

Evaluating qualitative research

Qualitative research, like all scientific research, consists of comparing ideas with observations. In good research, the ideas are thereby changed—strengthened, weakened, qualified, or elaborated. Criteria for evaluating qualitative research focus both on the process and on the product—that is, on the research methods that are used and on the changed ideas themselves (the interpretation).

Many qualitative investigators explicitly reject the possibility of absolute objectivity and truth. 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. Instead, good practice in research seeks to ensure that understanding is permeated by observation. Investigator bias can be reframed as *impermeability* (interpretations not permeated by empirical observations). Good practice in reporting seeks to show readers how understanding has been changed. The traditional goal of truth of statements is replaced by the goal of *understanding by people*. Thus, the validity of an interpretation is always in relation to some person, and criteria for assessing validity depend on who that person is (eg, reader, investigator, research participant).

Qualitative research differs from traditional quantitative research on human experience in several ways. Results are typically reported in words rather than primarily in numbers. This may take the form of narratives (eg, case studies) and typically includes a rich array of descriptive terms, rather than focusing on a few common dimensions or scales. Investigators use their (imperfect) empathic understanding of participants' inner experiences as data. Events are understood and reported in their unique context. Materials may be chosen for study because they are good examples rather than because they are representative of some larger population. Sample size and composition may be informed by emerging results (eg, cases chosen to fill gaps; data gathering continued until new cases seem redundant). Emancipation or enhancement of participants may be considered as a legitimate purpose of the research. As a consequence of these characteristics, interpretations are always tentative and bound by context.

The outline that follows was based on a review of how qualitative research on human experience was being conducted and reported.¹ It includes lists of evaluative criteria for assessing (a) good practice in conducting the research and (b) validity of interpretations. These lists have been drawn from many sources,¹ and overlapping lists have been published by others.²⁻⁸ I encourage readers to consult these complementary sources. I have tried to make the lists inclusive, but they are neither exhaustive nor mutually exclusive, nor do all items apply equally to every piece of qualitative research. The term "qualitative research" refers to method rather than topic, and these lists may be understood differently in the context of research on different topics.

Good practice criteria

Like all research reports, qualitative research reports should clearly describe and justify the investigator's choices:

- *Are the study's questions or topics clearly stated?* Qualitative studies may not begin with specific hypotheses; however, the domain of inquiry and the study's goals should be clearly stated in the introduction.

- *Is the selection of participants or materials clearly justified?* Unlike statistically based studies, qualitative studies may focus on informative exemplars (good examples) rather than representative samples. Selection of later cases may depend partly on observations of earlier cases. In the report, the bases for selection and the process by which these bases changed in the course of the study (if they changed) should be explicit.

- *Are the methods for gathering and analysing observations clearly described?* Descriptions of data gathering procedures should be sufficiently detailed to permit replication. Because analytic procedures are less standardised than in statistical hypothesis testing studies, descriptions of qualitative analytical procedures may need to be relatively more detailed, particularly addressing the points noted below.

In addition to these general principles, a more specific canon of good practice in qualitative research aims to enhance permeability and help readers to assess how well observations have permeated investigators' understanding. The following criteria concern analytical practices that enhance permeability:

- *Engagement with the material:* did procedures include intense personal contact with participants? Intimate familiarity with a text? Prolonged and persistent observation? Discussion of interpretations with other investigators or with participants? Checking participants' reactions? Seeking disconfirming data?

- *Iteration:* did investigators cycle between interpretation and observation, repeatedly reformulating and examining revised interpretations in light of further observation or examination of evidence?

- *Grounding:* were there systematic procedures for linking (relatively abstract) interpretations with (relatively concrete) observations? Were clear examples presented?

- *Asking "what," not "why."* In interviews, did investigators seek information that participants had (eg, what they experienced)? Participants' interpretations (eg, theories about the causes of their experience) are sometimes of interest, but they do not substitute for investigators' interpretations.

Another set of criteria concern reporting practices that help readers to assess permeability. By knowing the personal and social context of the study, readers can make adjustments for differing preconceptions (biases) and can assess how well the observations permeated the interpretations.

