Evaluating Qualitative Research

Jeasik Cho and Allen Trent

The Oxford Handbook of Qualitative Research (1 ed.)
Edited by Patricia Leavy

Print Publication Date: Jul 2014
Subject: Psychology, Psychological Methods and Measurement
Online Publication Date: Aug 2014 DOI: 10.1093/oxfordhb/9780199811755.013.012

Abstract and Keywords

This chapter addresses a wide range of theories and practices related to the evaluation of qualitative research (EQR). First, six categories of EQR are presented: (1) a positivist category, (2) Lincoln and Guba's alternative category, (3) a "subtle-realist" category developed by Hammersley and Atkinson, and Seale, (4) a general EQR category, (5) a category of post-criteriology, and (6) a post-validity category. Second, evaluation strategies for EQR are offered by providing a variety of actual examples. Third, the chapter discusses a path forward for EQR that includes both internal and external elements. The chapter concludes with a holistic view of EQR needed to collectively construct/confront inner and outer challenges to qualitative paradigms in the twenty-first century. Twenty-first-century criteria supported include thought-provoking ideas, innovative methodology, performative writing, and global ethics and justice mindedness.

Keywords: Evaluation criteria, validity, checklists, rubrics, politics of evidence, twenty-first-century criteria

> Quality is elusive, hard to specify, but we often feel we know it when we see it. In this respect research is like art rather than science.
>
> – *Seale, 2002, p. 102*
>
> Criteria in the 21st century are not one-dimensional.
>
> – *Lichtman, 2006, p. 197*

We feel exactly the same way that frontier scholars of grounded theory Juliet Corbin and Anselm Strauss (2008) feel regarding the evaluation of qualitative research:

> I feel paralyzed, unsure of where to begin, or what to write. As I search the literature, I find that evaluation is necessary but there is little consensus about what that evaluation should consist of. Are we judging for "validity" or would it be bet-

> ter to use terms like “rigor”...“trustworthiness,”... or “goodness,”... or something called “integrity”... when referring to qualitative evaluation? (p. 297)

Let us select the term *validity*. “Validity has been referred to many ways, including successor validity, catalytic validity, interrogated validity, transgressive validity, imperial validity, simulacra/ironic validity, situated validity, and voluptuous validity” (Altheide & Johnson, 2011, pp. 584–585), and a review of the qualitative literature tells us that there are many more definitions.

Why is it so hard to get started “evaluating” qualitative research? Patton (2002) notes that “some of the confusion that people have in assessing qualitative research stems from thinking it represents a uniform perspective, especially in contrast to (p. 678) quantitative research. This makes it hard for them to make sense of the competing approaches within qualitative inquiry” (p. 543).

So, to evaluate qualitative research, shall we simply follow Altheide and Johnson’s (2011) lead? In their chapter in the *Sage Handbook of Qualitative Research*, entitled “Reflections on Interpretive Adequacy in Qualitative Research,” their approach was threefold: they updated their well-known article “Criteria for Assessing Interpretive Validity in Qualitative Research” (Altheide & Johnson, 1994), they called their ideas about this job “analytical realism,” and they proposed an “evidentiary narrative” embedded in “a symbolic interactionist perspective” (p. 582) that goes against neo-positivist, scientific, or evidence-based research. We are impressed with their deep philosophical, provocative ideas on developing a new grand quality criterion in response to the current scientific, evidence-based movement that devalues an ideal of qualitative research, but we are more interested in exploring a broader sense of evaluative criteria in qualitative research. We call our approach *evaluating qualitative research* (EQR).

By and large, we are baffled by at least three issues regarding the evaluation of qualitative research: little agreement with the nature of evaluation in qualitative research, a continuous impact of traditional positivist evaluation criteria on qualitative research, and a broad political discourse on the politics of evidence. At least, however, we agree with Schwandt’s (2002) viewpoint that constructing an evaluation lens that involves general and specific accounts of what we might hope to find in a good study is exciting intellectual work. Schwandt’s four general approaches to evaluating qualitative research are to use (1) universal conventional criteria, (2) alternative criteria of trustworthiness and authenticity, (3) pragmatic criteria, and (4) subtle realist criteria of validity and relevance. Although we are impressed with his scheme for a developmental perspective on EQR, our feeling is that this kind of framework is, by itself, something like recreating what has already been deemed disagreeable in this field.

Despite the field’s confusion, disagreements, and our perplexed reaction, our thesis on EQR in this chapter is clear. We express a very simple but meaningful perspective on the evaluation of the processes and products of qualitative research. Our perspective is threefold. First, because we observe that EQR is seen as a relatively cohesive discourse (e.g., a huge number of journal articles and book chapters start with Lincoln and Guba’s [1985]

seminal book, *Naturalistic Inquiry,* and an equally large number of qualitative studies reference and rely on Lincoln and Guba's [1985] construction of *trustworthiness* criteria), we want to provide "a sketch of EQR" to categorically describe qualitative differences among many different theoretical and practical ideas related to qualitative research evaluation. Second, we provide several evaluation strategies for EQR. And third, we discuss a path forward for EQR that includes both internal and external elements. We conclude this chapter with a beehive metaphor, which gives a holistic view of the kind of EQR needed to collectively construct, collaborate, and confront inner and outer challenges to qualitative paradigms in the twenty-first century.

Evaluation of Qualitative Research: Six Categories

Under the umbrella of qualitative research over the past three decades, the EQR subfield of study has gradually developed in breadth and depth, along with the blossoming of qualitative inquiry adopted in almost all fields of social science. Relatively speaking, EQR is seen as cohesive because Lincoln and Guba's (1985) discourse on trustworthiness criteria has been accepted as *the* platform for EQR. Even though these trustworthiness criteria are still considered essential in discussing the quality of qualitative research, different discourses are available.

As talk of paradigm has broadened, the platform for EQR has changed as well (Guba & Lincoln, 1989). Those who operate from post-modern and post-structuralist traditions criticize trustworthiness criteria as another version of traditional or foundational approaches (Scheurich, 1996). Defining *validity* is another issue. Some use the terms "criteria" and "validity" interchangeably, drawing on a philosophical and/or evaluation discourse (Creswell, 2006; Schwandt, 2002; Seale, 1999). Others use validity as a broad epistemological concept to justify an ideal of qualitative inquiry (Lather, 1986). As mentioned earlier, such terms as validity, rigor, trustworthiness, goodness, integrity, and so on are interpreted in many different ways by many different people. Lichtman's (2006, 2009) position on EQR provides a good explanation:

> At this point, I caution you to be careful as you review criteria for judging qualitative research. Several viewpoints are in play. One group contends that we need to return to research that is more scientific, but I believe that is not necessarily the majority viewpoints. Others see the field as still (p. 679) in a state of flux... The climate of the world of educational research is such that there is increased accountability and standardization and control. The field has become more politicized than it once was.... It is not possible, nor is it desirable, to reach any kind of consensus about what standards should be adopted... the field is not unified... reviewers of journals often embrace a kind of generic criteria. Although they review articles in the health field, the points they make are applicable to education....[Al-

> though] the issue of judging, quality, and rigor is very much alive… it is clear that the issue of quality is not yet resolved.
>
> (2006, pp. 231–232)

Considering the field's disparity, as well as the seeming urgent need for some sort of resolution, our sketch of EQR is categorical in pointing out qualitative differences among many different theoretical and practical ideas. We present six categories of EQR: (1) a positivist category; (2) Lincoln and Guba's alternative category; (3) a "subtle-realist" category developed by Hammersley and Atkinson, and Seale; (4) a general EQR category; (5) a category of post-criteriology; and (6) a post-validity category.

We hope these categories are a useful and meaningful way of sketching a broad view of EQR. We see the six categories as a map that one can use to start making sense of EQR. This sketched map is our own, and others may see the field of EQR differently. We interpret the field of EQR as evolving at present because choosing a set of evaluative criteria in and of itself is socially constructed and politically driven in nature. Therefore, these six categories should not be interpreted as either hierarchical or linear. Simply put, each is a distinctly different category relying on its own specific criteria (Cho & Trent, 2006; Tracy, 2010). We would like our six categories to be seen as providing a holistic perspective, one that continues to evolve but still moves forward, addressing the complex nature of qualitative research and bringing new insights as we collectively draw a broader picture of EQR.

