guidelines for exemplary qualitative research: *theory*, the various philosophical and theoretical traditions that both support communities of practice and develop from them. This work can be *prescriptive* because a priori theory may impose a structure of assumptions. Each community of practice using qualitative research traditions adheres to its own particular theoretical assumptions. For example, one of the standards of the Chicago school of qualitative sociology, begun in the 1920s, is to develop social science theory through research participants' descriptions of their own experiences, thus making explicit the "invisibility of everyday life" (Erickson, 1986, p. 121). In this tradition, the practice of checking a researcher's interpretations and representations with participants prior to publication is valued (e.g., Duneier, 1999).

More recently, the interdisciplinary nature of qualitative inquiry allowed it to welcome the epistemologies (e.g., queer theories, feminist theories, race-based theories) and accompanying methodologies of groups who organized the social movements of

the 1960s and 1970s to protest their absence from public policy, academic scholarship, and positions of power and influence throughout our culture. These as well as more mainstream epistemologies, such as critical and interpretive theories, vary both in what they consider standards of evidence and the priorities they give to shared standards.

In recognition of these variations, AERA (2002), in a commentary on its own ethical standards for editing, reviewing, and appraising research, has emphasized that it is important "to ensure fair treatment of those who submit manuscripts for publication and to promote the advancement of educational research through attention to the quality of research and *the preservation of the robust methodological pluralism of educational inquiry* [italics added]" (p. 103). Thus, the "methodological adequacy of a piece of work" must be judged in relation to the "methodological requirements of its type . . . [and] the significance of its results in the context of the problems internal to its own tradition, and not the requirements and aspirations of types to which it does not belong" (p. 105). And, of course, philosophies and theories, like scientific beliefs and research practices, change over time; thus, AERA valorizes emergent traditions in its commentary.

In qualitative inquiry, the last century's "linguistic turn" (e.g., Rorty, 1967) that critiqued the stability and transparency of language contributed to the proliferation of critical theories and the development of postmodern and poststructural approaches to science, research, and scholarship. These critiques question and unsettle our notions of truth and knowledge, objectivity and subjectivity, and science and evidence and warn of the danger of calls for norms and standards. As we have noted, such theories have their own logics that can in turn be interrogated for implicit and explicit standards of practice.

It is difficult to keep practice and theory from norming each other and thus shutting down innovation in methods of inquiry. Communities of practice inevitably develop and enforce standards. But if these communities are to thrive, they must establish procedures for the ongoing interrogation of those standards. Working in the tension of simultaneously doing science and troubling it is not always easy, and this approach requires curiosity and generosity as we encounter different and what may seem at the time incommensurable, threatening, and even dangerous theories and practices, with their accompanying standards. Yet this balancing act enables conditions for good science. How then, have qualitative researchers discussed what good science is and how to assess the quality of their work?

### Validity

Qualitative researchers have always discussed how to evaluate their science, the quality of their analyses and theoretical interpretations of data. They disagree, however, over the terms used in these discussions: *validity*, *reliability*, *rigor*, and parallel terms such as *trustworthiness*, *credibility*, *transferability*, *verisimilitude*, *relevance*, *plausibility*, and *confirmability*. The literature contrasts qualitative quality standards to those used in the scientific method (Eisenhart & Howe, 1992; Howe & Eisenhart, 1990); discusses validity in general (Kvale, 1995; Lincoln, 1995; Maxwell, 1996; Miles & Huberman, 1984; Seale, 2004; J. K. Smith & Deemer, 2000;

J. K. Smith & Hodkinson, 2005); and considers it within specific qualitative traditions, such as ethnography (Altheide & Johnson, 1994; LeCompte & Goetz, 1982; Willis & Trondman, 2000), action research (Anderson & Herr, 1999; Melrose, 2001), alternative ethnographic designs (Denzin, 2000; Ellis, 2000; Richardson, 2000), mixed theory designs (Koro-Ljungberg, 2004), postmodern, feminist, critical, and poststructural designs (Lather, 2001; Lenzo, 1995; Scheurich, 1993), and self-study designs (Bullough & Pinnegar, 2001; Feldman, 2003). Lather (1993) has even claimed to be obsessed with validity. Validity has been defined and described in a variety of ways and for a variety of purposes (LeCompte, Preissle, & Tesch, 1993; Maxwell, 1992; Talburt, 2004). Because *validity* as a term has been discussed, debated, contested, and redefined by qualitative researchers, we focus on validity as characteristic of the standards of evidence discourse as it relates to qualitative inquiry.

