



5

ASSESSING FOCUS GROUP RESEARCH

HOW CAN THE quality of focus group research be assessed? This question is part of the broader issue of assessing qualitative research in general. Many scholars agree on the need for quality assessment for qualitative research, particularly with its growing use across multiple disciplines; however, the strategy for assessing quality remains an area of much debate and divergence. The traditional criteria for quality assessment of scientific research (objectivity, validity, and reliability) are often seen as inappropriate for assessing qualitative research because they are based on assumptions about research that stem from the positivist paradigm of quantitative measurement and experimental research. Some scholars have therefore proposed alternative criteria for the assessment of qualitative research that are more reflective of the interpretive paradigm that underlies qualitative approaches. However, others maintain that the concepts of validity and reliability are important quality measures, but they require a different approach when applied to qualitative research. In addition, multiple criteria for assessing qualitative research have been developed, but there remains no broad consensus on the suitability of generic criteria to assess the diverse range of approaches used in qualitative research. Discussions on appropriate and effective strategies for assessing the quality of qualitative research are therefore ongoing in academic literature.

This chapter begins by providing an overview of some of the challenges in assessing qualitative research in general, and the drawbacks of using generic criteria to assess qualitative research. The difficulties in applying the traditional criteria of validity and reliability for assessing qualitative research are explained, but this chapter focuses on the importance of these concepts for quality assessment and how each can be effectively used to assess qualitative research. The chapter concludes by outlining a framework for assessing focus group research by following stages of the research process to assess research design, data collection and interpretation, and presentation of research findings. These method-specific guidelines can be used to assess research articles using focus group discussions or to maintain rigor in the design of a focus group study.

Assessing Quality in Qualitative Research

Qualitative research is increasingly being used and published in a diverse range of disciplines. As a result, a greater variety of academic researchers, editors, reviewers, and funders are becoming exposed to qualitative research, yet many may have limited experience in qualitative research and its underlying principles. This has prompted renewed interest for guidance in assessing the quality of qualitative research, in particular the call for more formal criteria for quality assessment.

The call for assessment criteria for qualitative research has come from multiple sources. Academic researchers across a broad range of disciplines are now using qualitative research. The increase in mixed methods research and a movement toward interdisciplinary research has led to researchers in diverse disciplines becoming exposed to qualitative research. Researchers from varied disciplines request guidance on how to use and assess qualitative research. In response, a host of articles have been published in a variety of academic journals that highlight the value of qualitative research for a certain discipline and provide guidance on quality assessment. Qualitative research is also increasingly being published in biomedical journals; however, journal editors and reviewers often lack training in social science research, and fewer have specific expertise in qualitative research methods. This has promoted the need for criteria for quality assessment to guide

the review process and inform publication decisions on qualitative manuscripts (Green & Thorogood, 2009; Flick, 2007). Some academic journals now provide guidelines for authors wishing to submit qualitative research, which often become the internal evaluation criteria for publication decisions. Research funding bodies also need to judge the quality and feasibility of research proposals that use qualitative methods. Leading funding bodies in the United Kingdom (e.g., Economic and Social Research Council) and the United States (e.g., National Institutes of Health, National Science Foundation) now provide documents to guide the assessment of qualitative research to foster a more transparent review process. Furthermore, much qualitative research is still conducted in the health sciences, which has experienced a major shift toward evidence-based health care, whereby health policy and practice is based on research evidence. Green and Thorogood (2009) state that if findings from qualitative research are to be included in research evidence that will inform clinical practice and health-care decision-making, there needs to be some assessment of the quality of the evidence presented in qualitative studies. Policy and practice decisions based on low-quality research may lead to ineffective changes in health service delivery and wasted healthcare resources (Dixon-Woods, Shaw, Agarwal, & Smith, 2004). Overall, as qualitative research is increasingly being conducted and evaluated by disciplines less familiar with the principles and procedures of the approach, the need for transparent quality assessment strategies is becoming more pressing.

The traditional criteria for assessing scientific research (objectivity, validity, and reliability) are widely used across multiple disciplines. These concepts are derived from the natural sciences and are therefore most relevant to assessing quantitative and experimental research studies. The direct application of these concepts to qualitative research is problematic because of its interpretive approach, the iterative research process used, and the subjective nature of qualitative methods (discussed later). Therefore, many scholars have argued the need for different criteria to assess qualitative research from that used for quantitative studies; however, the question of how to assess qualitative research continues to be a challenge (Flick, 2007; Silverman, 2011a). In recent decades many criteria, guidelines, and checklists specifically tailored for assessing qualitative research have been proposed in academic literature or

have emerged from scientific journals or research funding bodies in response to the needs described previously. Dixon-Woods et al. (2004) identified more than 100 different proposals on assessing quality in qualitative research, some of which, they state, adopt incompatible positions on certain issues.

Despite these attempts to develop criteria for quality assessment, there remains little consensus on appropriate strategies for assessing qualitative research. In part, this challenge relates to the nature of qualitative research itself, which is not a unified field but comprises a diverse range of methods, methodological approaches, and theoretical perspectives, making a criteria-based approach to assessment particularly difficult. For example, conducting in-depth interviews within the grounded theory approach may require different assessment criteria from in-depth interviews conducted within discourse analysis or within community-based participatory action research. Furthermore, assessing a study that used grounded theory is itself problematic. Not only is grounded theory difficult to implement in its original form, but the approach itself has evolved with each of its developers (Glaser and Strauss) taking the approach in different directions. Strauss developed the more structured procedural aspects of grounded theory, whereas Glaser retained the components of emergent discovery of the approach. Providing criteria to assess a grounded theory study is therefore far from straightforward. Several studies may have used the same method (e.g., interviews) or methodological approach (e.g., grounded theory), but require a different assessment strategy to acknowledge the diversity of methodological approaches used. Furthermore, developing a unified set of criteria for assessing qualitative research may have an undesirable outcome of favoring certain approaches over others, potentially leading researchers to write to the criteria to “tick the box” and maximize publication rather than reflect how validity was actually achieved in a study (Barbour, 2001).