Types of validity in qualitative research

Group of people	Impact of interpretation on preconceptions or bias	
	Fit or agreement	Change or growth
Readers	Coherence	Uncovering; self evidence
Participants	Testimonial validity	Catalytic validity
Investigators	Consensus; replication	Reflexive validity

● *Disclosure of investigators' forestructure*: did the report reveal the investigators' initial orientation? Preconceptions or expectations for the study? Values? Preferred theories? Relevant personal background?

● *Explication of social and cultural context*: did the report examine assumptions shared between investigators and participants? Relevant cultural values? Circumstances of data gathering? Meaning of the research to the participants?

● *Description of investigators' internal processes*: did the report describe the investigators' personal experience during the investigation? Relationships with participants? Personal impact of findings?

Validity criteria

As the table shows, criteria for evaluating interpretations based on qualitative investigations can be cross classified: (a) according to whether the interpretation's impact is on the readers of the research report, on the research participants, or on the investigators and (b) according to whether the impact is one of simple fit versus one of change or growth in understanding. This focus on the impact of research interpretations on the understanding of specific people is not a rejection or replacement of traditional validity criteria, which may have powerful effects on people's evaluations of a study's interpretations. Instead it represents a recognition that people's understanding may be also be affected by other qualities of the work.

Of course, no one criterion is, by itself, an adequate test of an interpretation's validity. Meeting any one criterion to some degree implies only that, all else equal, an interpretation is somewhat more trustworthy than if it had not met that criterion to that degree. Not every criterion applies to every study, but convergence across several perspectives and types of validity (table) may represent a stronger claim than does any one alone. Such convergence is sometimes called *triangulation*, seeking information from multiple data sources, multiple methods, and multiple earlier theories or interpretations, to arrive at an evaluation of an interpretation's trustworthiness.

Criteria that must be judged from the reader's perspective include:

● *Coherence*: is the interpretation internally consistent? Is it comprehensive; that is, does it encompass all of the relevant elements and the relations between elements? Will it be useful in encompassing new elements as they come into view? A better interpretation encompasses its rivals—confirming, supplementing, elaborating, simplifying, or superseding them.

● *Uncovering; self evidence*: is the interpretation a solution to the concern that motivated the reader's interest (ie, did it "uncover" something that was previously hidden or unknown)? Did it produce change or growth in the reader's perspective? Did it lead to action? Did the interpretation feel right in the context of the reader's other knowledge and beliefs (ie, did it seem self evident to readers after they read it)? Interpretations may be presented in some form to research participants or may be negotiated with participants. Their reactions bear on

the validity of the interpretations. Criteria that reflect the research participants' reactions to the interpretations include:

● *Testimonial validity*: did participants indicate that the interpretation accurately described their experience? For example, did they make direct or indirect allusions to feeling understood? Were their reactions to hearing the interpretation consistent with the interpretation's motifs? Did they reveal fresh and deeper material?

● *Catalytic validity*: did the research process reorient, focus, and energise participants? A catalytically valid interpretation produces change or growth in the people whose experience is being described. Were the participants empowered by the interpretation or the research process? Did they subsequently take more control of their lives? Criteria that reflect the impact of the study on the investigator and on the theory that motivated the study include:

● *Consensus; replication*: did multiple investigators who were familiar with the observations (eg, members of the research team; external reviewers or auditors) find the proposed interpretation convincing? Were the conclusions based on formal rules of evidence? Did they fit with widely accepted exemplars? Were the findings replicated? Note that replication always involves judgments, insofar as no event is ever repeated exactly; successful replication reflects an investigator's judgment that an interpretation encompasses new observations as well as previous ones.

● *Reflexive validity*: did the observations change the investigator's understanding or the theory (ie, did the study reflexively affect the initial understanding)? Was the resultant understanding different from the forestructure; was there evidence of permeability in the investigators' understanding? New ideas and goals emerge from a living theory as it encounters new data and is acted upon by new minds. A theory that becomes rigid (impermeable) and can no longer support this kind of dialectical interaction and change is scientifically dead.