A Positivist Category

Quality in qualitative research is multidimensional. If quality in quantitative research requires accuracy, precision, rightness, or directness, then quality in qualitative research requires context, locality, properness, and indirectness in addition to those required in quantitative research. This is mainly because qualitative research is value-laden or at least value-related. To help readers better understand our first category, we start with four goals or criteria that are important to consider in the traditional view of EQR. Simply, advocates of this category see qualitative and quantitative research as the same and so use the same criteria, ones based in quantitative research. In a similar vein, mixed-methods scholars identify a series of evaluation criteria necessary for measuring the product and process of mixed methods research (Dellinger & Leech, 2007; Leech, Dellinger, Tanaka, & Brannagan, 2010; Tashakkori & Teddlie, 2003, 2008). Sale and Brazil (2004) present a review of criteria for critically appraising mixed-methods research. In their review, they give a very comprehensive list of literature that identifies criteria for evaluating quantitative and qualitative methods in terms of the four conventional validity goals: *internal validity, external validity, reliability,* and *objectivity*.

Lincoln and Guba's (1985) Alternative Category

Perhaps the field of EQR would not be as advanced without Lincoln and Guba's (1985) alternative approach to judging qualitative research. This approach is well known and, as noted earlier, is still greatly influencing the discourse on EQR. In addressing the traditional goals or criteria of internal validity, external validity, reliability, and objectivity seen in the first category, Lincoln and Guba propose *credibility, transferability, dependability,* and *confirmability*, respectively. In Table 32.1, we briefly explain these parallel goals (Thomas & Magilvy, 2011, pp. 152–154).

A "Subtle-Realist" Category

The subtle-realist approach is pragmatic in nature. British scholars Hammersley and Atkinson (1995) and Seale (1999) make a strong case for the necessity of compromise between various extremes. Their philosophical stance in this regard lies between idealism and realism, claiming that neither of them properly addresses the continuing tension of contemporary research, particularly in ethnography. Seale notes, "The widespread appeal of alternative conceptions of research is based upon some fundamental dissatisfactions with the scientific world view" (p. 7). Those who reside in this camp of thought believe that quality in qualitative research is "a somewhat elusive phenomenon that cannot be pre-specified by methodological rules" (p. 7). That is, those concerned with quality in qualitative research don't necessarily "give up on scientific aims as conventionally conceived, but also draw on the insights of postscientific conceptions of (p. 680) social research" (Seale, p. x). For them, objectivism is seen as "a resource that can be used productively as an attitude of mind by social researchers" (p. 25). Consequently, the discourse on EQR is not fixed but "open to the possibility that conclusions may need to be revised in the light of new evidence" (p. x). A subtle-realist category that is conceptualized in this pragmatic stance is convergent with the following point of view:

> Criteriology is, at root, an impossible project if it is intended to reflect an internally logical line of argument that simultaneously reconciles philosophical and political positions with the great variety of research practices which people may wish to pursue. The challenge appears to be to construct some general account of what we might hope to find in a good study that is, on the one hand, open enough to include this variety, and, on the other hand, not so loosely specified as to be no value in providing guidance.
>
> (Seale, 1999, p. 47)

Table 32.1 Lincoln and Guba's alternative criteria for evaluating qualitative research

Traditional	Alternative	Key Points
Internal Validity	Credibility	The elements that allow others to recognize the experiences contained within the study through the interpretation of participants' experiences; checking for the representativeness of the data as a whole; member checking involving returning to the participants to ensure that the interpretations of the researcher are accurate representations of participants' experiences; peer debriefing; prolonged engagement
External Validity	Transferability	The ability to transfer research findings from one group to another; thick description used to provide the reader with detailed contextual information; transfer of understanding is believed to occur if both contexts are similar
Reliability	Dependability	When other research follows the decision trail used by the researcher; having peers participate in the analysis process
Objectivity	Confirmability	Self-critical attitude on the part of the researcher about how one's own preconceptions affect the research

The relationship between claim and evidence is a starting point for the subtle-realist approach to EQR. Triangulating data, in itself, cannot warrant the credibility of a research report; although triangulation is useful to consider, subtle realists argue that "member validation offers a method for testing researcher's claims by gathering new evidence" (Seale, 1999, p. 71). The quality of qualitative research results from the degree of members' involvement, whether weak or strong. Thus, openness to the possibility that conclusions may need to be revised in the light of new evidence is determined by the extent to which members are involved in the closeness between evidences and claims.

A General EQR Category

As Seale (1999) noted, a dilemma exists for EQR: the field needs a set of criteria broad enough to include a variety of qualitative research traditions. The field of qualitative research is broad in history, paradigms, theories, and practices. Each qualitative research tradition has its own rationale for quality considerations (Creswell, 2006). Although discipline-specific criteria for these research traditions are available, a majority of the literature on EQR attempts to provide general criteria or validity applicable to qualitative research generally. These attempts are likely to be encountered in many research articles, some of which will be discussed in the next section of this chapter. We define this attempt as belonging to a general EQR category that proposes evaluative guidelines intended to assist reviewers or committee members in judging the quality of qualitative research of any type. It could be seen as too general for some particular types of qualitative research and perhaps too specific for others.

A Category of Post-Criteriology

The post-criteriology category is seen as radical to some extent because those who reside in this category believe that it is neither desirable to use validity or criteria from the conventional positivist standpoint nor even possible to set up predetermined criteria for qualitative research that uncovers complex meaning-making processes. (p. 681)

> Is it possible to devise a set of goodness criteria that might apply to an inquiry regardless of the paradigm within which it was conducted? Or is it the case... that goodness criteria are themselves generated from and legitimated by the self-same assumptions that undergrid each inquiry paradigm, and hence are unique to each paradigm?
>
> (Guba, 1988, p. 16, cited in Smith, 1990, p. 168)

Smith (1990) reviewed three alternative paradigms and criteria—post-empiricism or post-positivism, constructivism, and critical theory—and found an overall regulative ideal for inquiry: “objectivity, solidarity, and emancipation,” (p. 183) respectively. His criticism is focused on the assumption that “each paradigm has dispensed with the idea of an absolutely authoritative foundation for knowledge. This nonfoundationalism greatly complicates the criteria issue” (p. 183). There are at least three points common to these different perspectives. First, there is no possibility that a mechanical decision-making procedure can be applied to distinguish valid from invalid research. Second, methodology or procedures, in and of themselves, are not sufficient for making decisions about the quality of inquiry. Finally, although only briefly noted earlier, an appeal to consistently successful prediction is not a live option, in that none of the three perspectives has done very well in this area.

A Post-Validity Category

Before explaining this last category, clarifying the difference between a general sense of credibility used in qualitative research and the theoretical sense of validity used in this section is needed. All the earlier five categories of EQR are more or less direct, straightforward, or less abstract in suggesting ways of judging quality or goodness criteria on qualitative research. The post-validity category has its roots in Patti Lather's (1986) seminal article, "Issues of Validity in Openly Ideological Research: Between a Rock and a Soft Place," in which she redefines goodness criteria in ways that make evaluation meaningful for value-based research programs such as feminist research, neo-Marxist ethnography, and Freirian empowering research. She argues that for these research programs to be properly assessed, goodness criteria such as triangulation, construct validity, face validity, and catalytic validity must be built into research designs. That is, critical research programs need accurate data credibility, a researcher's systematized reflexivity, respect for participants' interpretation on data (called member-checking), and evidence of participants' consciousness change.

Later, Scheurich's (1996) article, entitled "The Mask of Validity: A Deconstructive Investigation," takes Lather's value-based research programs a step further, arguing that the conventional approach and Lincoln and Guba's naturalistic approach are fundamentally similar. That is, the general techniques Lincoln and Guba invented have the same orthodox voices that originated in the positivist paradigm. Social transformational research is validated in ways that require a celebration of the play of multiplicity and difference in data collection, analysis, and interpretation. All in all, EQR in this regard is subject to locality or contextuality, in which meaning is de- or reconstructed toward social justice.

Different Strategies for EQR

Here, having reviewed our sketch of the six general categorical approaches to EQR, we present a series of common strategies for qualitative research evaluation. From the many possible, we select five major strategies for EQR that are different in form and content from one another. In the first, scholars develop a list of criteria or checklist that follows a series of research procedures. In the second, a professional organization sets a high level of research standards. In the third, a reviewer is provided with a rubric or scoring guide to review a journal article. In the fourth, an analysis tool is used to evaluate key aspects of the process and the product of qualitative research. And in the last, we include a set of criteria against which art-based research and performance studies are evaluated.