A further challenge in developing quality criteria lies in determining how to assess the more interpretive elements of qualitative research. Some of the central tasks of qualitative research involve interpretation, such as code development and coding data, which can be difficult to describe and more challenging to assess. It remains difficult to develop indicators for readers to recognize the interpretive components of qualitative research,

which can also be effectively operationalized so that different readers agree on whether these criteria have been met. The dilemma is that some of the most important qualities of qualitative research can be the hardest to assess (Dixon-Woods et al., 2004). Therefore, a concern of using criteria for quality assessment is the risk of giving less prominence to the interpretive elements because of their measurement difficulties, while giving undue prominence to the more tangible procedural tasks in qualitative research. One outcome may be that studies following appropriate procedures but with poor interpretation are considered better quality than those with less procedural detail but presenting rich and compelling interpretive detail (Dixon-Woods et al., 2004). Formalizing quality assessment through criteria may thus suppress the important interpretive components that are central to qualitative research.

Using generic criteria for assessing qualitative research therefore remains challenging, as does directly applying the traditional criteria of validity and reliability. However, the concepts of validity and reliability remain important for assessing qualitative research, but require a different application to embrace the interpretive paradigm and qualitative research. The following sections describe the challenges in using validity and reliability in their original form for assessing qualitative research, and then describe how each concept can be applied to qualitative research to effectively assess quality and scientific rigor.

Applying Validity and Reliability to Qualitative Research

Validity and reliability are well established measures of scientific rigor. However, they originate from quantitative research and their direct application to qualitative, interpretive research can be problematic, as indicated previously. In response, a range of alternative terms for assessing qualitative research have been proposed, such as “trustworthiness,” “credibility,” and “legitimacy,” instead of validity and the terms “dependability,” “consistency,” “stability,” and “representativeness” as alternative terms for reliability (Guest, MacQueen, & Namey, 2012). Although there are good arguments for using alternative terms, Morse, Barrett, Mayan, Olson, and Spiers (2002, p. 8) state that “the terms reliability and validity remain pertinent

in qualitative inquiry and should be maintained,” while also understanding their limitations for assessing qualitative research. Furthermore, Morse et al. (2002) argue that by creating alternative terms for these measures, qualitative research may be marginalized from mainstream science and the associated legitimacy it has. Despite their imperfect fit to qualitative research, the concepts of validity and reliability remain equally important for assessing qualitative research. Therefore, the remainder of this chapter uses the terms validity and reliability and describes how these concepts can be applied to qualitative research. Discussions of validity, reliability, and quality in qualitative research are extensive in published literature; this chapter provides only a summary of these concepts, their limitations, and how they can be applied to qualitative research. Although the strategies described next should enhance the credibility of a study, they are not sufficient to guarantee scientific rigor and quality, and cannot rectify a poorly conceived study, ineffective research instruments, or lack of critical analyses of data.

Validity

Scientific validity refers to “truth” or “accuracy” and may be described as “the extent to which an account accurately represents the social phenomenon to which it refers” (Hammersley, 1990, p. 57). The concept of validity originates from the positivist paradigm and applies most directly to quantitative research and the extent to which a study captures the true phenomenon. There are two components of validity: internal and external validity (Ritchie & Lewis, 2003). Internal validity refers to the extent to which a study measures what it intended to measure. External validity refers to the extent to which study findings are generalizable to a broader population outside the study itself. These constructs of validity are clearly based on the positivist paradigm of measurement and objectivity, and are less appropriate to qualitative research in their original form for the following reasons. The concept of validity assumes that there exists a single “truth” that can be captured through a research instrument, such as a survey. However, the underlying assumption of the interpretive paradigm is that there is not one truth but multiple perspectives on reality when examining social phenomenon. Therefore, validating the accuracy of an account is difficult to apply to qualitative research where multiple accounts of the same

phenomenon are possible and it is the range of different perspectives that are valued. Internal validity is also based on the assumption that certain variables in a study, notably contextual factors, can be controlled in statistical tests to ensure that the analysis measures the specific variables of interest without any confounding factors. Such analytic approaches require standardized data collection and analytic techniques not available, or appropriate, for qualitative research. Furthermore, external validity involves the ability to generalize study findings, which typically requires drawing a random sample so that the study findings can be extrapolated to a broader population; however, qualitative research uses purposive (non-random) sampling aimed at seeking depth and richness of information not representativeness.

Although the concept of validity has its origins in measuring validity in quantitative research, “it is widely recognized that [validity] is an equally significant issue for qualitative research. But the questions posed are different ones and relate more to the validity of representation, understanding and interpretation” (Ritchie & Lewis, 2003, p. 273). Overall, validity has a different focus when applied to qualitative research where it is used to assess “the credibility and accuracy of process and outcomes associated with a research study” (Guest et al., 2012, p. 84). Internal validity involves assessing the credibility of a study, to examine whether the data and its interpretation are trustworthy and effectively portray the phenomenon examined. Providing transparency in the research process by describing all procedural tasks and decisions can demonstrate scientific rigor, which contributes to the trustworthiness of the data. Further strategies can be used to demonstrate the validity of interpretation of qualitative data, to show that they are valid representations of a phenomenon. In addition, the transferability of qualitative research findings is often used to describe external validity, to assess the context in which the results of qualitative research can be transferred to other settings or populations. Some approaches to demonstrate validity of qualitative data and the validity of its interpretation are summarized next.