Validity is generally understood by educational researchers as “the trustworthiness of inferences drawn from data” (Eisenhart & Howe, 1992, p. 644). Indeed, Scheurich (1993) remarked that validity is the boundary line (that differs among epistemologies) for what is acceptable and not acceptable in research. Both Moss (1996) and Mishler (1990) echoed Scheurich’s point that validity is an epistemological issue that methodological procedures can barely begin to address. Thus, Moss warned of the imposition of “a priori criteria abstracted from existing practice” (p. 26) because of the very different theoretical work done under the umbrella of qualitative research. Lather (2001) summarized these cautions with her reminder that ever since the publication of “Cronbach and Meehl’s 1955 essay on the problems with construct validity in psychological testing, validity has been the problem, not the solution” (p. 243). Driving these discussions is the need for researchers to address a central question in any kind of research inquiry: “Why should I believe this?” (Wallace & Wray, 2006, p. 28). How then do qualitative researchers deal with validity in their research? In the following sections, we show how qualitative researchers use data as evidence to warrant claims within different theoretical frameworks and specific communities of practice. These activities are separable only heuristically; in research practice, they intertwine and implicate one another throughout the course of any study.

## Data and Evidence

Qualitative researchers’ concerns about the quality of their work are evident in discussions about formulating both research design and questions within explicit theoretical and philosophical traditions; accessing and entering settings; selecting, collecting, and analyzing data; and building a case for conclusions. Quality is constructed and maintained continuously throughout the life of a research project and includes decisions that researchers make as they interact with those they study and as they consider their analyses, interpretations, and representations of data. Qualitative researchers in education have used, generated, and redefined various terms that relate to the assessment of quality shared with other social, human, and professional science researchers: *credibility*, *validity*, *triangulation*, *trustworthiness*, *truth*, and *verification*. Qualitative scholars have seldom used the terms *standards*,

*evidence*, *claims*, and *warrants*, terms derived from formal logic that are currently used in policy documents and reports (for a notable exception, see Morse, Swanson, & Kuzel, 2001).<sup>3</sup>

In methodological writing, the term *qualitative data* is generally taken to encompass the “rough materials researchers collect from the world they are studying” (Bogdan & Biklen, 2006, p. 117), including field notes, documents, transcriptions of interviews and interactions, and artifacts. Data are produced from social interactions and are therefore constructions or interpretations. There are no “pure,” “raw” data, uncontaminated by human thought and action, and the significance of data depends on how material fits into the architecture of corroborating data. “Data analysis leads to a *reconstruction* of those constructions” (Lincoln & Guba, 1985, p. 132). In other words, qualitative data and information are always already interpretations made by participants as they answer questions or by researchers as they write up their observations. Neither research participants nor researchers can be neutral, because, as emphasized earlier, they are always positioned culturally, historically, and theoretically. There is no Archimedean standpoint outside human activity (Alexander, 2006; Hartsock, 1983) from which to claim neutrality and produce value-free data.

In a now classic consideration of qualitative methods in education, Erickson (1986) asserted that

the corpus of materials collected in the field are not data themselves, but resources for data. Fieldnotes, videotapes, and site documents are not data. Even interview transcripts are not data. All these are documentary materials from which data must be constructed through some formal means of analysis. (p. 149)

Nevertheless, most qualitative textbooks now use a definition of data similar to Bogdan and Biklen’s preceding formulation. Although Bogdan and Biklen claimed that “data are both the evidence and the clues” (p. 117), Lincoln (2002) cautioned that “data and information are not evidence until two things happen: first, someone recognizes it as data, and second, an inquirer subjects it to some form of systematic analysis, which turns it into evidence directed toward some question or argument” (p. 6).

## Claims and Interpretations

In research, claims are statements of meaning grounded in evidence and theory. Claims describe, interpret, deconstruct, critique, predict, and explain lived experience. Claims are statements that connect the world bounded by our data to our interpreted understanding of that data.

In 1962, Raoul Naroll, an anthropologist, working from concerns about the misuse of data, called for data quality control. Naroll tried to assess information by the conditions under which it was generated and the extent to which it compared and contrasted with other information. Because he studied patterns of cross-cultural behavior and experience, he was concerned about the accuracy of reports describing cultures across the world and whether they were biased by incomprehension, misapprehension, or downright malice. Although in the 21st century, people are better equipped to record information almost everywhere about almost everyone, misunderstandings, alternative understandings,

and conflicting standpoints continue to complicate the generation of knowledge, and most scholars recognize that the multiple layers of meaning in human experience can be inconsistent, incoherent, and even incomprehensible as well as sometimes consistent, coherent, and comprehensible. The material that constitutes data and the constituting of the data are limited and fallible, and the scholar's task is to identify the limitations and faults of information along with its value.