How good is good enough?

As in qualitative research itself, summary judgments about the quality of qualitative studies depend not on the number of criteria met but on the importance and balance of multiple criteria. Checklist items, such as those that the appendix shows, may usefully remind reviewers about specific criteria, but they should not be mechanically scored and summed in so far as some issues may be far more important than others in particular studies. The broader criteria are the trustworthiness of the method and interpretation. After considering the methods, observations, and interpretations in light of the foregoing criteria, how well can readers trust the methods to have adequately exposed the investigators' ideas to empirical observations and how well can they trust the interpretations to improve people's understanding of the phenomena that were investigated?

WILLIAM B STILES, PhD
Department of Psychology
Miami University
Oxford, Ohio, USA

Please address correspondence to William B. Stiles, Department of Psychology, Miami University, Oxford, OH 43056, USA. Email stileswb@muohio.edu.

Thanks to Beth Harrick, Roger Knudson, and the Qualitative Research Group at Miami University's Department of Psychology for comments on a draft of this paper. An earlier version was presented at the Society for Psychotherapy Research meeting, Pittsburgh, Pennsylvania, June, 1993.

1 Stiles WB. Quality control in qualitative research. *Clinical Psychology Review* 1993;13:593–618.

- 2 Altheide DL, Johnson JM. Criteria for assessing interpretive validity in qualitative research. In: Denzin NK, Lincoln YS, editors. *Handbook of qualitative research*. Thousand Oaks, California: Sage, 1994:485–99.
- 3 Elliott R, Fischer CT, Rennie DL. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Br J Clin Psychol*; in press.
- 4 Giacomini MK, Cook DJ. A user's guide to qualitative research in health care: Part II: what are the results and how do they help me care for my patients? Unpublished manuscript, Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Ontario, Canada, 1998.
- 5 Guba E, Lincoln Y. *Fourth generation evaluation*. Newbury Park, California: Sage, 1989.
- 6 Kvale S. *Interviews: an introduction to qualitative research interviewing*. Thousand Oaks, California: Sage, 1996.
- 7 Mishler EG. Validation in inquiry-guided research: the role of exemplars in narrative studies. *Harvard Educational Review* 1990;**60**:415–42.
- 8 Packer MJ, Addison RB. Evaluating an interpretive account. Packer MJ, Addison RB, editors. *Entering the circle: hermeneutic investigation in psychology*. Albany, New York: State University of New York Press, 1989: 275–92.

Appendix

Evaluating qualitative research

Worksheet

Article authors and title:

A. Good practice criteria

Clarity of description and justification of:

Study questions or topics

Low 1 2 3 4 5 6 7 High NA

Selection of participants or materials

Low 1 2 3 4 5 6 7 High NA

Methods for gathering and analysing observations

Low 1 2 3 4 5 6 7 High NA

Analytical practices that enhance permeability:

Engagement with the material

Low 1 2 3 4 5 6 7 High NA

Iteration

Low 1 2 3 4 5 6 7 High NA

Grounding

Low 1 2 3 4 5 6 7 High NA

Asking "what," not "why"

Low 1 2 3 4 5 6 7 High NA

Reporting practices for assessing permeability:

Disclosure of investigators' forestructure

Low 1 2 3 4 5 6 7 High NA

Explication of social and cultural context

Low 1 2 3 4 5 6 7 High NA

Description of investigators' internal processes

Low 1 2 3 4 5 6 7 High NA

B. Validity criteria

Impact on the readers:

Coherence

Low 1 2 3 4 5 6 7 High NA

Uncovering; self evidence

Low 1 2 3 4 5 6 7 High NA

Impact on the research participants:

Testimonial validity

Low 1 2 3 4 5 6 7 High NA

Catalytic validity

Low 1 2 3 4 5 6 7 High NA

Impact on the investigators and the theory:

Consensus; replication

Low 1 2 3 4 5 6 7 High NA

Reflexive validity

Low 1 2 3 4 5 6 7 High NA