Validity of Data

The validity of qualitative data refers to the extent to which the data effectively portray the phenomenon under investigation. Are the

data trustworthy? Do they accurately represent the phenomenon, its variation, and nuance? Were data generated from an inductive process? The validity of data is clearly dependent on the rigor of the research process from which it was produced and the effective application of inductive data collection; therefore, it is necessary to demonstrate research procedures used to enable validity to be assessed. This requires researchers to clearly document the research process so that others can judge the credibility of the research and the trustworthiness of the data. Demonstrating credibility and showing transparency are key strategies that contribute to assessing the validity of qualitative data, as described next.

Credibility (Lincoln & Guba, 1985) refers to the trustworthiness of the study to generate valid data that accurately represent the phenomenon studied. Credibility refers to the “confidence in the truth of the findings, including an accurate understanding of the context” (Ulin, Robinson, & Tolley, 2005, p. 25). Assessing whether data collected effectively reflect the views of study participants is central to assessing qualitative research. The credibility of the study is directly related to the research process, the methodological procedures and decisions, and the steps taken to ensure scientific rigor. Therefore, validity can be enhanced during all stages of the research process from developing an appropriate research design, selecting research methods, following inductive data collection, and using effective strategies to analyze and interpret data. At each of these stages of the research process scientific rigor can be enhanced, such as adequate training of moderators on rapport development and probing; pilot-testing discussion questions; following inductive data collection (described later); using accepted procedures for data analysis to ensure interpretations are evidence based; and implementing ethical procedures. Scientific rigor through all stages of the research process is critical for data validity. Therefore, validity is not only assessed at the completion of a study but addressed during each task in the research process. Providing transparency in reporting the research process (see discussion below) can demonstrate that the research is robust and that procedural validity was enhanced throughout the study. For focus group research, the elements of effective study design and implementation have been described in previous chapters of this book. Figure 5.1 details specific questions that can be asked of a focus group study to assess its credibility.

Inductive data collection, using open questions and probing participants, inherently facilitates valid responses from participants, particularly compared with structured quantitative data collection. Guest et al. (2012) state that the open-ended nature of interview questions and inductive probing used in qualitative research allow researchers to gain more precise responses from participants than closed category questions on a survey instrument. For example, a participant's response to a survey question may not be offered as one of the closed category options in quantitative research, particularly when there is no "other" category. This leaves the participants or the interviewer to assign a response category, which may not be entirely valid. This problem is avoided in qualitative research where participants can provide open and elaborate responses in their own words without the constraint of researchers' categories, which may more accurately capture their views and provide greater depth and nuance, thereby improving data validity and overall quality. In addition, the interviewer has the flexibility to probe a participant for clarity or rephrase a question if it seems unclear to a participant. Therefore, the interactive and inductive processes of qualitative interviewing inherently contribute to data validity.

Transparency involves clearly reporting the research process to provide an "audit trail" of procedures and decisions from which others can assess the validity of the study and the data generated. An audit trail lays bare the rationale for the study design, the process of data generation, and the analytic procedures used. It not only provides procedural information on what was done and who was involved, but also the reasoning for choices made during the research process. In addition, an audit trail can "show the conceptual process by which meaning or interpretation has been attributed or theory developed" (Ritchie & Lewis, 2003, p. 276). Therefore, an effective audit trail highlights the analytic steps that led to assertions made in the findings, so that there are no seemingly unsupported leaps of logic in the final results presented. Providing transparency in the research process is particularly important in qualitative research given the diversity of approaches used and the iterative process of data collection. This diversity means that studies using similar research methods may have applied a different field approach, which underscores the need for each study to clearly document their process, procedures,

and rationale. Although such documentation does not guarantee validity, it does provide important information for readers to make an informed assessment of the scientific rigor of the study, and thereby the credibility of the study findings and interpretations. The importance of transparency is not only for external quality assessment, but can also encourage researchers to be more systematic and deliberate in their approach and provide clear rationale for the methodological decisions, thereby also increasing research quality throughout the study (Guest et al., 2012).

Validity of Interpretation

Much qualitative research is based on understanding meaning and interpretation of data. Therefore, demonstrating the validity of these interpretations is critical to the trustworthiness of qualitative research findings. How valid are researchers' understandings and interpretation of the data collected? How can researchers' manage subjective interpretation of data? How can validity of concepts and explanations be demonstrated? These are critical questions in assessing the validity of interpretation in qualitative research. Several strategies are commonly used to demonstrate the validity of researcher's interpretations of qualitative data, as described next.

Respondent validation (also called "member checking") involves presenting a summary of the study findings to a selection of study participants, other members of the study population, or key informants familiar with the culture or context of the research. These informants are asked to respond to the research findings, often confirming or clarifying results presented, and verifying their accuracy within the study population. This strategy provides some external validation that study results and their interpretation are valid and recognizable by members of the study community themselves. It provides an important safeguard against interpretation bias. When this strategy is used it is typically included in the methods section of a research article, highlighting the nature of the respondents providing comment and whether any discrepancies in interpretation were identified.