Justifying claims also depends on demonstrating and reporting appropriate and adequate methods of data generation. Naroll (1962) discussed sets of general criteria that might characterize competent work (for critiques of what is called criteriology, see, e.g., Schwandt, 1996; J. K. Smith & Deemer, 2000; J. K. Smith & Hodgkinson, 2005). From expectations of careful documentation of systematic fieldwork such as those developed by Naroll, succeeding scholars have developed expectations for the conduct of quality research. These include attention to (a) thorough description of design and methods in reports, (b) adequate demonstration of the relationship of claims to data, and (c) thoughtful consideration by the researcher of the strengths and limitations of the study. Next, we consider each of these issues in turn.

Thorough description of design and methods is the effort to represent decisions, procedures, and researcher thinking in ways that audiences find recognizable and comprehensible. Practices that support this effort can be contrasted with the mystique fostered among some artists who fear that demystifying or describing their work in detail might threaten the product rather than enhance it. Thus, researchers commonly provide detailed descriptions of how they went about their studies, the problems they encountered, and the reasoning on which they based their decisions (e.g., Fine, 1991; Fordham, 1996; Peshkin, 1986; Valenzuela, 1999). These methodological explorations often appear in anthologies of researchers' accounts of their studies (e.g., Behar & Gordon, 1995; Lareau & Shultz, 1996; St. Pierre & Pillow, 2000).

The relationship of data and claims is demonstrated in research reports by offering adequate and appropriate information for readers to reexamine and assess a researcher's assertions and interpretations. At a minimum, sufficient data are cited in reports to support each claim (e.g., Lortie's [1975] study of teachers demonstrates the patterns he claimed by extensive quotations from what the teachers had to say); at a maximum, researchers make data available to others for secondary research (e.g., Goldman-Segall, 1998, has made video data from her digital ethnography available to readers on the Web). How much material is provided and whether entire collections of field notes, interviews, and documents are made available to the public are conventions that vary by qualitative tradition. Oral historians and ethnographers often archive their data, for example.

Finally, how researchers address the strengths and limitations of their studies includes issues about the relationships of researchers and participants, the roles of researchers in their studies, and the ethics and politics of representation. Of course, these decisions intersect with decisions about how much detail to disclose about methods and design and how to select data to support assertions. For example, researchers may decide to limit access to material from their research studies for ethical reasons. In particular, researchers working with indigenous groups whose traditional

knowledge and lore have long been exploited by researchers are well advised to consider the ethical and political implications of decisions concerning all aspects of doing research (e.g., L.T. Smith, 1999, 2005).

Only recently have some qualitative scholars begun to explicitly address the issue of quality using the terminology of "standards of evidence." For example, Wilson (1994) proposed five criteria that address the nature of the information, how it is acquired, and how it is interpreted: "evidence should be consistent with a researcher's chosen epistemology or perspective" (p. 26), "evidence should be observable" (p. 28), "evidence should be gathered through systematic procedures" (p. 29), "evidence should be shared and made public" (p. 30), and "evidence should be compelling" (p. 30).

Lincoln (2002) offered another set of criteria: (a) "researchers should have been deeply involved and closely connected to the scene"; (b) "researchers should achieve enough distance from the phenomenon to permit recording action and interpretations relatively free of the researcher's own stake"; (c) "claims should be based on an adequate selection of the total corpus of data"; (d) "data should come, at least partly, from publicly accessible observation records"; and (e) "data and analysis should include consideration of inferences and interpretations, as well as concrete phenomena" (p. 9). These are two possibilities for assessing standards of evidence that may fit some qualitative traditions.

The introductory texts scholars use to teach qualitative inquiry also offer suggestions for how researchers use evidence to support their claims. Because the relational aspects of qualitative work are so important, scholars value extended time in the field, what Wax (1971) called immersion. "Being there" matters. Qualitative methodologists also encourage member checks: going back to participants and asking them, "Have I got it right?" Working with other researchers—peer debriefers and research groups—to help think about the complexity and ethics of the work is also recommended (Lincoln & Guba, 1985). And because most qualitative research is grounded in descriptive claims about the work, not only are sufficient data to support claims crucial, but researchers must be able to, in Geertz's (1973, p. 10) words, "contrive somehow first to grasp" what is going on before they can represent it for others. Writing of anthropological ethnography, Geertz added,

Doing ethnography is like trying to read (in the sense of "construct a reading of") a manuscript—foreign, faded, full of ellipses, incoherencies, suspicious emendations, and tendentious commentaries, but written not in conventionalized graphs of sound but in transient examples of shaped behavior. (p. 10).