Ten Commandments

How does one evaluate dissertation studies or journal articles? We find the following list a very typical set of criteria (Cobb & Hagemaster, 1987). We'll call these the *ten evaluative commandments*:

1. Expertise

2. Problem and/or research question
3. Purpose
4. Literature review
5. Context
6. Sample
7. Data collection
8. Data processing and plans for analysis
9. Human subject
10. Importance to the field

To our knowledge, almost all researchers, scholars, and teachers took an introductory class to learn how to conduct research (Ambert, Adler, Adler, & Detzner, 1995; Burns, 1989; Duncan & Harrop, 2006; Elliott, Fischer, & Rennie, 1999; Forchuk & Roberts, 1993; (p. 682) Greenhalgh, 1997). What students usually learn is that research goes through a process something like problems → questions/purposes → literature review → context/setting → sample/participants → data collection/display/analysis/interpretation → significance of research. Additionally, students learn about the human subject review process. Reviewing a research project in light of typical research procedures and components is common (Popay, Rogers, & Williams, 1998; Yin, 1999). The following review guideline is used in *The Asian Journal of Educational Research and Synergy,* and it highlights a typical research process using *key evaluative criteria* (this journal accepts both quantitative and qualitative research):

General Considerations

1. Importance and interest to the journal's readers
- What does the paper contribute to the field of education?
- Is it significant to the target community?
- Does it present a new and significant contribution to the literature?
- Is it timely and relevant?

2. Originality of the paper
- Is the study innovative? Interesting?

3. What were the author(s) trying to accomplish and were they successful?

Specific Considerations

1. Presentation
- Does the paper present a cohesive argument?
- What is the basic logic of the presentation?
- Are the ideas clearly presented?

2. Writing
- Is the writing concise and easy to follow?

3. Length

• What portions of the paper should be expanded? Removed? Condensed? Summarized? Combined?

4. Title

• Is the title informative?

5. Abstract and introduction

• Do the abstract and introduction accurately reflect the points made in the paper?

6. Literature review

• Are the cited articles/papers current?

• Is the literature review comprehensive?

• Does the literature review contain a coherent argument supported by literature (as opposed to a list of studies)?

7. Methods for studies involving primary data collection

• Does the author provide enough detail of the methodology?

• Are the methods described clearly enough to facilitate replication (where applicable)?

• Is there a sound research methodology?

• Are the methods appropriate?

8. Data presentation

• Could the design be conveyed more easily?

• Are the data clearly presented?

• Can the reported results be verified easily by reference to tables and/or figures?

• Would another form of presentation help?

• Are illustrations instructive?

• Are all tables and figures clearly labeled? Necessary? Well-planned?

9. Analysis and interpretation

• Does the organization of results promote understanding?

• Are the analyses appropriate and logical? Are they described in enough detail?

10. Discussion

• Are the discussion and conclusions made by the author supported by the data?

• Does the writer understand the limitations of his or her work?

- Is there enough breadth and depth in the implications of his or her study?

This detailed guideline is intended to help a reviewer examine a journal article and is similar to the ten evaluative commandments presented earlier. We find two considerations interesting in this guideline: originality and discussion. The discussion part covers conclusion, limitations, and implications, all of which are worth being assessed. The originality part, expressed as *innovative* or *interesting*, is definitely something important for the reviewer to consider. Arguably, those concerned with a general set of criteria are interested in constructing a checklist inherent in logic, specificity, or thoroughness in form and content. In other words, this kind of checklist-type evaluation strategy is appreciated on the grounds that any research can be assessed in a way that follows a linear sense of logic, specificity, and thoroughness. The next is an example of a review checklist by Clive Seale (1999), who wrote a seminal book about evaluating the quality of qualitative research. Seale organizes his major checklist items in terms of introduction (two criteria), methods (five criteria), analysis (six criteria), presentation (six criteria), and ethics (one criterion), (p. 683) along with an additional thirty-six subcriteria following these major criteria:

Criteria for the evaluation of qualitative research papers

1. Are the methods of the research appropriate to the nature of the question being asked?
2. Is the connection to an existing body of knowledge or theory clear?

Methods

3. Are there clear accounts of the criteria used for the selection of subjects for study and of the data collection and analysis?
4. Is the selection of cases or participants theoretically justified?
5. Does the sensitivity of the methods match the needs of the research questions?
6. Has the relationship between fieldworkers and subjects been considered, and is there evidence that the research was presented and explained to its subjects?
7. Was the data collection and record keeping systematic?

Analysis

8. Is reference made to accepted procedures for analysis?
9. How systematic is the analysis?
10. Is there adequate discussion of how themes, concepts, and categories were derived from the data?
11. Is there adequate discussion of the evidence both for and against the researcher’s arguments?
12. Have measures been taken to test the validity of the findings?

13. Have any steps been taken to see whether the analysis would be comprehensible to the participants, if this is possible and relevant?

Presentation

14. Is the research clearly contextualized?
15. Are the data presented systematically?
16. Is a clear distinction made between the data and their interpretation?
17. Is sufficient of the original evidence presented to satisfy the reader of the relationship between the evidence and the conclusions?
18. Is the author's own position clearly stated?
19. Are the results credible and appropriate?

Ethics

20. Have ethical issues been adequately considered?

To elaborate, under the Methods heading, Seale (1999) addresses typical issues related to procedures, such as the selection of subjects, theoretical sampling, the relationship between fieldworkers and subjects, and systematic ways of data collection and record keeping. Under the heading of Analysis, he points out basic steps to follow: data analysis procedures (reliability); a degree of systematic analysis; adequate discussion of themes, concepts, and categories; negative case analysis; validity; and checking meaning with respondents. Last, the heading of Presentation discusses a synthesis of data that indicates context-specific, systematic data display; proper interpretation; evidence-based conclusion; the researcher's position; and credible results. Some subcriteria are: *Could a quantitative approach have addressed the issue better? To what extent are any definitions or agenda taken for granted, rather than being critically examined or left open? Has reliability been considered, ideally by independent repetition? Has the meaning of their accounts been explored with respondents? Are quotations, fieldnotes, etc. identified in a way which enables the reader to judge the range of evidence used? Have the consequences of the research… been considered?*

Research Standards and Descriptive/Prescriptive Rating Scales

A rigorous attempt to identify a set of general checklist criteria embedded in a linear sense of logic, specificity, and thoroughness is clearly evident in the recent publication of the American Educational Research Association's (AERA) (2006) Standards for Reporting on Empirical Social Science Research. AERA uses the word *standards* and organizes its checklist under two overarching themes, *warrantability* and *transparency*. Table 32.2 is an excerpt of the AERA research standards, showing the great emphasis placed on analysis and interpretation.

The general research standards in the left column deal with reliability, analysis methods, inference, and conclusion. The specific standards for qualitative research in the right column are focused largely on analysis and interpretation; they are strongly geared toward

"being transparent" in the process of developing the descriptions, claims, interpretations, evidence that serves as a warrant for each claim, practices used to develop and enhance the warrant for the claims, and interpretive commentary. Presumably, these two core themes, warrantability and transparency, proclaimed by the world's largest educational research association, have significant impact on the qualitative research community in many ways. Warranted claims and transparent procedures could be construed as political in nature (p. 684) and have been used in recent years in the name of scientific, evidence-based research by political conservatives typically thought to oppose the use and funding of qualitative research (Denzin, 2012). However, many qualitative researchers appear to endorse the word *transparency* as a newly emerging and important criterion in conducting and evaluating qualitative research.

Table 32.2 Standards for reporting on empirical social science research

General Research Standards	Qualitative Standards intended to make the process of analysis transparent for reviewers and readers
5.1. The *procedures used for analysis* should be precisely and transparently described from the beginning of the study through presentation of the outcomes. Reporting should make clear how the analysis procedures address the research question or problem and lead to the outcomes reported. The relevance of the analysis procedures to the problem formulation should be made clear.	**5.11**. The *process of developing the descriptions, claims, and interpretations* should be clearly described and illustrated. The description should make it possible to follow the course of decisions about the pattern descriptions, claims, and interpretations from the beginning to the end of the analysis process. Sufficient detail should be included to make the process transparent and engender confidence that the results are warranted.

5.2. *Analytic techniques* should be described in sufficient detail to permit understanding of how the data were analyzed and the processes and assumptions underlying specific techniques (e.g., techniques used to undertake content analysis, discourse or text analysis, deliberation analysis, time use analysis, network analysis, or event history analysis). **5.3**. The analysis and presentation of the outcomes of the analysis should make clear how they *support claims or conclusions* drawn in the research.	**5.12**. The *evidence that serves as a warrant for each claim* should be presented. The sources of evidence and the strength and variety of evidence supporting each claim should be described. Qualifications and conditions should be specified; significant counter-examples should be reported. Claims should be illustrated with concrete examples (e.g., fieldnote excerpts, interview quotes, or narrative vignettes), and descriptions of the social context in which they occurred should be provided. If a warranted claim entails a generalizing statement (e.g., of typicality), it should be supported with evidence of its relative frequency. Speculations that go beyond the available evidence should be clearly represented as such. **5.13**. *Practices used to develop and enhance the warrant for the claims* should be described, including the search for disconfirming evidence and alternative interpretations of the same evidence. Significant limitations due, for instance, to insufficient or conflicting evidence, should be described.