Although respondent validation has some appeal, it presents numerous challenges. Data are collected from individual participants, yet respondent validation involves verifying the collective study results that comprise a synthesis of multiple experiences

and perspectives. Some have questioned the effectiveness of this approach. Can study participants effectively verify analytic outcomes of a study, which result from cross-case comparison and detailed immersion in data? Will study participants understand the collective results of academic research? Some scholars (Morse et al., 2002; Barbour, 2001; Mays & Pope, 2000) caution that respondent validation may be problematic because an individual's response may not be visible in aggregated research results. This becomes particularly challenging when study results are more conceptual or present explanatory frameworks, such as in grounded theory, because concepts and processes become more abstract than the individual experiences from which they are generated and may therefore be difficult to recognize by the study community. In contrast, Guest et al. (2012) believe that even though an individual's response is not explicitly visible, participants would be able to recognize some of the issues that their contribution helped to create. A related challenge in using this method of validation is that it assumes that a single reality is being verified; however, multiple experiences were captured in the data. This may cause participants to disagree with the experiences or viewpoints presented in the results that they are unfamiliar with or that differ from their own perspectives. This does not mean that these results are incorrect, but reflects that multiple perspectives exist. Thus, participants may validate their own perspective but not that of others as presented in study results. Furthermore, respondent validation may not be logistically feasible when it is not possible to return to the study population, such as for in international research or when resources are limited.

Peer review involves assessing validity by asking researchers outside the research team to examine study data and the interpretations derived. This provides an assessment of "external validity." Peers are instructed to assess the logic and consistency of the analysis to identify potential interpretation bias (Guest et al., 2012). Peer review provides assessment of the researchers' interpretations and whether these are well-grounded in the data itself, thereby keeping researchers' subjectivity in check. A similar strategy can be conducted within the study team, whereby several team members analyze the same section of data independently to assess the consistency with which they are able to generate similar interpretations of the data.

Negative and deviant case analyses are strategies for increasing analytic rigor to minimize researcher's interpretation bias. Qualitative research is commonly criticized for using data selectively (or "cherry-picking") to support an argument proposed by the researcher. Negative and deviant case analyses are analytic tasks that challenge researchers to be critically self-reflective in interpreting data by challenging their interpretations. Negative case analysis involves actively seeking data that may contradict themes identified or an explanation proposed, and highlighting or explaining these negative cases. Contradictory data can be challenging to manage but explicitly seeking and incorporating these data into the study results indicates that data interpretations are indeed reflective of complex qualitative analysis rather than being used selectively to support a particular perspective. Similarly, deviant case analysis involves identifying outliers that do not fit an emerging interpretation of the data. These outliers are then examined explicitly, whereby researchers may adjust their interpretations to incorporate outliers or understand why these cases are different. These strategies may be reported in the data analysis section of a research article to demonstrate how the interpretations presented "fit" the study data.

Delimiting interpretations to make explicit the context in which they are valid is a simple strategy to increase the validity of interpretations presented. Not all study findings are relevant to the entire study population. Some explanations apply to a defined subset of participants (e.g., young males only), whereas others are valid only under certain conditions or circumstances (e.g., only participants who use public transport explained difficulties in accessing facilities). Delineating the boundaries, or scope, of an explanation provides specificity on when an explanation is valid and the conditions under which this interpretation holds true, thereby increasing the validity of interpretation.

Analytic induction is a process of analyzing qualitative data involving iterative interpretation to ensure that explanations "fit" the data. Analytic induction (Silverman, 2011a; Flick, 2009; Fielding, 1988) involves developing a provisional explanation of phenomenon based on initial analyses, then examining data case by case to assess whether the explanation fits each case. As each case is examined the explanation is adjusted to

incorporate nuances of specific cases, so that the explanation evolves from the analytic process. When a case does not fit, the explanation is refined and this process continues until all data can be accounted for with the final explanation. Analytic induction involves constant comparison of cases; examination of negative cases; and incorporation of outliers (described above). The use of analytic induction strengthens study findings by demonstrating that they originate from data itself and not researchers' subjective interpretation. Although the details of each iteration are not reported in a research article, the use of analytic induction and any major adjustments to an explanation may be noted in a description of data analysis and theory building.

Triangulation refers to "combining multiple theories, methods, observers and empirical materials, to produce a more accurate, comprehensive and objective representation of the object of study" (Silverman, 2011a, p. 369). It is a common strategy for validating research findings that is based on the premise that when findings from multiple independent sources converge it provides confidence that the findings are trustworthy and valid.

Triangulation may contribute to validity in two ways: by confirmation of the study findings or by providing completeness of the findings. Denzin (1989) suggests four ways to use triangulation as confirmation of study findings in qualitative research by triangulating: (1) between research approaches (e.g., quantitative and qualitative); (2) between methods (e.g., interviews and group discussions); (3) between researchers (e.g., using multiple interviewers or analysts); and (4) theory triangulation, whereby data are viewed through different theoretical lenses. Using multiple methods and multiple approaches are perhaps the most common applications of triangulation in qualitative research, in addition to comparing coding from independent analysts to confirm similar understanding and interpretation of data between researchers. A further strategy involves comparing study findings with themes, concepts, and interpretations provided in extant literature among similar study populations.

Triangulation may also be used to improve the completeness of qualitative study findings. Triangulation may be used as a means to increase understanding of phenomenon by examining it from different perspectives, thereby exploiting the variation sought in

qualitative research. This use of triangulation does not validate or confirm findings but “is best understood as a strategy that adds rigor, breadth, complexity, richness and depth to any enquiry” (Denzin & Lincoln, 2000, p. 5). This approach adds rigor to the research by exploring phenomenon from multiple perspectives, examining contradictions and inconsistencies that exist in qualitative data to provide a fuller understanding of the issues.