Geertz has explained elsewhere (as cited in an interview with Olson, 1991) that description as a rhetorical marker of validity surfaces in representations of research with a "sense of circumstantiality and of *power in reserve* [italics added] (if an anecdote or an example doesn't sound strained but sounds like you've got fifty others and this is the best one you chose)" (p. 249).

Using multiple researchers, multiple methods of data collection, and multiple theoretical analyses to complicate rather than simplify knowledge production also provide warrants for our claims. And, of course, peer review both during the research process and prior to publication has been standard practice. There

is no single marker of validity in qualitative inquiry, and the best qualitative research uses many of the strategies just described and invents others specific to the particular study. Thus, validity cannot be defined in advance by a certain procedure but must be attended to at all times as the study shifts and turns.

In contrast to the emphasis on what qualitative researchers ought to do, Erickson (1986) offered “five major types of evidentiary *inadequacy* [italics added]”: (a) “inadequate amounts of evidence,” (b) “inadequate variety in kinds of evidence,” (c) “faulty interpretive status of evidence,” (d) “inadequate disconfirming evidence,” and (e) “inadequate discrepant case analysis” (p. 140). Researchers, of course, must establish an evidentiary warrant for assertions or claims they make. Erickson wrote that assertions generated during fieldwork “are tested and retested against the data base: the corpus of fieldnotes, interview protocols, site documents” (p. 146), and so forth. Here, Erickson did use the language of warrants and evidence to discuss the importance of disconfirming cases:

To test the evidentiary warrant for an assertion the researcher conducts a systematic search of the entire data corpus, looking for disconfirming and confirming evidence. . . . If the discrepant cases outnumbered those that fitted the assertion, the assertion would not be warranted by the data. Even if most of the cases fitted the assertion, the discrepant instances would be noted for subsequent analysis. (p. 146)

Yet qualitative researchers working from other theoretical perspectives disagree that there is such a thing as a disconfirming case (e.g., Lather, 1993; Scheurich, 2001). Another way to think about data or evidence that fail to fit emerging patterns is to rework the patterns to better represent the data. This work produces general statements, interpretations, and theories.

### Generalizing and Theorizing

The way claims are warranted in qualitative research typically centers on the data, but researchers also use the literature relevant to studies, information collected by other scholars, and a variety of other sources to justify their claims. People unfamiliar with qualitative research assume that knowledge produced is not generalizable in the sense that it does not make what Kaplan (1964), a philosopher of science, called nomological generalizations, assertions that are “truly universal, unrestricted as to time and space . . . always and everywhere the case, provided only that the appropriate conditions are satisfied” (p. 91). Kaplan himself, however, distinguished among a variety of generalizations, assertions, and claims. Nomological or lawlike generalization rarely carries the weight in interpretive research that it does in positivist research, because the goal is not to generalize to predict and control but rather to describe what people do and say within local contexts. These particularistic generalizations or assertions on the basis of the direct experience of observation and interviewing, called generic propositions by Kaplan and naturalistic generalizations by Stake and Trumbull (1982), are the goal of much qualitative work. Lewis and Ritchie (2003) used the term *representational generalization* for similar assertions to emphasize that the particularistic generalization should be representative of the context and participants studied.

Qualitative researchers often make connections across studies to establish the applicability of their work. For these, Lewis and

Ritchie (2003) distinguished inferential generalization from theoretical generalization. The former applies a proposition to settings and people other than those studied, and the latter applies constructs developed in a study to the generation and refinement of theory. The legitimacy of these claims depends on how qualitative researchers use the literature and other sources to build their arguments. Inferential generalization in qualitative work revolves around careful comparison across settings and people. Theoretical generalization, in contrast, may depend on how convincingly a researcher accounts for diverse patterns with a compelling construct. The ultimate judge for some claims is, of course, the readers of the research, who decide whether claims made apply to their situations.