5.4. Analysis and interpretation should include information about any *intended or unintended circumstances* that may have significant implications for interpretation of the outcomes, limit their applicability, or compromise their validity. Such circumstances may include, but are not limited to, key actors leaving the site, changes in membership of the group, or withdrawal of access to any part of the study or to people in the study.	**5.14**. *Interpretive commentary* should provide a deeper understanding of the claims—how and why the patterns described may have occurred; the social, cultural, or historical contexts in which they occurred; how they relate to one another; how they relate to (support or challenge) theory and findings from previous research; and what alternative claims or counter-claims were considered.
5.5. The *presentation of conclusions* should (a) provide a statement of how claims and interpretations address the research problem, question, or issue underlying the research; (b) show how the conclusions connect to support, elaborate, or challenge conclusions in earlier scholarship; and (c) emphasize the theoretical, practical, or methodological implications of the study.	

Table 32.3 is a review form for evaluating AERA annual conference proposals. It addresses the research standards alluded to earlier by specifying warrantability and transparen-

cy. The evaluation contents or criteria of this review form are aligned with general research procedures, just like those of checklists, but they are much more descriptive and prescriptive. The form describes what each research component is like (e.g., perspectives or theoretical framework) and, at the same time, it prescribes what must be expected by a reviewer (e.g., evidence, substantiation or warrants for arguments, and scientific significance). Additionally, it gives a 1–5 rating scale. Typically, reviewers are eventually asked to make a decision. To our knowledge, providing written comments is typical, along with stating a decision that falls within one of four judgmental calls: accepted as is, accepted with minor revision, accepted with major revision, or rejected. The AERA proposal review evaluation form has a binary decision rule—accepted or rejected—and includes comments for both writer and division chair. **(p. 685)**

Table 32.3 AERA annual conference proposal review form	
Objectives or purposes	Min (insignificant) 1 2 3 4 5 Max (Critically significant)
Perspective(s) or theoretical framework	Min (Not well executed) 1 2 3 4 5 Max (Well executed)
Methods, techniques, or modes of inquiry	Min (Not well executed) 1 2 3 4 5 Max (Well executed)
Data sources, evidence, objects, or materials	Min (Inappropriate) 1 2 3 4 5 Max (Appropriate)
Results and/or substantiated conclusions or warrants for arguments/ point of view	Min (Ungrounded) 1 2 3 4 5 Max (Well grounded)
Scientific or scholarly significance of the study or work	Min (Routine) 1 2 3 4 5 Max (Highly original)
Comments to the program chair (This field is mandatory; you must comment)	
Comments to the author/submitter (This field is mandatory; you must comment)	
Reviewer Recommendation Accept ( ) Reject ( )	

Evaluative Rubrics

Table 32.4 is a rubric-type review form for the journal *Multicultural Perspectives*. This journal accepts both quantitative and qualitative work, but mostly includes qualitative research articles.

This evaluation rubric reviews journal articles in the context of multiculturalism (race, gender, ethnicity, etc.); because multicultural education includes several dimensions that deal with *general thematic criteria*, these differ from generally encountered criteria like questions, purposes, literature, analysis, and conclusion. Given a number of different no-

tions of multiculturalism and multicultural education, this journal's evaluation rubric adopts (p. 686) such general thematic criteria as provocative content (new and thought-provoking) and organized/focused, clear/comprehensive, or interesting reading, along with commonly addressed criteria such as significant topic, clear purpose/scope and methods, and appropriateness to the journal. This review rubric, or general thematic rubric, with its nine dimensions/criteria, not only assists reviewers in evaluating broad ranges of research articles submitted to this interdisciplinary journal, but also seeks a high level of article quality by emphasizing strong qualitative evaluation criteria (e.g., "new and thought-provoking").

Table 32.4 A review form used in the *Journal of Multicultural Perspective*

Rating Dimension	Ex	G	M	W	Comments
Significant Topic					
Clear Purpose and Scope					
Provocative Content (new and thought-provoking)					
Analytical (theoretical, empirical, conceptual, philosophical)					
Organized and Focused					
Clear and Comprehensive					

Conclusions Valid					
Interesting Reading					
Appropriate for *Multicultural Perspectives*					
Written Comments					

Directions: Place an “X” for each dimension: *Ex* = Excellent; *G* = Good; *M* = Marginal; *W* = Weak. Jot notes in the “comments” section and incorporate these into the narrative.

Criteria: By, For, and Of the Readers, Participants, and Investigators

In the matter of evaluating content and form, we have thus far examined a series of criteria set forth in checklists, standards, and rubrics. We would like to draw attention to another, different form of evaluation. If the previous strategies and discussions on determining the inclusion of evaluation criteria are straightforward and directive in terms of what qualitative research is like and how it proceeds, then the argument that Stiles (1999) makes is insightful and relational:

> The concept of objectivity is replaced by the concept of *permeability*, the capacity of understanding to be changed by encounters with observations. Investigators argue that we cannot view reality from outside of our own frame of reference. Investigator bias can be reframed as *impermeability*.... Good practice in reporting seeks to show readers how understanding has been changed. The traditional goal of truth of statement is replaced by the goal of *understanding by people*. Thus, the validity of an interpretation is always in relation to some people, and criteria for assessing validity depend on who that person is (e.g., reader, investigator, research participant).
>
> (p. 99; emphasis in original)

To elaborate, according to Stiles (1999), EQR involves two sets of judgments on quality: *good practice criteria* and *validity criteria*. Here, we briefly explain the first: judgmental quality criteria. It is likely that all sorts of criteria mentioned in the previous types of evaluation thus far are convergent with what Stiles refers to as good practice criteria in light of the investigator's choice, sound analytical practices, and disclosures of the investigator's forestructure. Some example criteria include: "Are research questions clearly stated? Are prolonged and persistent observation made? Did the investigator make a disclosure of his or her orientation or assumptions?" (p. 99). These judgmental criteria and their subcriteria are intended to evaluate the degree of what is generally called "credibility" or claims of truthfulness.

What makes Stiles's (1999) strategy unique in the matter of EQR is the "validity criteria" (p. 100) that are mainly concerned with who is impacted by the researchers' interpretations and how the impact of interpretation is utilized and for what purpose. The table of analytic evaluation developed by Stiles is seen in Table 32.5.

The 3×2 grid analysis tool in Table 32.5 involves three different stakeholders and two different purposes of interpretation. For example, if the purpose of interpretation is to determine readers' agreement with regard to what is found in the research, then one major criterion should be *coherence*, which includes follow-up questions like "Is the interpretation internally consistent? Is it comprehensive?... Does it encompass all of the relevant elements and the relations between elements?" (p. 100). If the purpose of interpretation is to make readers rethink their existing belief system, then they should have *revealing* or *self-evident* learning experiences as they read a text. Subquestions related to this level of evaluation include "Is the interpretation a solution to the concern that motivated the

reader's interest?... Did it produce change or growth in the reader's perspective? Did it lead to action?" (p. 100).

At the level of evaluation criteria to be applied to research participants, the major criterion is *testimony*, which allows participants to express their voices from (p. 687) their own perspectives. Follow-up questions are "Did participants indicate that the interpretation accurately described their experience?... Were their reactions to hearing the interpretation consistent with the interpretation's motifs? Did they reveal fresh and deeper material?" (Stiles, 1999, p. 100). *Catalytic validity*, one of Guba and Lincoln's (1989) five authenticity criteria, is used if the purpose of interpretation is to empower the participants' life worlds and to have them "take more control of their lives" (p. 100). This catalytic validity is also more purposefully and critically used in emancipatory social science research (e.g., feminist research, neo-Marxist critical ethnography, and Freirian research). In effect, Lather (1986) radically redefines catalytic validity as indicating "not only... a recognition of the reality-altering impact of the research process itself, but also... the need to consciously channel this impact so that respondents gain self-understanding and, ideally, self-determination through research participation" (p. 67).