Using triangulation to confirm study findings has ready appeal, but it can be difficult to conduct effectively. Data from different methods come in very different forms (e.g., observations vs. group discussions vs. survey data) that may not be directly comparable. Furthermore, the generation of similar findings from different methods of data collection provides some confirmation of those findings, but the absence of corroboration does not suggest lack of validity in qualitative research, because different methods produce different views of the phenomenon under study (Barbour, 2001). There is also some debate among qualitative researchers on the value of triangulation for confirming study findings, because it assumes there is a single objective truth that can be validated. However, “in cultural research, which focuses on social reality, the object of knowledge *is* different from different perspectives. And the different points of view cannot be merged, into a single, ‘true’ and ‘certain’ representation of the object” (Moisander & Valtonen, 2006, p. 45). Hammersley (1992) argues that one cannot know for certain that accounts given in social research are true because there is no independent and reliable access to “reality,” therefore all accounts can be true even when divergent, because they represent different perspectives on reality. It is often the goal of qualitative research to seek out variant views and diverse experiences, therefore assessing validity by convergence (through triangulation) seems contradictory to the purpose of qualitative research. Nevertheless, some applications of triangulation can validate study findings in qualitative research, so triangulation should not be dismissed but applied with awareness of its limitations for validating social research.

External Validity

External validity typically refers to the ability to generalize study findings to a broader population. It is an important criterion for

validity in quantitative research. Generalizability is based on the expectation of a sufficiently large sample whereby participants are randomly selected and standardized data collection is used. As such, generalizability is difficult to apply to qualitative research, which focuses on a small number of participants selected non-randomly and data are collected using responsive (non-standardized) interviewing, because the goal is to seek information richness not representation. Some scholars state that generalizability is not applicable to qualitative research because it is based purely on description and focuses on select cases. Others state that generalization is a relevant and achievable task in qualitative research, albeit conducted in a different way than in quantitative studies. For example, Padgett (2012, p. 206) states that “[qualitative] findings can have transferability and resonance without being ‘generalizable’ in a statistical sense based on how the sample was selected,” and Mason (1996, p. 6) states that “qualitative research should produce explanations which are generalizable in some way, or have wider resonance.” Given that the concept of generalizability may be difficult to apply to qualitative research, Lincoln and Guba (1985) suggest the term “transferability of study findings” as more appropriate for qualitative research.

In qualitative research, generalization is approached differently than in quantitative research. Ritchie and Lewis (2003) describe three forms of generalizability for qualitative research: (1) representational generalizability, (2) inferential generalizability, and (3) theoretical generalizability. Representational generalizability refers to the extent to which study findings can be inferred to the parent population from which they were sampled; however, the basis for such representation is different in qualitative research. In qualitative research “it is not the prevalence of particular views or experiences... about which wider inferences can be drawn. Rather, it is the content or ‘map’ of the range of views, experiences, outcomes or other phenomena under study and the factors and circumstances that shape and influence them, that can be inferred to the research population... It is at the level of categories, concepts and explanation that generalization can take place” (Ritchie & Lewis, 2003, p. 269). Achieving representational generalizability draws on issues of validity of the research process, including the degree to which the sample captures diversity within the parent population and the accuracy with which phenomenon have been

identified and interpreted (as described in previous sections). For example, using recruitment strategies for maximum variation (e.g., purposive and theoretical sampling) captures diversity in the context and conditions of the issues examined, thereby capturing the heterogeneity within the parent population that enables inferences to become applicable. This aspect of generalizability reflects the principle of statistical inference but without using probability criteria. It refers more to achieving inclusivity and diversity in the dimensions and properties of the issues examined (Silverman, 2011a). The validity and scientific rigor of a study therefore has a critical influence on achieving representational generalizability.

Inferential generalizability refers to the relevance of the study findings to other contexts and populations beyond the study setting itself. For example, as the extent to which findings from a study on injecting drug users in New York City can be applied to injecting drug users in other large US cities. The core distinction from representational generalizability is that findings are not so much being assessed as “representing” the parent population, although this is implicit, but being applied (or “inferred”) to a new context or population. Inferential generalizability is achieved in qualitative research by generating broader level concepts, processes, explanations, or theoretical frameworks that have relevance outside the specific context from which they were derived. For example, such general concepts as “fear,” “stigma,” “peer pressure,” or “bullying” are transferable to other contexts, but the specific examples from which these concepts were derived remain context- or case-specific. Therefore, “[inferential] generalization in qualitative research is the gradual transfer of findings from case studies and their context to more general and abstract relations, for example a typology” (Flick, 2009, p. 408). This involves reducing the contextual relevance of the study findings by developing broader conceptual results that can be transferred to other settings. Flick (2009, p. 407) states that “this attachment to contexts often allows qualitative research a specific expressiveness. However, when attempts are made at generalizing the findings, this context link has to be given up in order to find out whether the findings are valid independently of and outside specific contexts.” Some approaches to qualitative research, such as grounded theory, are more suited to generating conceptual results from individual narrative experiences. It is important to document the analytic process by which

broader concepts were derived to demonstrate the internal validity of those concepts being transferred. The effectiveness of inferential generalizability undoubtedly also depends on external factors, such as the degree of congruence between the study context and the context to which findings are being inferred. Providing as much contextual detail as possible on the study itself allows others to determine the appropriateness of inferring findings to other contexts.

Finally, theoretical generalizability refers to the use of study results to develop empirical theory, by the development of new theory or the contribution of new concepts to existing theory. In this sense the theory developed is context-free and thus generalizable in the universal sense as a contribution to scientific inquiry. However generalization is used in qualitative research, the type of generalization and the basis for its relevance needs to be made clear so that appropriate claims of generalizability can be made.

Reliability

Reliability refers to the replicability of a study, whereby if the study was repeated using the same methods and approach it could produce the same results. Reliability responds to the question of “whether or not some future researchers could repeat the research project and come up with the same results, interpretations and claims” (Silverman, 2011a, p. 360). It is an objective measure of consistency, which essentially demonstrates that the study findings are independent of any accidental circumstance of their production, and are therefore free of subjectivity or bias (Kirk & Miller, 1986). Reliability originates from the natural sciences and is most appropriate to more standardized quantitative research and experimental design, whereby taking repeated measures that show consistent results demonstrates the reliability of those readings.

