Establishing Validity

Introduction to Validity

As the researcher puts together a dataset based on interviews, observations, and other material, she needs to keep in mind the criteria that are used to demonstrate validity of the research.

When we talk about validity, we refer how much an empirical claim from the data actually reflects the meaning of the concept we are studying, as agreed by the community of researchers. Of course, like most subjects, validity can be quite complex and has been defined as having four aspects: face, criterion related, construct, and content validity. I urge you to check out these distinctions in a standard research methods text, for example, Earl Babbie, The Practice of Social Research.

We are not only looking for accuracy or validity but also reliability. Reliability refers to the capacity of the data collection method or tool to consistently get the same results. Conversational interviewing, however, while designed to study a particular topic or set of topics, is not necessarily designed to get the same answers from every respondent.

Although qualitative methods are increasingly being used by researchers across a wide range of fields, persistent skepticism remains about the reliability as well as validity of qualitative data. This skepticism is especially strong among some who rely on controlled experiments and quantitative data for their own research, unfamiliar with the techniques for producing reliable and valid qualitative data.

Thus, before we turn to some of the particular techniques of conversational interviewing, understanding how validity is established in qualitative research is important, both for our own skill in using qualitative methods, but also for our communication with other scholars both within and outside of our own fields of study.

We will use five criteria for establishing validity for qualitative research:

1. Descriptive Validity
2. Interpretative Validity
3. Theoretical Validity
4. Generalizability
5. Evaluative Validity

(1) Descriptive validity

Descriptive validity refers to the researcher's successful capture and representation of the facts on the ground.

Has the research been complete and honest in the representation of what he observed?

Would any other observer present in the scene have described the same interactions, the same statements, and effect the same facts?

Is there other evidence to back up the researcher's description?

All other categories of validity depend on descriptive validity. And consequently, robust data collection and data gathering is the foundation, the essential necessary condition, the first step and basic criterion for all qualitative research.

It's impossible to overstate this, the need for detailed, accurate description, as it also creates additional legitimacy for conversational interviewing, rather than survey research or prescriptive face-to-face interviews.

Failures to achieve descriptive validity are frequently the result of both insufficient data gathering and insufficient reporting of what was observed, regardless of whether the researcher is conducting interviews or doing participant observation fieldwork.

Failures to achieve descriptive validity do not, however, necessarily come about through active dishonesty or when researchers may make mistakes or misremember events.

More often, descriptive validity is compromised by the insufficiency of the reported data.

Descriptions must be as detailed as possible, which is often associated with the phrase "thick description." One test for descriptive validity is whether different observers or participants in the same event produce similar accounts of that event.

Does everyone present or observing report the same actions and conditions?

Here we are not talking about interpretations or evaluations, but simply who said what, when. What was the arrangement or movements of the people in the spaces, the sequences of action?

Let's elaborate on this by thinking about what we do when observers of a scene, an event, or an organization, for example, do produce different accounts.

Researcher A produces an account of a group meeting. Researcher B visits the same group at another time and finds something different and even perhaps contradictory. Thus questions arise about the descriptive validity of A's account or the validity of B's account of the group meeting.

How do we make sense of this difference? Did the group change its activities between the visits?

Was the first observation an outlier, different than the usual activities?

Did the researchers pay attention to different parts of the meeting?

These questions can sometimes be resolved by examining the time and context of the observations and other knowable conditions that might have affected the setting when it was observed or the selection of actions described if they were not all that actually took place.

The differences must, however, be reconciled, because the social scientist seeks to produce as complete and accurate a representation of the phenomena--in this example, the group meeting, as can possibly be accomplished.

Perhaps the two different observations were not produced from observations at different times but from different persons participating or observing at the same time.

It is possible that even in the same interaction, the perspective, the position of the person, may obscure what another can see. Again, the descriptions must be factual, not evaluative, and must be reconcilable.

(2) Interpretative Validity

Social (and management) scientists are not concerned, however, only with descriptions of objects, situations, and interactions. They are interested in what these objects and situations mean to the participants being studied.

In this example, the social scientists would be interested in what participants in the group meetings experience and how these meanings organize their actions.