Table 32.5 Types of validity in qualitative research

Impact of interpretation on preconceptions or bias		
Group of people	**Fit or agreement**	**Change or growth**
Reader	Coherence	Uncovering; self-evidence
Participants	Testimonial validity	Catalytic validity
Investigators	Consensus; replications	Reflexivity validity

Criteria for Art-Based Research and Performance Studies

In recent years, art-based researchers have proposed six evaluative criteria. Barone and Eisner (2012) note with some critical comment on existing inquiry into EQR that "employing a quantitative metric enables one to enumerate or to summarize quantity.... Criteria [for arts-based qualitative work] are much more slippery" (Barone & Eisner, 2012, p. 147). With specific art-based evaluative criteria in mind, they propose the following set of criteria:

- *Incisiveness*: The degree to which research gets to the core essence of a social issue; Barone and Eisner (2012) assert that incisive research: "offer[s] the potential for wak-

ing the reader up to a strange world that appears new and yet always existed in the shadowy corners of the city that they had never explored on their own" (p. 149)

• *Concision*: The degree to which research occupies the minimal amount of space; "any additional material simply diminishes the capacity of the piece to achieve that purpose, waters down the power of the work, and hence its effectiveness" (pp. 149–150)

• *Coherence:* The creation of a work of arts-based research whose features hang together as a strong form (pp. 150–151)

• *Generativity:* The ways in which the work enables one to see or act on phenomena, even though it represents a kind of case study with an n of only 1 (pp. 151–152)

• *Social significance*: Something that matters, ideas that count, important questions to be raised (p. 153)

• *Evocation and illumination*: Feeling or defamiliarizing an object so that it can be seen in a way that is entirely different from the ways in which customary modes of perception operate (p. 154)

Barone and Eisner (2012) add that these six criteria should be seen as "a cue for perception" (p. 154), one that assists observers or audiences in making a better evaluation of an art product. Therefore, they offer these criteria merely as a starting point for thinking about the appraisal of works of art-based research. Getting locked into criteria that constrain innovation and dampen imagination is undesirable. As with the other scholars mentioned earlier, Barone and Eisner take a deliberative, balanced perspective on EQR. Barone and Eisner assert,

> We do not believe that we can have an effective arts based research program without some degree of common reflection over what might be attended to in looking at such work. Thus, in a certain sense, we compromise between, on the one hand, common criteria and, on the other, criteria that are idiosyncratic to the work itself. This may appear a dilemma, but it is a reality.
>
> (p. 155)

The compromise alluded to is indeed a reality, one that those involved in qualitative research deal with regularly. Seeing qualitative research as art is not new. But following a "recipe" to produce art-based research is like using a recipe to produce a chocolate cake to a particular standard. The problem is that "the more detailed and prescriptive the recipe, the more likely that the cakes made from that recipe will be indistinguishable from one another" (Barone & Eisner, 2012, p. 155). Eventually, Barone and Eisner "invite you, the readers, to use your own judgment in applying these criteria to the examples of the works of arts based research" (p. 155).

Cho and Trent (2009) suggest validity criteria for assessing performance-related studies. Performance is often viewed as an "object" or the presentation of the results of analysis (Hamera & Conquergood, 2006, p. 420). In this view, qualitative researchers think, plan, select, and show through performance their inquiry findings as the last phase of assign-

ment/research project completion. Although Cho and Trent support this traditional role for performance in qualitative research, they claim that their conceptualization is broader, incorporating performance aspects at all stages of the inquiry process. (p. 688) They acknowledge the meaning of performance both in and as qualitative research because the boundary between performance and qualitative research blurs as researchers/teachers and students/audience or researchers and reviewers come to see “conducting qualitative research” as an inevitably personal, social, and political performative process. They advocate for in-depth dialogues and scaffolding to support audiences’ and other researchers’ introduction to the possibility of constructing and utilizing performance in/as qualitative inquiry (Hamera, 2006).

Cho and Trent (2009) offer validity criteria for performance that are critically oriented, culturally responsive, and pedagogically sound. The rubric they construct is not only evaluative but also pedagogical in nature (see Table 32.6). The rubric outlines criteria for all three phases of the performative process: pre-, during, and post-performance.

Pre-performance as imaginative rehearsal is an ongoing textual rehearsal process as the researcher finalizes the analysis and interpretation of the data collected. The focus of imaginative rehearsal is on making the voices of subjects relational and evocative as the researcher constructs texts as scripts. Criteria needed to evaluate this imaginative textual practice involve data sufficiency, level of critical interpretation, and degree of script craftsmanship. The stage of performance-in-use, associated with artistic re/presentation, involves transacting the lived experiences of others with audiences by means of the voices and bodies of the performer(s). One of the main criteria is degree of understandability of the performance being re/presented. With clear delivery in mind, this criterion is one that cautions that some performance is too complex to understand. The post-performance stage is nurtured by a co-reflexive member-checking process among subjects, performers, and audience. It is important to link artistic re/presentation with degrees of intensive experience and closeness between the performer and the audience. Post-performance is seen as a beginning, not an ending, because the effect of a performance on the performer and the audience may be rearranged as both parties share their understandings with one another. The performer should be very clear about his or her rationale for checking validity: Whose authority? Whose artistic achievements? And, whose evaluative validity is of most importance at this time in this place? Which choices promote the primary aim of attaining a deeper, empathic understanding across participants (both performers and audience members)? These co-constructive validity-seeking questions may help audiences reflect critically, not so much on aesthetics at the surface level as on hidden messages underpinning the performance.

Evaluating Qualitative Research: Politics of Evidence for the Twenty-First Century

The criteria for judging a good account have never been settled and are changing.

(Clifford, 1986, p. 9)

The question of whether it is possible to measure the value of qualitative research from the standpoint of conventional evaluation criteria has resurfaced. Those who accept a positivist paradigm assume that reality can be objectively measured. These researchers are reigniting the paradigm wars in ways that repeat old arguments in new form. The new focus of the attack is on the shaky nature of evidence drawn from qualitative research. As has been explored throughout this chapter, much scholarship has been focused on EQR in recent decades. As a consequence, more accurate, meaningful (p. 689) ways of evaluating qualitative research have been established.

Table 32.6 Validity criteria designed to guide the development, enactment, and assessment of dialogical performance of possibilities

	Pre-performance	During-performance	Post-performance
Process	• Imaginative • Textual rehearsal	• Artistic representation • Situated engagement	• Co-reflexive member checking • Caring/empowered/non violent
Major Criteria	• Data-sufficiency • Critical interpretation • Script craftsmanship • Multiple voices • Persuasive • Advocacy	• Aesthetic • Dialogical engagement • Understandability • Improvisational • Empathetic/authentic	• Divergent reactions • Focus on major concrete issues • Generation of possible solutions • Co-construction of further questions • Un/learning about social justice • Promotion of continued conversation and action

Despite the evolution of robust evaluation frameworks, work remains on at least two fronts. Internally, as a community of qualitative researchers, we need to continue to focus on the purposes of our scholarly work and the ways we legitimize it both within and outside our fields. This, necessarily, is a never-ending conversation, and one in which all researchers should participate. Externally, we need to continue to focus on appropriate responses to those who diminish the rigorously obtained knowledge that results from naturalistic inquiry. Those who prioritize only randomized, generalizable work with numerical findings (despite the inherent associated problems) ignore a robust knowledge base that, pedagogically, has often more to offer than a statistical analysis of decontextualized "data." This knowledge base presents in narrative form, as stories, and, as humans and inquirers, it is among our most basic ways of knowing. Unfortunately, as we discuss later in this chapter, those who perpetuate paradigm wars also wield a great deal of power in research and policy communities.

Evaluating Qualitative Research: Moving Forward in Contemporary Contexts

Thus far, our review has illuminated the wide variety of approaches to EQR, including wide-ranging epistemological underpinnings, as well as a broad array of strategies and processes for evaluating qualitative work. Still, despite extant models and frameworks, researchers work in dynamic, always changing contexts—socially, personally, and politically. As noted earlier, there is and always will be a need to continue to examine emergent evaluation prescriptions and proposals and to juxtapose these with contemporary evolutions in context and culture. Richardson's evolving work on this topic provides a good example. Richardson, in 2000, offered five criteria against which to assess the validity/quality of ethnographic texts:

- *Substantive contribution*: Does this piece contribute to our understanding of social life? Does the writer demonstrate a deeply grounded (if embedded) human-world understanding and perspective? How has this perspective informed the construction of the text?
- *Aesthetic merit*: Does this piece succeed aesthetically? Does the use of creative analytical practices open up the text, invite interpretive responses? Is the text artistically shaped, satisfying, complex, and not boring?
- *Reflexivity*: How did the author come to write this text? How was the information gathered? Ethical issues? How has the author's subjectivity, as both a producer and a product of this text, been addressed? Is there adequate self-awareness and self-exposure for the reader to make judgments about the point of view? Do the authors hold themselves accountable to the standards of knowing and telling of the people they have studied?
- *Impact*: Does this affect me? Emotionally? Intellectually? Generate new questions? Move me to write? Move me to try to new research practices? Move me to action?