The ideal of objective replicability is often seen as unobtainable in qualitative research because of the subjective nature of social research and the iterative research process used. Qualitative research is often conducted to understand complex social phenomenon, to explore contextual influences on social behavior, and to seek diversity in participants’ experiences and characteristics. To do this it requires an iterative process of discovery that is responsive and dynamic so that researchers can follow leads as the

research process unfolds. This approach is unlikely to be repeated exactly; therefore, the goal of objective replication may be naive and unachievable in qualitative research (Lincoln & Guba, 1985). Even when using semi-structured research instruments, which often ask the same open questions in the same order, an interviewer often uses a great deal of responsive probing thereby taking each interview in potentially very different directions depending on the participant's experiences (Guest et al., 2012). Therefore, even though structure may exist in qualitative research instruments, inductive probing means that responses may not be replicable, but this is not to say that these responses are not reliable.

Replicability may also be challenging because of the interpretive nature of qualitative research. Interpretation is a central component of qualitative research, but it introduces subjective influences to the research process, which challenge the goal of replicability. These concerns mean that the goal of objective replicability, which is implicit in reliability, is not appropriate to qualitative research. However, this does not mean that the concept of reliability should be abandoned altogether for qualitative studies, but rather discussed in terms that have greater resonance with the principles and procedures of qualitative research. As a result of these issues alternative terms have been proposed for assessing reliability in qualitative research, for example "confirmability" (Ritchie & Lewis, 2003); "consistency" (Hammersley, 1990); "trustworthiness" (Glaser & Strauss, 1967); "dependability" (Lincoln & Guba, 1985); and "transparency" (Silverman, 2011a). These terms highlight central characteristics of reliability but do not focus on objective replicability.

Reliability is often seen as less important than validity in qualitative research because replication, which is at the heart of reliability, is not a goal of qualitative research. In applying reliability to qualitative research, it is important to understand the elements of qualitative research that can be consistent and confirmed, and that could reoccur with some certainty. Ritchie and Lewis (2003, p. 271) state that "it is the collective nature of the phenomena that have been generated by the study participants and the meanings that they have attached to them that would be expected to repeat." Therefore, replicability may be sought in the core concepts identified in qualitative data and the consistency in understanding the meanings that participants attach to these concepts; "thus the reliability of the findings

depends on the likely recurrence of the original data and the way they are interpreted” (Ritchie & Lewis, 2003, p. 271). Reliability rests in part on rigorous application of qualitative research procedures to identify inductive constructs and their interpretation, and on transparency and documentation of these procedures. In this way reliability also enhances the validity of a study. Reliability becomes particularly relevant to qualitative research when comparison is sought, for example between groups or locations, requiring some consistency in study procedures. Providing structure in research procedures facilitates reliability and comparison. Guest et al. (2012, p. 88) highlight that “instruments, questions and processes with more structure enable a more meaningful comparative analysis. With no structure one cannot make claims that any differences observed are due to actual differences between groups, since all or most of the variability could just as easily be due to differences in the way questions were asked.” Structure may be achieved by using systematic procedures in data collection and analysis; however, these should not compromise the inductive discovery that is characteristic of qualitative inquiry. Strategies for building structure in qualitative research to enhance comparability and reliability have been well documented and are summarized next.

Transparency of research procedures is as important for reliability as it is for validity (as described previously). Detailed documentation of the scientific procedures used in data collection and analysis are critical for the replicability of a study, and demonstrate that data support the claims made in the study findings. Furthermore, Silverman (2011a) indicates the importance of “theoretical transparency” for assessing reliability, whereby researchers make explicit the theoretical stance from which interpretation was conducted, and show how this led to the particular interpretations of data that were presented and excluded other interpretations, thereby guiding the reader on the theoretical framework within which the study could be reliably replicated. It is therefore incumbent on researchers to provide sufficient detail and transparency on the research process, procedures, and the theoretical stance of the study to assess its reliability (Kirk & Miller, 1986).

Reflexivity involves researchers indicating subjective characteristics or circumstances that may have influenced data collection or interpretation. Reflexivity is an important element of reliability because it can indicate whether there were certain characteristics

of the researcher (e.g., ethnicity, experience, perspectives, or language ability) that may have influenced the nature of data collected. Similarly, specific circumstances may be unique to a study, for example a study on people's perceptions of disaster relief immediately after experiencing an earthquake, drought, or other natural disaster. These characteristics and circumstances may be unique to a particular study and not replicable in future studies. Therefore, including reflexivity in a research report is important for its influence on the replicability of a study. Reflexivity therefore extends the transparency of the research process and contributes to assessing reliability.

Using systematic procedures provides consistency in the research process, which supports reliability. The systematic procedures that may be used for focus group discussions are described more fully in Chapter 2. For example, using a semistructured discussion guide ensures participants are asked the same set of open questions, while still allowing the moderator to explore issues raised in the discussion. Field-testing the discussion guide provides a check on whether participants consistently understand the questions in the same way. Training moderators on the intent of each question on the discussion guide facilitates relevant and consistent probing, because "without a sense of purpose, inductive probing lacks both direction and relevance" (Guest et al., 2012, p. 86). Monitoring data as they are collected improves consistency and overall data quality, as it provides an opportunity to review transcripts or debrief with moderators to refocus questioning strategies when needed. Recording interviews, verbatim transcription, and accurate translation of data have become the norm in many qualitative studies. These procedures ensure that data represent participants' own words and intents as concretely as possible and therefore generate more accurate data. Using a transcription protocol adds consistency and systematic rigor, as does checking transcripts for accuracy and completeness (Hennink, 2007, 2008). Translating data adds complexity to data reliability, but training translators on developing verbatim translated transcripts that retain the intent of the speaker contributes to consistency. Use of data analysis software can also improve systematic data analysis. These procedures provide structure and reliability checking at important stages during the process of generating focus group data.