Theory and practice are inseparable in doing qualitative research. As we have shown, researchers generate theory from their data through a complex process of warranting their claims. Geertz (1973) wrote that this kind of theory needs to “stay rather closer to the ground” (p. 24). Such “theoretical formulations hover so low over the interpretations they govern that they don’t make much sense or hold much interest apart from them” (p. 25). But researchers also produce evidence on the basis of a priori theory. These include macro-level theory such as positivism, social constructionism, Marxism, and feminism, as well as midlevel theories such as cognitive and linguistic theories. Making use of data and information as evidence means “relying on background knowledge and auxiliary hypotheses, of lading data with theory” (Wylie, 2002, p. 169).

So the social, material, political, and theoretical contexts of research are ever present, because all forms of evidence “presuppose a society within which they are symbolically meaningful” (Alford, 1998, p. 36). The contingency of evidence may trouble those positivists, logical positivists, and scientific realists who produced an alternative to logical positivism in the 1960s and 1970s. Those scholars use an objectivist, realist, and foundationalist epistemology and seem to believe that because qualitative research acknowledges that science, evidence, and truth are contingent, it is not valid, or, as the What Works Clearinghouse implies, it is “weak.” Interpretivists, on the other hand, suspect that positivists and scientific realists deny the role of theory, culture, and politics in research when they assume, for instance, that “to describe the physical and social world scientifically . . . multiple observers can agree on what they see” (NRC, 2002, p. 25), with no consideration for how the knower shapes the known.

### Conclusion

As we suggest in the title of this article and throughout, we intend our work here to be the thoughts of one group of qualitative researchers, not an authoritative account of standards of evidence in qualitative inquiry. Although we have described in some depth the many systematic and scientific ways qualitative researchers conduct their studies, we emphasize that we are not advocating for a set of standards of evidence that may be taken up by others and used as a checklist to police our work—quite the contrary. We call on other researchers, both qualitative and quantitative, to resist current political forces seeking to impose a set of restrictive standards on educational research that serve only to control what research gets funded and conducted and, at the same time, to inhibit the creation of new research methodologies.

Although not all qualitative researchers agree about the application or purpose of quality standards or even whether there should be standards, several common concerns stand out as having significantly shaped the dialogue and debate on this issue. Because the strength of qualitative research is the close contact and connection that qualitative researchers have with research participants, this contact has played a central role in shaping the principles and quality standards that guide them. A guiding question has been, How can we best listen to, work with, and represent the people our work is intended to serve?

Elliott (2006), for example, favors educational research as a practical science and rejects current conceptions of “practical rationality” cast in terms of “the science of measurement” (pp. 180–181). Federal government agencies such as the Institute of Education Sciences, policy makers, funding agencies, and sources external to the educational research community and their constituents expect educational research to produce generalizable, unambiguous, and immediately applicable solutions to complex educational problems. In contrast to this engineering model, qualitative researchers have long been engaged in a more pragmatic conversation. How can research generate useful, informational, and thought-provoking feedback or knowledge to relevant and interested communities of scholars and practitioners? Rather than focusing on eliminating the subjectivity of the researcher in a fruitless effort to attain objective knowledge, qualitative researchers pursue how best to work with the fruitful positionings that each researcher brings to a project (e.g., Peshkin, 1988).

Throughout this article, we have emphasized the heterogeneity of qualitative research practice. We five authors do not all agree on or even view similarly what we have written here. What we have tried to do is identify common notions of qualitative evidence and the standards qualitative researchers use to judge whether the claims they make are warranted or justified by that evidence. As readers of qualitative work, we look for indications of researcher skepticism: Do authors of research reports reveal themselves as their own best critics? Do they discuss the limits and uncertainties of their work? How forthright are they about competing interpretations and explanations for the patterns they claim?

Representing the multiple layers of human experience is fraught with challenge, alternative, and limitation. But everything is not, as some argue, “just a matter of opinion,” nor is what makes a qualitative study good a simple matter of meeting a checklist of criteria. First, not all criteria are relevant to the range and variety of qualitative work. Second, pedestrian work can satisfy criteria without producing anything worth knowing. Third, applying even the limited standards we have discussed in this article is no small endeavor: It requires expert judgment. Fourth, expectations for good qualitative work have historically been tailored to the particular theoretical frameworks, methodological traditions, and substantive issues addressed. We think that this is the level at which conversations about quality should remain. For example, standards to judge quality for life history research may be inappropriate for microethnography. Fifth, what we have tried to do in this article is to indicate how research practice and theory interact with relationships built with participants such that decisions about quality are made as the research progresses. We hope that our discussion here will be another

incitement in the continuing conversations about truth and validity that have always preoccupied qualitative researchers as we struggle to generate epistemologies and methodologies that enable us to grapple with the complex world in which we live and do science.