This is as essential to qualitative research as is the description of persons, groups, and organizations.

Most people engage in qualitative research primarily to uncover the meanings that circulate in the local culture. Beyond describing what people are doing and saying, the researcher wants to know what is the participant's perspective or point of view.

What does this activity, this relationship, or this meeting, or phenomena mean to the actors?

When we use the word "meaning," we are talking about the understanding, the intentions, the cognition, affect, belief, evaluation, and anything else that could be encompassed by what is broadly termed the participant's perspective or point of view.

How is this topic, the group meeting, the activity, the interaction part of the actor's working knowledge of the world?

So how do we go about finding the meanings of social action? How does the researcher articulate a participant's experience of this phenomena?

How does the researcher provide an interpretation of the scene or of, in this case, the group meeting?

The researcher listens and observes the subjects in various exchanges with others or with the researcher in a conversation about the topic, or a series of events, or through various stories.

Then the researcher attempts to summarize, across the described conversations or events, the simplest synthesis or generalization across the exchanges and actions.

This synthesis is the researcher's interpretation of what the events mean or signify to the subject.

How do we test whether the researcher's interpretation is valid? To assess interpretive validity, we have no choice but to go back to the description. Interpretation is not a matter of the researcher's opinion or preferences.

It must be supported by the description that has already been provided.

Does the asserted interpretation accurately synthesize the descriptions as they have been presented?

Does the description provide sufficient evidence to support the interpretation?

We can assess the validity of a description by determining agreement among participants and observers as to the bare facts. Who is there and what happened? However, we cannot fully assess the validity of an interpretation by seeking agreement of the participants.

A research subject may not be fully aware of the range of data grounding the interpretation.

This is because the researcher is developing an interpretation, a claim about the meaning of some actions, on the basis of many examples, not all of which are within the experience of individual or particular actors.

The interpretation of social action and phenomena, the meaning for the actors, is a synthesis across many described examples. If, however, the descriptive materials are rich in detail that is thick,

it may be possible to produce multiple interpretations by focusing on different aspects of the data.

Nonetheless, those multiple foci will have to be reconcilable, made consistent with each other, because it is important that an interpretation not ignore parts of the data.

(3) Theoretical Validity

While interpretive validity refers to the meaning for the actors--both conscious as well as unconscious or tacit intentions, beliefs, concepts--of the people being studied, theoretical validity, a third criteria for qualitative work, refers to what this means to the researchers, not to the actors being studied.

Interpretive validity is established in part by the subject's recognizing themselves in the researcher's descriptions and by the congruence between the description and the researcher's interpretive synthesis.

Theoretical validity is established in terms of the perspectives of other social scientists. Theoretical validity moves the research to a different domain outside the natural setting that is being studied.

A theoretical framework asks, what is this an example of? What is the larger category of phenomena or processes that this data is useful for understanding? How does the researcher's account of this setting or group fit within what is already known? How might what we already know in other scholars' work be amended in light of what is learned in this project?

What are the analytical, somewhat abstract and general categories and concepts that these observations and interpretations illustrate?

Theoretical validity offers the researcher's explanation of the social setting being studied.

More specifically, theoretical validity refers to the concepts the researcher brings to or develops during the study.

Theoretical concepts are what sometimes make social science seem obscure and filled with jargon.

The audience for theoretical validity is not the group being studied, nor is it necessarily the general public, but other social scientists.

Theoretical validity asks what the description and interpretation of the setting mean to the community of scholars observing that activity, what contributions the research makes to the conceptual and theoretical resources of the scholarly community.

The test of theoretical validity occurs when the researcher locates her account and interpretation within a more general set of concepts and models of human action.

Validation occurs through the critique of the final research output, presented at professional seminars, at conferences, through peer review publication, where anonymous reviewers might raise questions and challenge an author's claim.

Most of the time, authors have the opportunity to refine their arguments to meet these challenges or to persuade the scholarly community that a new or different way of viewing a phenomena is necessary and that this new work fills a lacuna in the corpus of knowledge.