Verbatim quotations connect a reader directly to the words of a study participant, and provide a direct link between the issues raised by participants and their interpretation by researchers. Quotations therefore provide a powerful contribution to reliability in qualitative research, because “quotes lay bare the emergent themes for all to see” (Guest et al., 2012, p. 95). They enable issues to be verified in the direct words of a participant, rather than relying only on researchers’ interpretation of the issues (Seale, 1999), thereby improving transparency and reliability. However, quotations should always be part of a narrative; without this it can have a detrimental effect on reliability because readers are then open to draw their own interpretations based on select quotations and without having had the benefit of analyzing the whole data set (see Chapter 4 on presenting quotations from focus group research).

Inter-coder reliability is perhaps the most commonly described measure of reliability in qualitative research (Guest et al., 2012; Silverman, 2011a; Barbour, 2001). Inter-coder reliability typically involves two analysts who independently code the same transcripts with an identical set of codes and then compare the consistency of their coding. It provides a strong reliability check to assess whether different researchers interpret qualitative data in the same way by identifying the same core concepts and then code data in a consistent way, thus minimizing subjective interpretation. Developing a codebook that lists all codes and a detailed description of their application adds consistency between coders. There are three common methods to assess inter-coder agreement (Guest et al., 2012; Silverman, 2011a).

Subjective assessment involves coders simply discussing the double-coded text to identify discrepancies in interpretation of data and application of codes, and then revising the codebook or coding strategies if needed. No metrics are generated with this method.

Percent agreement calculates simple percentage agreement based on the tally of agreements over the total number of comparisons. Some qualitative data analysis software has this function and generates tables of agreement by codes or coders, and calculates an overall percentage agreement. This method is useful for identifying problematic codes, coders,

or coding strategies used. An overall percentage agreement of 80% or higher is considered good agreement.

Cohen's kappa calculates the level of agreement taking into account chance agreement and is therefore considered more accurate. A kappa score of 0.8 or greater is considered high, but often difficult to achieve. Cohen's kappa is not effective for small samples and an overreliance on a single statistical score can detract from focusing on the causes of coding discrepancies that influence reliability.

With all methods of intercoder reliability an important component is identifying the causes of coding discrepancies, such as problems with codes, different interpretation of data, or variations in coding styles, and then rectifying these issues. Typically codes that have less than 80% simple agreement or a kappa score of below 0.8 are discussed and revised to improve consistency. Although producing a quantitative measure of intercoder agreement has its place there are also limitations. For example, even a slight difference in the size of a coded text segment coded is considered inconsistent agreement reducing the level of agreement generated. Therefore, simple subjective assessment often is sufficient and retains the focus on discussing discrepancies to increase consistency.

Assessing Focus Group Research

The applications of validity and reliability discussed previously are relevant for assessing focus group research. The strategies described in this chapter can be used to determine the overall credibility of a focus group study to produce valid data and to assess the trustworthiness of a researcher's interpretations of data. Following the procedural guidance given throughout this book will assist in making effective methodological decisions to improve the quality and rigor of a focus group study and contribute to valid and reliable study outcomes.

In assessing focus group research it can be helpful to adopt a process approach to examine whether aspects of validity and reliability are addressed at different stages of the research process. Figure 5.1 integrates relevant procedural advice on conducting focus group research (given in this book) with indicators of validity and reliability (discussed in this chapter) and presents questions that can be asked at different stages of the research process

STUDY PURPOSE	
Study Purpose	<p>Is the study purpose clearly stated?</p> <p>Does the article title effectively capture the study purpose?</p> <p>Are focus group discussions the most effective method for the study purpose?</p> <p>Is the significance of the study clearly articulated?</p>
Research Question	<p>Is the research question clear and focused?</p> <p>Is the research question suitable for qualitative research?</p> <p>Are focus group discussions suitable to answer the research question?</p>
RESEARCH DESIGN	
Theoretical Framework	<p>Is the study embedded in a theoretical framework?</p>
Study Design	<p>Is the study design clearly identified?</p> <p>Does the study design operationalize the theoretical framework?</p> <p>Are focus group discussions appropriate for the study design?</p> <p>Is the use of focus group discussions clearly justified?</p> <p>Is the role of focus group discussions clear in a mixed methods study design?</p>
DATA COLLECTION	
Context	<p>Is the study context sufficiently described?</p> <p>Is the selection of study sites described?</p>
Field Team	<p>Is it clear who collected the data and their characteristics?</p> <p>Are the moderator(s) characteristics described?</p> <p>Was there a note-taker present at the focus groups?</p> <p>Is training of the field team described?</p>
Participants	<p>Is the study population defined?</p> <p>Was the study population segmented in any way?</p> <p>Are the types of study participants recruited appropriate for the study purpose?</p> <p>Do they appear to be the most 'information rich' sources on the topic?</p>

Figure 5.1. Assessing focus group research.