## NOTES

<sup>1</sup>The disciplinary sources for qualitative research are multiple and diverse, including social and cultural anthropology; qualitative sociology; professional studies such as education, law, business, counseling, social work, library science, medicine, nursing, and the health professions; history; psychology, especially clinical, developmental, and cognitive studies; case traditions in areas such as political science and economics; communications and journalism, especially investigative reporting and media studies; fieldwork in the natural sciences, especially biology, geology, astronomy; and the arts and humanities.

<sup>2</sup>Approaches to research design are likewise multiple and overlapping, including (but not limited to) ethnography, field study, community study, case study, life history and biographical study, phenomenological approaches, conversation analysis, oral history, the variety of kinds of interview studies, document analysis and other historical study, survey study, autoethnography, narrative inquiry, connoisseurship, portraiture, action research, insider research, collaborative research, observational study, and multimethods research including unobtrusive research.

<sup>3</sup>We support this claim with our analysis of the qualitative literature (Preissle, Freeman, deMarrais, Roulson, & St. Pierre, n.d.)

## REFERENCES

- Alexander, H. A. (2006). A view from somewhere: Explaining the paradigms of educational research. *Journal of Philosophy of Education, 40*(2), 205–221.
- Alford, R. R. (1998). *The craft of inquiry: Theories, methods, evidence*. New York: Oxford University Press.
- Altheide, D. L., & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485–499). Thousand Oaks, CA: Sage.
- American Educational Research Association (2002). *Ethical standards of the American Educational Research Association: Cases and commentary*. Washington, DC: Author.
- American Educational Research Association. (2006). *Standards for reporting on empirical social science research in AERA publications*. Retrieved September 22, 2006, from [http://www.aera.net/uploadedFiles/Opportunities/StandardsforReportingEmpiricalSocialScience\\_PDF.pdf](http://www.aera.net/uploadedFiles/Opportunities/StandardsforReportingEmpiricalSocialScience_PDF.pdf)
- Anderson, G. L., & Herr, K. (1999). The new paradigm wars: Is there room for rigorous practitioner knowledge in schools and universities? *Educational Researcher, 28*, 12–21, 40.
- Behar, R., & Gordon, D. A. (Eds.). (1995). *Women writing culture*. Berkeley: University of California Press.
- Benner, P. (2001). *From novice to expert: Excellence and power in clinical nursing practice* (Comm. ed). Upper Saddle River, NJ: Prentice Hall.
- Bogdan, R. C., & Biklen, S. K. (2006). *Qualitative research for education: An introduction to theories and methods* (5th ed.). Boston: Pearson Education Group.
- Bullough, R. V., Jr., & Pinnegar, S. (2001). Guidelines for quality in autobiographical forms of self-study research. *Educational Researcher, 30*, 13–21.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity. *Psychological Bulletin, 52*, 281–302.
- Denzin, N. K. (2000). Aesthetics and the practices of qualitative inquiry. *Qualitative Inquiry, 6*, 256–265.