• *Express a reality*: Does this text embody a fleshed out, embodied sense of lived experience? Does it seem "true" —a credible account of a cultural, social, individual, or communal sense of the "real"? (p. 254)

The evaluative questions listed here concerning ethnographic texts can be applied to judging most qualitative texts. As a journal referee, one must be concerned with the degree of contribution, a sense of aesthetics, the level of a researcher's reflection, the learning of the reader, and indications of credibility. It appears, however, that Richardson's criteria have changed. When writing with St. Pierre in 2005, Richardson and St. Pierre exclude the last criterion, *express a reality*. There is no explanation as to why the last criterion about credibility is no longer included in this later version.

It is typical for research methodologists to offer a set of evaluative criteria that are claimed to be relevant and necessary based on their theoretical underpinnings. Many of the scholars highlighted in this chapter have done so, and these criteria sets illustrate that some criteria are commonly used, whereas other criteria are used uniquely, depending on the different purposes and uses of the evaluation. Yet, by looking at the matter of EQR from a broader perspective, we may end up concluding that EQR, like other theoretical constructions in social science, is simultaneously contextual, cultural, and political. When a reviewer evaluates a manuscript, the process is individualistic, and it is hard to describe the multiple influences impacting the reviewer's perspective. These individualistic and hidden meanings used by a reviewer do not necessarily match neatly with a set of criteria provided by a journal editor, colleague, or conference organizer. Assessment tools in this complex process are used for formality, (p. 690) convenience, and as a standardized means to ensuring fairness in determining contributors. In the end, it is the reviewer's construction of meaning (or lack of) around the text that matters.

By the same token, an inclusion or exclusion of goodness criterion is socially constructed. The earlier noted discrepancy between Richardson (2000) and Richardson and St. Pierre (2005) serves as an example. The omission must be more than random. The co-authors likely included those criteria on which they agreed and co-constructed understandings, and omitted those on which they did not. A consistent theme across both authors' individual and collaborative work is the joining of art and science in the production of qualitative texts. "Science is one lens, and creative art is another. We see more deeply using two lenses. I want to look through both lenses to see 'a social science art form' —a radically interpretive form of representation" (Richardson & St. Pierre, p. 964). Perhaps the qualitative research community accepts these scholars' social science art form, which is similar to what Lather (1986) refers to as "a new rigor of softness… validity of knowledge in process… an objective subjectivity" (p. 78). A social science art form or an objective subjectivity is something that continues to evolve. A constant deliberation on the inclusion and exclusion of criteria in EQR is necessary to better address the changing nature of knowledge and aesthetics in sociocultural contexts.

Evaluation, Criteria, and Power

Scholars continue the conversation about evaluating research. Tracy (2010) presents a recent proposal for a model to ensure "excellent qualitative research." Tracy's model is a solid synthesis of what has been researched and theorized about in recent history. Alternatively, Lichtman's (2006) review of evaluating qualitative research includes *personal criteria*, which are based on her philosophy and assumptions regarding a good piece of qualitative research. Thus, Lichtman attempts to make her personal philosophy explicit by reflecting on the self, the other, and interactions of the self and other. Lichtman argues that "an understanding of the other does not come about without an understanding of the self and how the self and other connect" (p. 192). Then she goes on, "I believe each is transformed through this research process" (p. 192). In contrast, Tracy takes an objective stance in establishing her model's rationale for education establishment power holders:

> In addition to providing a parsimonious pedagogical tool, I hope my conceptualization may aid in garnering respect for qualitative methods from power holders who know little about our work. Despite the gains of qualitative research in the late 20th century, a methodological conservatism has crept upon social science over the last 10 years... evidenced in governmental and funding agencies' preference for research that is quantitative, experimental, and statistically generalizable.... High ranking decision makers—in powerful governmental, funding, and institutional review board positions—are often unprepared and unable to appropriately evaluate qualitative analyses that feature ethnography, case study, and naturalistic data.
>
> (Tracy, 2010, pp. 837–838)

With these pedagogical and political purposes in mind, Tracy (2010) provides eight universal hallmarks for high-quality qualitative methods across paradigms, suggesting that each criterion of quality can be approached via a variety of paths and crafts, the combination of which depends on the specific researcher, context, theoretical affiliation, and project. Her eight "big-tent" criteria for excellent qualitative research are listed in Table 32.7.

We'll examine two of these criterion for clarification: "rich rigor" and "meaningful coherence." The nature of rigor is tricky and difficult for evaluators to define. Rigor in qualitative research differs from that in quantitative research. *Rigor* literally means "stiffness," from the Latin word *rigere,* to be stiff, and it implies rigidity, harshness, strict precision, an unyielding quality, or inflexibility. The term *qualitative rigor,* then, is an oxymoron, considering that qualitative research is "a journey of explanation and discovery that does not lead to stiff boundaries" (Thomas & Magilvy, 2011, p. 151). Thus, the word *rigor* involves many dimensions that must be considered. In qualitative research, rigor often refers to the thorough, ethical conduct of a study of a social phenomenon. We argue that all criteria—rigor and numerous others—used (or considered) in evaluating qualitative research are necessary but may not be sufficient. Tracey's (2010) thesis, therefore, is in line with

the tricky nature of rigor, which also reflects what Richardson mentioned earlier, a wish to have a social science art form of EQR:

> Like all components in this conceptualization—rich rigor is a *necessary but not sufficient* marker of qualitative quality. For qualitative research to be of high quality, it *must* be rigorous. However, a head full of theories and a case full of data does not automatically result in high quality work. Qualitative methodology is as much art as it is effort, piles of data, and time in the field. And just (p. 691) (p. 692) like following a recipe does not guarantee perfect presentation, or completing a vigorous training plan does not guarantee race-day success, rigor does not guarantee a brilliant final product. That being said, rigor does increase the odds for high quality, and the methodological craft skills developed through rigorous practice transcend any single research project, providing a base of qualitative fitness that may enrich future projects.

Table 32.7 Eight “big-tent” criteria for excellent qualitative research

Criteria for quality	Various means, practices, and methods through which to achieve (end goal)
Worthy topic	The topic of the research • Relevant • Timely • Significant • Interesting
Rich rigor	The study uses sufficient, abundant, appropriate, and complex • Theoretical constructs • Data and time in the field • Sample(s) • Context(s) • Data collection and analysis processes
Sincerity	The study is characterized by • Self-reflexivity about subjective values, biases, and inclinations of the researcher(s) • Transparency about the methods and challenges
Credibility	The research is marked by • Thick description, concrete detail, explication of tacit (nontextual) knowledge, and showing rather than telling • Triangulation or crystallization • Multivocality • Member reflections

Resonance	The research influences, affects, or moves particular readers or a variety of audiences through • Aesthetic, evocative representation • Naturalistic generalizations • Transferable findings
Significant contribution	The research provides a significant contribution • Conceptually/theoretically • Practically • Morally • Methodologically • Heuristically
Ethical	The research considers
	• Procedural ethics (such as human subjects) • Situational and culturally specific ethics • Relational ethics • Exiting ethics (leaving the scene and sharing the research)
Meaningful coherence	The study
	• Achieves what it purports to be about • Uses methods and procedures that fit its stated goals • Meaningfully interconnects literature, research questions/foci, findings, and interpretations with each

(Tracy, 2010, p. 841; emphasis in original)

Tracy (2010) uses metaphors of art and recipes to point out that a claim for rigor involves *a closer investigation*. Its promise and limitations coexist. The politics of "being rigorous" is clearly evident in many types of qualitative research. Likewise, techniques to ensure "rigor," such as advanced statistical analyses, do not guarantee brilliant quantitative research, either. It is the perception of reviewers or assessors that decides what makes research "good research." All judgment calls involves a complex mix of relative, contextual, political, and/or ethical criteria. In this regard, "tools, frameworks, and criteria are not value free" (Tracy, 2010, p. 838).

Meaningful coherence, Tracy's final criterion, is accomplished when "the study achieves what it purports to be about, uses methods and procedures that fit its stated goals, and meaningfully interconnects literature, research questions/foci, findings, and interpretations with each" (p. 839). Thus, this criterion is likely to be seen as a summary of overall judgments in a typical evaluation tool.