This open and participatory process for achieving theoretical validity helps to establish the state of art and social science embodied in the literature from which the students and professional researchers acquire their methodological and theoretical tools, as well as the information recognized as consensually agreed and accurate understandings about the social world.

(4) Generalization

The fourth criteria of validity is generalizability, or what is sometimes called external validity, which refers to the degree to which a particular finding applies to situations or phenomena beyond those of the original study and from which the explanation has been generated.

Is this setting typical of the theoretical or abstracted category of phenomena?

The criteria of generalizability is a matter of identifying the typicality of the setting or persons about and from whom the data were collected.

Can we make inferences from this case or population to other cases or larger populations?

How is the population characterized?

What are the commonalities across examples to which we can apply what was learned in this research?

Of all the possible settings, groups, organizations, or phenomena to observe, what justifies this selection, and how is this sample the same or different from others?

How did the researcher select the particular examples to study?

More to the point - under what conditions are the observations from one setting applicable to another?

Or what limits the use of this data in another?

Does an explanation of human behaviour, say, in the 19th century still apply to behaviour observed in the 20th?

Does a theory developed out of a study of an office environment also usefully explain behaviour observed on a farm?

Does research conducted in Sri Lanka apply to Bangladesh?

I want to pause here to emphasize an exception to this criteria of generalizability.

It is important to realize that in-depth qualitative research, especially ethnography of a particular group or organization, does not need to be generalizable.

It may be a unique setting, an outlier from the normal.

The researcher has decided to look at this situation in order to learn about how this example differs from others.

It is essential, however, that the researcher must be explicit and clear about whether the particular setting is meant to be typical or not.

Much can be learned about uncommon situations, as well as typical situations.

The researcher should not, however, claim generality for findings that are particular to the specific case.

Generalizability is, however, also assessed with regard to the analysis of the data that was collected.

This is sometimes referred to as internal validity.

This dimension or criteria of validity draws attention to the researcher's process of choosing which observations to use to formulate and distribute as descriptions, interpretations, and then theoretical claims of this particular setting, whether or not the setting is common or unique.

Internal validity for qualitative research refers in effect to the methods of sampling within the data collected, rather than selecting a site or cases to study.

Unlike quantitative and statistical data analysis, where all the data are represented in the summary statistics and subsequent analyses, all the qualitative data collected cannot be offered as is to the audience.

The descriptive data need to be culled to provide a sufficiently thick but synthesized account of what was observed, something to serve in place of summary statistics, means, and medians.

Are the situations described, the quotations used, the examples offered in the study report representative of that particular setting, that organization, or that group?

And how did the researcher pick which observations and interviews to report?

In other words, a project will be assessed not only with regard to the typicality of the setting or to its unusualness, but also with regard to the researcher's process for synthesizing and reporting the data.

(5) Evaluative validity

Evaluative validity is about the process of forming judgments, often moral, or normative assessments about the social behaviours or phenomena under study.

Although researchers may have preferences that drive them to study one thing, rather than another, or to prefer one set of social outcomes, rather than others.

In the conduct of social science research, scholars generally avoid moral judgment and concentrate primarily on explanation, and the formulation of useful theories. There are, however, schools of research that purposely take normative stances with regard to the condition of the worlds they study.

For example, there is a strong and large segment of social science research that systematically addresses questions of inequality, power, and justice, with lively and extensive traditions of critical theory.

Thus, students are likely to encounter social science research that does make moral or political judgments, implicitly or explicitly, judgments about the activities of their research subjects.

The presence of normative judgment in a researcher's account does not immediately invalidate the argument, but it does open the door to critiques and discussions of its evaluative validity.

Has the researcher laid out his preferences, his standards for moral evaluation of the actor's experiences and interpretations?

Unless explicitly framed as a political critique, the focus of most social science research and writing is on explanation, not moral judgment.

That doesn't mean your judgments or moral assessments about your topic are irrelevant, but they are not considered scientific, because they cannot be falsified, and they ought to be kept outside the research process.

A social scientist with different preferences should produce the same results from the data collection and analysis, if the researcher has successfully put aside his preferences.

Issues of generalized ability, and evaluative validity, do not vary significantly between qualitative and quantitative research methods.