- Duneier, M. (1999). *Sidewalk*. New York: Farrar, Straus.
- Eisenhart, M. A., & Howe, K. R. (1992). Validity in educational research. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 644–680). San Diego, CA: Academic Press.
- Elliott, J. (2006). Educational research as a form of democratic rationality. *Journal of Philosophy of Education*, 40(2), 169–185.
- Ellis, C. (2000). *Creating criteria: An ethnographic short story*. *Qualitative Inquiry*, 6, 273–277.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York: Macmillan.
- Feldman, A. (2003). Validity and quality in self-study. *Educational Researcher*, 32, 26–28.
- Fine, M. (1991). *Framing dropouts: Notes on the politics of an urban public high school*. Albany: State University of New York Press.
- Flyvbjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge, UK: Cambridge University Press.
- Fordham, S. (1996). *Blacked out: Dilemmas of race, identity, and success at Capital High*. Chicago: University of Chicago Press.
- Foucault, M. (1979). *Discipline and punish: The birth of the prison* (Alan Sheridan, Trans.). New York: Vintage. (Original work published 1975)
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz, *The interpretation of cultures: Selected essays* (pp. 3–30). New York: Basic Books.
- Goldman-Segall, R. (1998). *Points of viewing children's thinking: A digital ethnographer's journey*. Mahwah, NJ: Lawrence Erlbaum.
- Hartsock, N. C. M. (1983). The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In S. Harding & M. B. Hintikka (Eds.), *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology and philosophy of science* (pp. 283–310). Boston: D. Reidel.
- Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher*, 19, 2–9.
- Kaestle, C. F. (1993). The awful reputation of education research. *Educational Researcher*, 22, 23–31.
- Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. Scranton, PA: Chandler.
- Koro-Ljungberg, M. (2004). Impossibilities of reconciliation: Validity in mixed theory projects. *Qualitative Inquiry*, 10, 601–621.
- Kvale, S. (1995). The social construction of validity. *Qualitative Inquiry*, 1, 19–40.
- Lagemann, E. C. (2000). *An elusive science: The troubling history of education research*. Chicago: University of Chicago Press.
- Lareau, A., & Shultz, J. (Eds.). (1996). *Journeys through ethnography: Realistic accounts of fieldwork*. Boulder, CO: Westview.
- Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, 34(4), 673–693.
- Lather, P. (2001). Validity as an incitement to discourse: Qualitative research and the crisis of legitimation. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 241–250). Washington, DC: American Educational Research Association.
- Lather, P. (2004, April). *Getting lost: Feminist efforts toward a double(d) science*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Lather, P. (2006). Foucauldian scientificity: Rethinking the nexus of qualitative research and educational policy analysis. *International Journal of Qualitative Studies in Education*, 19, 783–791.
- LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research*, 52(1), 31–60.
- LeCompte, M. D., Preissle, J., & Tesch, R. (1993). Evaluating qualitative design. In M. D. LeCompte, & J. Preissle, (Eds.), *Ethnography and qualitative design in educational research* (2nd ed., pp. 315–356). San Diego, CA: Academic Press.
- Lenzo, K. (1995). Validity and self-reflexivity meet poststructuralism: Scientific ethos and the transgressive self. *Educational Researcher*, 24, 17–23, 45.
- Lewis, J., & Ritchie, J. (2003). Generalising from qualitative research. In J. Ritchie & J. Lewis (Eds.), *Qualitative research practice: A guide for social science students and researchers* (pp. 263–286). London: Sage.
- Lincoln, Y. S. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry*, 1, 275–289.
- Lincoln, Y. S. (2002, November). *On the nature of qualitative evidence*. Paper presented at the annual meeting of the Association for the Study of Higher Education, Sacramento, CA.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62(3), 279–300.
- Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.
- Melrose, M. J. (2001). Maximizing the rigor of action research: Why would you want to? How could you? *Field Methods*, 13(2), 160–180.
- Miles, M. B., & Huberman, A. M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher*, 13, 20–30.
- Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review*, 60(4), 415–442.
- Morse, J. M., Swanson, J. M., & Kuzel, A. J. (Eds.). (2001). *The nature of qualitative evidence*. Thousand Oaks, CA: Sage.
- Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25, 20–28.
- Naroll, R. (1962). *Data quality control—A new research technique: Prolegomena to a cross-cultural study of culture stress*. New York: Free Press.
- The National Academies, Division of Behavioral and Social Sciences and Education. (2006). *Standards of evidence: Strategic planning initiative*. Retrieved June 30, 2006, from <http://www7.nationalacademies.org/dbasse/Standards%20of%20Evidence%20Description.html>
- National Research Council. (2002). *Scientific research in education*. R. J. Shavelson & L. Towne (Eds.). Committee on Scientific Principles for Education Research. Washington, DC: National Academy Press
- National Research Council. (2005). *Advancing Scientific Research in Education*. L. Towne, L. L. Wise & T. M. Winters (Eds.). Committee on Research in Education. Washington, DC: National Academy Press.
- Olson, G. (1991). The social scientist as author: Clifford Geertz on ethnography and social construction. *Journal of Advanced Composition*, 11(2), 245–268.
- Peshkin, A. (1986). *God's choice: The total world of a fundamentalist Christian school*. Chicago: University of Chicago Press.
- Peshkin, A. (1988). In search of subjectivity: One's own. *Educational Researcher*, 17, 17–22.
- Preissle, J., Freeman, M., deMarras, K., Roulston, K., & St. Pierre, E. A. (n.d.). *How qualitative researchers discuss quality work*. Retrieved September 23, 2006, from <http://www.coe.uga.edu/leap/qualresearch/index.html>
- Richardson, L. (2000). Evaluating ethnography. *Qualitative Inquiry*, 6, 253–255.
- Rorty, R. M. (Ed). (1967). *The linguistic turn: Essays in philosophical method*. Chicago: University of Chicago Press.