Tracey's "big-tent" set of criteria is a synthesis of other scholars' constructions of existing goodness criteria. These criteria may usefully remind reviewers about a variety of judgmental aspects in their attempts to determine "how good is good enough," but it is also important to think about the fact that qualitative research "should not be mechanically scored and summed insofar as some issues may be far more important than others in particular studies" (Stiles, 1999, p. 100). In the end, it is necessary to develop some kind of standardized form of evaluative criteria to be used in qualitative research. Such constructions provide us with meaningful evaluation tools or guidelines, aligned with key criteria, which determines the degree of credibility in qualitative research. Yet, is it really possible to develop standardized forms of evaluation applicable to any type of qualitative research? Tracy (2010) thinks it is:

> Perhaps the most controversial part of this conceptualization is the notion of universal criteria for qualitative quality. However, I believe that we need not be so tied to epistemology or ontology (or the philosophy of the world) that we cannot agree on several common end goals of good qualitative research. Qualitative methodologists range across postpositivist, critical, interpretive, and poststructural communities. In contrast,... researcher reflexivity is a validity procedure clearly positioned within the critical paradigm where individuals reflect on the social, cultural, and historical forces that shape their interpretation..., I would argue instead that researcher reflexivity—like many other practices for goodness—serves as an important means toward sincerity for research in a number of paradigms. Its utility need not be bound only to critical research.
>
> (Tracy, 2010, p. 849)

Nonetheless, we find that some prestigious qualitative journals don't provide these kinds of criteria or guidelines for their reviewers. Instead, reviewers invited by these journals are provided with very general guidelines. Table 32.8 is an example of the *International Journal of Qualitative Studies in Education* (QES) review form.

As Table 32.8 shows, there are no specific criteria to be used in reviewing manuscripts in this prestigious qualitative research journal. Nonetheless, editorial manager Gonzalez (2012) has confidence in this open process: “Reviewers are free to send any comment to the author. We have very strong scholars to agree to review and most of the times, our reviewers are very detailed (without asking them) in their reviews from grammar, to format, to content... many of them go and make comments to each section of the manuscript (intro, methodology, results, conclusions)” (personal communication, August 16, 2012). In this review process, what we find is a sense of autonomy, fit, trust, and professional ethics. Reviewers who have expertise (p. 693) know what is worth assessing and how good is good enough.

Table 32.8 A review form used in *QSE: The International Journal of Qualitative Studies in Education*

Recommendation

( ) Accept
( ) Accept with minor revisions
( ) Accept with major revisions
( ) Reject and encourage resubmission
( ) Reject

Would you be willing to review a revision of this manuscript?

( ) Yes ( ) No

Comments (Confidential Comments to Editors)
Comments to the Author

Yet, there are external forces that question not only quality in qualitative research but also its legitimacy. For example, mixed-methods scholars and researchers try not to see themselves as post-positivists in the research paradigms that have been well established over the past several decades (Guba & Lincoln, 1994; Lincoln & Guba, 2000, 2005; Lincoln, Lynham, & Guba, 2011) but instead seek to create their own hybrid epistemology, one that they prioritize over qualitative research. The current neo-conservative initiatives —the National Research Council (NRC) or the Society for Research on Educational Effectiveness (SREE) (see Denzin, 2009, for more detail)—diminish the tradition of qualitative inquiry that values understanding in human science by narrowly defining what research is and how it should be assessed. Denzin (2009) points to the necessity of casting big-tent criteria to evaluate qualitative research in the context of a changing epistemological and political context:

> [W]e must expand the size of our tent, indeed we need a bigger tent! We cannot afford to fight with one another. Mixed-methods scholars have carefully studied the many different branches of the poststructural tree.... The same cannot be said for the poststructuralists. Nor can we allow the arguments from the SBR [Scientif-

> ically Based Research] community to divide us. We must learn from the paradigm conflicts of the 1980s to not over-reach, to not engage in polemics, to not become too self-satisfied. We need to develop and work with our own concepts of science, knowledge and quality inquiry. We need to remind the resurgent postpositivists that their criterion of good work applies only to work within their paradigm, not ours.
>
> (pp. 32–33)

As implied here, current discourse on the politics of evidence is mostly a resurrection of old-fashioned epistemological debates, which are initiated from several organizations or councils at the national level in the United States (e.g., NRC, SREE, the Cochrane Clearinghouse, the Campbell Methods Group, or the What Works Clearinghouse). These trends are generally called scientifically based Research (SBR) or evidence-based movement (EBM). The extended discussion goes beyond the scope of this chapter. The main epistemological questions that need to be asked, just as they were forty years ago, are: “Whose science? Whose scientific principles?” (Denzin, 2009, p. 141). Related to the inquiry of this chapter, we ask, “Whose criteria?”

Tracy’s (2010) eight “big-tent” criteria and the QSE’s simple scholarly decision recommendation with its open-ended comments are two extreme approaches within our qualitative research community. Those situated in the positivist epistemology and mixed-method scholars will likely prefer Tracy’s (2010) “big-tent” criteria for excellent qualitative research over the QSE’s simple form. This is not because Tracy’s reconstruction of other scholars’ constructions is absolutely truthful or valid in itself, but because Tracy approaches it procedurally, in terms of a logical flow of what a reviewer needs to do. The beauty of “big-tent” procedural criteria is that it is normative, to the extent that a reviewer should not disregard the work of an author due to a disagreement with the author’s epistemology. This also applies to the other evaluative extreme, such as the QSE’s recommendation sheet with open-ended comments, in which a reviewer has the freedom to make a scholarly judgment. In our opinion, the current debate on the politics of evidence is too heavily focused on ideology while giving too little attention to ethical concerns.

Conclusion

> In demonstrating methodological excellence, we need to take care of ourselves in the process of taking care of others. The most successful researchers are willingly self-critical, viewing their own actions through the eyes of others while also maintaining resilience and energy through acute sensitivity to their own well-being.
>
> (Tracy, 2010, p. 849)

Lincoln and Guba’s (1985) constructivist criteria to evaluate our qualitative research processes and products started a rich conversation and decades of scholarship designed to hone and refine those criteria initially proposed and to discover increasingly rich and

creative ways to address the challenge of evaluation. Lincoln and Guba argue that trustworthiness is always negotiable, not being a matter of final proof whereby readers are forced to accept an account. Therefore, the field of EQR is not an oxymoron. Much has been known about the nature of evaluative criteria in qualitative research. Some propose a general set of criteria, whereas others focus on specific sets of criteria. This chapter identifies six categories of EQR: (1) a positivist category, (2) Lincoln and Guba's alternative category, (3) a "subtle-realist" category, (4) a general EQR category, (5) a category of post-criteriology, and (6) a (p. 694) post-validity category. As seen in many strategies or examples of EQR, some make lists of questions about what is commonly expected in assessing the process and product of qualitative research, whereas others select key validity criteria against which an essence of qualitative research is identified, discussed, and evaluated. Still others adopt broad criteria across these different approaches to construct a comprehensive framework.

As evidence in educational research has continued to be more narrowly defined, many qualitative researchers propose clear counterarguments. Efforts will continue in the search for evaluative criteria from inside the qualitative research community, and the field of EQR will continue to grow, theoretically and practically.

What future directions can we expect for EQR in the twenty-first century? As discussed in this chapter, evaluating qualitative research is complex, challenging, and exciting all at the same time. What matters most is accepting this dilemma, celebrating the reality, and creating a holistic storyline (or a common playful intellectual ground) intended to invite those who have diverse backgrounds to bring different evaluative tools toward constructing flexible but firm evaluation theory, policy, and practice. The qualitative research community may do well to pay close attention to Barone and Eisner's (2012) compromise between common and unique criteria. The beginning of this holistic story has been written, and we hope that others jump in to constructively compete in searching for a common ground in evaluating qualitative research. One of authors of this chapter writes (Cho, 2010):

> The shape of a hexagon is naturalistic. Beehives, snowflakes, or molecules are some examples that can be found in nature. We like this hexagon shape just because it seems to represent a balance. A triangle implies a sense of absolute stability or a function of geometric equilibrium. A hexagon shows a sense of balance or harmony particularly when it is connected with others. It looks complicated and messy at a distance but patterned and fabricated when closely seen. Imagine that bees constantly move around the surface of beehive. A beehive is constructed in compactly connected hexagon shapes as bees diligently work with beeswax from their bodies. This analogy can lead qualitative researchers to be more creative in their practical engagement with validity. The shape of a hexagon is unique in that it leads to harmony and balance as it is tightened from, and connected to each other. (p. 4)

EQR is more than a sum of its parts. It goes beyond creating a set of checklists or recipes. Furthermore, it is more than paradigmatic idiosyncrasy. It should be holistic in nature. Our holistic approach to EQR doesn't seek a complete sense of convergence. Instead, it leaves some room, some unknown territory that may never be reached by the researcher. Like a bee that intuitively and holistically dances around and filters pollen into beeswax to construct a hive, a reviewer deeply imbibes both the process and the product of qualitative research to clearly ensure acceptable quality. Twenty-first century criteria that we support include (1) thought-provoking ideas, (2) innovative methodologies, (3) performative writing, and (4) global ethics and justice-mindedness. Riessman's (2008) reflection on truths and cautions is our ending in a new beginning:

> I prefer not to think in terms of standards or criteria, and warn students away from the "paradigm warfare" that exists out there in the literature. It can paralyze and... simplify what are complex validation and ethical issues all investigators face.... Narrative truths are always partial—committed and incomplete. (pp. 185, 186)

References

Altheide, D., & Johnson, J. (1994). Criteria for assessing interpretive validity in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485–499). Newbury Park, CA: Sage Publications.