Recruitment	<p>Are the inclusion/exclusion criteria made clear?</p> <p>Is participant selection theoretically justified?</p> <p>Is the process of participant recruitment described in detail?</p> <p>Are recruitment strategies relevant to the study population?</p> <p>Was participant recruitment iterative?</p>
Group Size, Composition & Number	<p>Is the size of focus groups appropriate? Are particularly small/large groups justified?</p> <p>Is group composition homogeneous? How was this achieved?</p> <p>Is the level of acquaintance between participants indicated (e.g. strangers or acquaintances)?</p> <p>Is the number of focus groups stated and justified?</p>
Data Collection	<p>When were data collected?</p> <p>Is there evidence of iterative data collection?</p> <p>How was group interaction encouraged?</p> <p>Is there evidence of responsive probing?</p> <p>How was information saturation achieved/determined?</p> <p>Were focus groups held in a suitable location?</p> <p>Is the context of data collection reflected?</p> <p>Is an audit trail of data collection and analysis evident?</p>
Research Instrument	<p>Was a discussion guide used?</p> <p>Does the discussion guide operationalize the study objectives?</p> <p>Are the topics or questions asked described? Are they suitable for a group discussion?</p> <p>Are questions open and designed to promote discussion?</p> <p>Are questions culturally appropriate for the study population?</p> <p>Is the number of questions appropriate?</p> <p>Was the discussion guide piloted?</p> <p>Was it translated and checked for accuracy?</p>

Figure 5.1. (Continued)

Reflexivity	Is there evidence of reflexivity during the research process? Do researchers reflect whether characteristics of the moderator, group location or contextual issues may have influenced data collection?
Ethics	Was ethical approval received? Are ethical issues adequately described? How was the research explained to participants? How was informed consent achieved? How were anonymity and confidentiality of participants protected? How were risks to participants minimized?
DATA	
Data Recording	How were data recorded? Were group discussions transcribed verbatim and/or translated? How were transcriptions/ translations checked for accuracy?
Data	Is there evidence of 'thick' data with depth, breadth, context and nuance? Does the data retain the 'voices' of the participants?
Analysis	Is the analytic approach stated and appropriate? Were data analyzed systematically? Is the process of data analysis clearly described, allowing the reader to see how analysis was conducted and results were derived? Is there adequate description of how codes and concepts were derived from the data? How was code development and coding of data validated? Is it clear whether data analysis was inductive? How were study findings validated? How was interpretation bias managed? Was a theoretical framework used to guide data analysis? Are negative cases discussed and explained?

Figure 5.1. (Continued)

Limitations	Are limitations of the data, method or study described?
STUDY RESULTS	
Clear & Coherent	<p>Are the study results described clearly?</p> <p>Do the study results reflect the application of the analytic approach stated?</p> <p>Is there logical coherence between the methods, results and study conclusions?</p> <p>Do results respond effectively to study purpose and research question?</p> <p>Is it clear how data collection and analysis arrived at the findings presented?</p>
Structure	<p>Is there a clear and logical structure, argument or central message conveyed?</p> <p>Is there a clear distinction between presentation of data and their interpretation?</p> <p>Are results from different methods presented effectively?</p> <p>Is the data source of results indicated (if different types of data are used)?</p>
New Knowledge	<p>Did new or unanticipated issues emerge from data?</p> <p>Is sufficient context of findings presented to determine their transferability?</p> <p>Is the significance of the results made clear?</p>
Variation	<p>Are a range of issues reported?</p> <p>Are diverse views described?</p> <p>Were effective comparisons made and reported?</p> <p>Was group interaction evident in the study results?</p>
Depth & Focus	<p>Are issues described in depth and detail? Are specific examples provided?</p> <p>Are nuances of issues described?</p> <p>Do results focus on responding to the research question?</p>

Figure 5.1. (Continued)

Context	Is the context of each issue described? Do the recommendations consider the broader socio-cultural or political context? Can participants 'voices' be distinguished from their interpretation by researchers? Are study findings placed in the context of the research literature on the topic?
Presentation	Are results presented appropriately within the interpretive paradigm? Are diagrams or visual displays of results effective? Are results presented ethically? Are quotations used effectively to support study findings?
Validity	Are participants' identities protected in reporting of quotations? Have effective strategies been used to validate data and its interpretation? Are study findings effectively grounded in the data? Are assertions made supported by the data? Is there sufficient evidence presented to validate findings?
DISCUSSION & CONCLUSION	
Appropriate	Are study implications clearly articulated? Are the implications adequately supported by study data? Is the transferability of results appropriately discussed? Are conclusions based on the research evidence? Is further research suggested?

Figure 5.1. (Continued)

to assess the quality and rigor of focus group research. This framework operationalizes the concepts of quality assessment into practical questions that can be used to judge the quality of focus group research. The framework may be used when writing or reviewing focus group research. The questions are presented as a guide rather than a mandate for assessing focus group research. Finally, given the word limits of journal articles not all procedural details

and justifications can be included in a research article; therefore, a balance between ideal content and practical realities also needs to be considered when evaluating focus group research.

Key Points

- Applying traditional quality assessment criteria (objectivity, validity, and reliability) to qualitative research can be problematic because of the interpretive approach, iterative research process, and subjectivity of qualitative research.
- Alternative strategies, checklists, criteria, and terminology for assessing qualitative research have been proposed, but there remains no agreement on appropriate assessment of qualitative research.
- The concepts of validity and reliability remain important for qualitative research, but they require a different application to effectively assess qualitative inquiry.
- Assessing validity of data and its interpretation is important in qualitative research. Rather than measuring validity, it is the validity of representation, understanding, and interpretation that is assessed in qualitative research.
- Strategies for assessing data validity include credibility of the research process and transparency in documenting research procedures.
- Strategies for assessing data interpretation include respondent validation, peer review, negative and deviant case analysis, delimiting interpretations, analytic induction, triangulation, and transferability.
- Assessing reliability in qualitative research focuses on identifying whether there is recurrence of core concepts and consistent meaning of these concepts in data.
- Strategies for assessing reliability in qualitative research include using structure and systematic procedures to identify core concepts (e.g., training data collectors, recording data, verbatim transcription, pilot testing instruments, inter-coder reliability, reporting verbatim quotations, and using reflexivity).
- Focus group research may be assessed using a process approach to identify how validity and reliability were addressed at different stages of the research process (see Figure 5.1).