- Ruccio, D. F., & Amariglio, J. (2003). *Postmodern moments in modern economics*. Princeton, NJ: Princeton University Press.
- Scheurich, J. J. (1993). The masks of validity: A deconstructive investigation. *International Journal of Qualitative Studies in Education*, 9(11), 49–60.
- Scheurich, J. J. (2001). *Research method in the postmodern*. Philadelphia, PA: Routledge.
- Schwandt, T. A. (1996). Farewell to criteriology. *Qualitative Inquiry*, 2, 58–72.
- Seale, C. (2004). Quality in qualitative research. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 409–419). London: Sage.
- Smith, J. K., & Deemer, D. K. (2000). The problem of criteria in the age of relativism. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2nd ed., pp. 877–896). Thousand Oaks, CA: Sage.
- Smith, J. K., & Hodkinson, P. (2005). Relativism, criteria, and politics. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (3rd ed., pp. 915–932). Thousand Oaks, CA: Sage.
- Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. London: Zed.
- Smith, L. T. (2005). On tricky ground: Researching the native in the age of uncertainty. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 85–108). Thousand Oaks: Sage.
- Soyfer, V. N. (1994). *Lysenko and the tragedy of Soviet science* (L. Gruliov & R. Gruliov, Trans.). New Brunswick, NJ: Rutgers University Press.
- Stake, R., & Trumbull, D. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science*, 7(1), 1–12.
- St. Pierre, E. A., & Pillow, W. S. (Eds.). (2000). *Working the ruins: Feminist poststructural theory and methods in education*. New York: Routledge.
- Talbur, S. (2004). Ethnographic responsibilities without the “real.” *Journal of Higher Education*, 75(1), 80–103.
- Towne, L., Wise, L. L., & Winters, T. M. (Eds.). (2005). *Advancing scientific research in education*. Washington, DC: The National Academies Press.
- Valenzuela, A. (1999). *Subtractive schooling: US-Mexican youth and the politics of caring*. Albany: State University of New York Press.
- Wallace, M., & Wray, A. (2006). *Critical reading and writing for post-graduates*. Thousand Oaks, CA: Sage.
- Wax, R. H. (1971). *Doing fieldwork: Warnings and advice*. Chicago: University of Chicago Press.
- Willis, P., & Trondman, M. (2000). Manifesto for ethnography. *Ethnography*, 1, 5–16.
- Wilson, B. J. (1994). A challenge to communication empiricists: Let’s be more forthcoming about what we do. *Western Journal of Communication*, 58(1), 25–31.
- Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage.
- Wylie, A. (2002). *Thinking from things: Essays in the philosophy of archaeology*. Berkeley: University of California Press.

## AUTHORS

**MELISSA FREEMAN** is an assistant professor in the Department of Lifelong Education, Administration, and Policy in the College of Education at The University of Georgia, 329 River’s Crossing, Athens, GA 30602; [freeman9@uga.edu](mailto:freeman9@uga.edu). Her research interests include transformational and relational theories and methodologies in qualitative research and evaluation, parental perceptions of public schooling, and philosophical hermeneutics.

**KATHLEEN DEMARRAIS** is a professor in the Department of Lifelong Education, Administration and Policy and associate dean in the College of Education at The University of Georgia, G3 Aderhold Hall, Athens, GA 30602; [kpd@uga.edu](mailto:kpd@uga.edu). Her research interests include the political context of qualitative research, qualitative pedagogy, and the influence of neoconservative philanthropy on educational policy.

**JUDITH PREISSE** is a professor in the Department of Lifelong Education, Administration, and Policy in the College of Education at The University of Georgia, 303 River’s Crossing, Athens, GA 30602, and an affiliated member of UGA’s Institute for Women’s Studies; [jude@uga.edu](mailto:jude@uga.edu). She teaches, researches, and writes in educational anthropology, qualitative research, feminist studies, and ethics.

**KATHRYN ROULSTON** is an associate professor in the Department of Lifelong Education, Administration, and Policy in the College of Education at The University of Georgia, 324 River’s Crossing, 850 College Station Road, Athens, GA 30602; [roulston@uga.edu](mailto:roulston@uga.edu). Her research interests include qualitative research methodology, qualitative interviewing, analysis of talk-in-interaction, and topics in music education.

**ELIZABETH A. ST. PIERRE** is a professor of language and literacy education and a qualitative methodologist at The University of Georgia, 125 Aderhold Hall, Athens, GA 30602; [stpierre@uga.edu](mailto:stpierre@uga.edu). Her research interests include poststructural theories of language and subjectivity and a critique of conventional qualitative inquiry.

Manuscript received June 30, 2006  
 Revision received September 25, 2006  
 Accepted September 28, 2006