Altheide, D., & Johnson, J. (2011). Reflections on interpretive adequacy in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *The SAGE Handbook of qualitative research* (4th ed.) (pp. 581–594). Thousand Oaks, CA: Sage Publications.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *American Educational Research Association Educational Researcher*, ***35***(6), 33–40.

Ambert, A., Adler, P., Adler, P., & Detzner, D. (1995). Understanding and evaluating qualitative research. *Journal of Marriage and Family*, ***57***(4), 879–893.

Barone, T., & Eisner, E. (2012). *Arts based research*. Thousands Oak, CA: Sage Publications.

Burns, N. (1989). Standards for qualitative research. *Nursing Science Quarterly*, ***2***(1), 44–52.

Cho, J. (2010). Searching for "good" data: A reflection on validity-in-practice with double-edged problems. Division D - Qualitative Inquiry Invited Symposium: *Working validity: Transactional, transformational, and holistic approaches to qualitative inquiry*. Paper presented at AERA (American Educational Research Association), May, Denver, CO.

Cho, J., & Trent, A. (2006). Validity in qualitative research revisited. *Journal of QualitativeResearch*, *6*(3), 319–340. (p. 695)

Cho, J., & Trent, A. (2009). Validity criteria for performance-related qualitative work: Toward a reflexive, evaluative, and co-constructive framework for performance in/as qualitative inquiry. *Qualitative Inquiry*, *15*(6), 1013–1041.

Clifford, J. (1986). Introduction: Partial truths. In J. Clifford & G. Marcus (Eds.), *Writing culture: The poetics and politics of ethnography* (pp. 1–26). CA: University of California Press.

Cobb, A., & Hagemaster, J. (1987). Ten criteria for evaluating qualitative research proposals. *Journal of Nursing Education*, ***26***(4), 138–143.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3rd ed.). Thousands Oak, CA: Sage Publications.

Creswell, J. (2006). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Dellinger, A., & Leech, N. (2007). Toward a Unified Validation Framework in Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(4), 309–332.

Denzin, N. (2009). The elephant in the living room: Or extending the conversation about the politics of evidence. *Qualitative Inquiry*, ***9***(2), 139–160.

Denzin, N. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, ***6***(2), 80–88.

Denzin, N., & Giardian, M. (2009). Introduction. In N. Denzin & M. Giardian (Eds.), *Qualitative inquiry and social justice* (pp. 11–50). Walnut Creek, CA: Left Coast Press, Inc.

Denzin, N. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, ***6***(2), 80–88.

Duncan, S., & Harrop, A. (2006). A user perspective on research quality. *International Journal of Social Research Methodology*, ***9***(2), 159–174).

Elliott, R., Fischer, C., & Rennie, D. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, ***38***(3), 215–299.

Forchuk, C., & Roberts, J. (1993). How to critique qualitative research articles. *Canadian Journal of Nursing Research*, ***25***(4), 47–56.

Greenhalgh, T. (1997). Assessing the methodological quality of published papers. *British Medical Journal*, ***315***(7103), 305–308.

Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park: Sage Publications.

Guba, E., & Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Newbury Park, CA: Sage Publications.

Hamera, J., & Conquergood, D. (2006). Performance and politics: Themes and arguments. In D. S. Madison & J. Hamera (Eds.), *The Sage handbook of performance studies* (pp. 419–425). Thousand Oaks, CA: Sage Publications.

Hamera, J. (2006). Introduction: Opening opening acts. In J. Hamera (Ed.), *Opening acts: Performance in/as communication and cultural studies* (pp. 1–10). Thousand Oaks, CA: Sage Publications.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice* (2nd ed.). London: Routledge.

Lather, P. (*1986*). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange*, ***17***(4), 63–84.

Leech, N., Dellinger, A., Tanaka, H., & Brannagan, K. (2010). Evaluating mixed research studies: A mixed methods approach. *Journal of Mixed Methods Research*, *4*(1), 17–31.

Lichtman, M. (2006). *Qualitative research in education: A user's guide*. Thousands, CA: Sage Publications.

Lichtman, M. (2009). *Qualitative research in education: A user's guide (2nd ed)*. Thousands, CA: Sage Publications.

Lincoln, Y., & Guba, G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.

Lincoln, Y. (1995). 'Emerging Criteria for Quality in Qualitative Research and Interpretive Research', *Qualitative Inquiry*, *1*(3), 275–289.

Lincoln, Y., & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed.) (pp. 163–188). Thousand Oaks, CA: Sage Publications.

Lincoln, Y., & Guba, E. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. Lincoln (Eds.), *The SAGE Handbook of qualitative research* (3rd ed.) (pp. 191–216). Thousand Oaks, CA: Sage Publications.

Lincoln, Y., Lynham, S., & Guba, E. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. Denzin & Y. Lincoln (Eds.), *The SAGE Handbook of qualitative research* (4th ed.) (pp. 97–128). Thousand Oaks, CA: Sage Publications.

Patton, M. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Patton, M. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousands Oak, CA: Sage Publications.

Popay, J., Rogers, A., & Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*, ***8***(3), 341–351.

Richardson, L. (2000). Writing: A method of inquiry. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed.) (pp. 923–948). Thousand Oaks, CA: Sage Publications.

Richardson, L., & St. Pierre, E. A. (2005). Writing: A method of inquiry. In N. Denzin & Y. Lincoln (Eds.), *The SAGE Handbook of qualitative research* (3rd ed.) (pp. 959–978). Thousand Oaks, CA: Sage Publications.

Riessman, C. (2008). *Narrative methods for the human sciences*. LA: Sage Publications.

Sale, J., & Brazil, K. (2004). A strategy to identify critical appraisal criteria for primary mixed-method studies. *Quality & Quantity*, *38*, 351–365.

Sale, J., Lohfeld, L., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and Quantity*, ***36***, 42–53.

Scheurich, J. J. (1996). The masks of validity: A deconstructive investigation. *International Journal of Qualitative Studies in Education*, ***9***(1), 49–60.

Schwandt, T. (2002). *Evaluation Practice Reconsidered*. New York: Peter Lang.

Seale, C. (1999). *The quality of qualitative research*. London: Sage Publications.

Seale, C. (2002). Qualitative issues in qualitative inquiry. *Qualitative Social Work*, ***1***(1), 97–110.

Smith, J. (1990). Alternative research paradigms and the problem of criteria. In E. Guba (Ed.), *The paradigm dialog* (pp. 167–187). Newbury, CA: Sage Publications.

Stiles, W. (1999). Evaluating qualitative research. *Evidence-based Mental Health*, ***2***(4), 99–101.

Tashakkori, A., & Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed (p. 696) model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671–702). *Thousand Oaks, CA*: *Sage Publications*.

Tashakkori, A., & Teddlie, C. (2008). Quality of inference in mixed methods research: Calling for an integrative framework. In M. Bergman (Ed.), *Advances in mixed research: Theories and applications* (pp. 101–119). London: Sage Publications.

Thomas, E., & Magilvy, J. (2011). Qualitative rigor or research validity in qualitative research. *Journal of Specialists in Pediatric Nursing*, ***16***, 151–155.

Tracy, S. (2010). Qualitative quality: Eight “Big-tent” criteria for excellent qualitative research. *Qualitative Inquiry*, *16*(10), 837–851.

Yin, R. (1999). Enhancing the quality of case studies in health services research. *Health Services Research*, ***345***(5), 1209–1224.

Jeasik Cho

Jeasik Cho, Department of Educational Studies, University of Wyoming

Allen Trent

Allen Trent, College of Education, University of Wyoming