Getting the Focus and the Group: Enhancing Analytical Rigor in Focus Group Research

Pamela S. Kidd
Mark B. Parshall

In the literature on focus groups, far more attention has been devoted to how groups are organized and conducted than to issues of analysis. Although exploitation of group dynamics is touted as a virtue of focus groups, there is very little guidance in the literature with respect to how differences between group and individual discourse impact the analysis and interpretation of focus group data. In this article, the authors describe analytical challenges inherent in the interpretation of focus group data and suggest approaches for enhancing the rigor of analysis and the reliability and validity of focus group findings.

In recent years, focus group methods have become increasingly popular as either an adjunctive or primary data collection approach in the social and health sciences and in evaluation research. Although the author of a recent article in *Newsweek* describes focus groups as a market research fad of the 1980s that is now passé (Kaufman, 1997), such a blithe dismissal ignores both a history dating back some 50 years and an unabated stream of publications pertaining to focus group methods over the past decade or so (see, for example, Agar & MacDonald, 1995; Basch, 1987; Bertrand, Brown, & Ward, 1992; Carey, 1995b; Carey & Smith, 1994; Catterall & Maclaran, 1997; DesRosier & Zellars, 1989; Drayton, Fahad, & Tynan, 1989; Festervand, 1985; Frey & Fontana, 1991; Kingry, Tiedje, & Friedman, 1990; Kitzinger, 1994; Krueger, 1994, 1997a, 1997b, 1997c; Krueger & King, 1997; Lederman, 1990; Madriz, 1998; McDaniel & Bach, 1996; Merton, 1987; Morgan, 1993a, 1993b, 1995, 1996, 1997a, 1997b, 1997c; Ramirez & Shepperd, 1988; Reed & Payton, 1997; Sim, 1998; Stewart & Shamdasani, 1990; Straw & Marks, 1995; Straw & Smith, 1995; Templeton, 1987). To paraphrase Mark Twain, reports of the demise of focus groups are exaggerated.

To some extent, the increased interest in and the use of focus groups are based on pragmatic issues of time and cost efficiency relative to individual interviews. However, these presumed savings may be illusory. Properly conducted focus

AUTHORS' NOTE: The authors wish to acknowledge Barbara Sachs of the University of Kentucky for her thoughtful review and criticism of an earlier draft of this article and to thank several anonymous reviewers for their suggestions. Work described in this article was supported in part by the National Institute for Occupational Safety and Health (NIOSH), 5 RO1CCR413067. Opinions expressed are those of the authors and do not represent policies or recommendations of NIOSH. Requests for reprints should be directed to Pamela S. Kidd, Ph.D., ARNP, CEN, University of Kentucky, KIPRC–Suite 202, 333 Waller Avenue, Lexington, KY 40504-2915; email pskidd1@pop.uky.edu

QUALITATIVE HEALTH RESEARCH, Vol. 10 No. 3, May 2000 293-308

groups are not necessarily inexpensive; unless one is in the business of conducting and analyzing focus groups, the time saved in interviewing may be lost in recruitment, logistics, and trying to make sense out of data that are complex and messy.

Enthusiasts tout the virtues of focus groups, but there are many who criticize focus group methods for a wide range of deficiencies. The few studies in which the claims of either side have been tested experimentally produced no evidence that focus groups are superior to either unmoderated (nominal) groups or individual interviews of equivalent numbers of subjects in terms of number or quality of ideas generated (Fern, 1982). There is some evidence that participants in focus groups find the experience more stimulating than participants in either self-administered open-ended surveys or structured group interviews with less spontaneous interaction (Bristol & Fern, 1996). There is also evidence that focus groups may be of value in studying issues in socially marginalized groups (Kitzinger, 1994; Madriz, 1998).

Having conducted and analyzed data from many focus groups in various occupational injury prevention and training evaluation projects, we have found repeatedly that individuals in groups do not speak or answer questions in the same way that they do in other settings. Agar and MacDonald (1995) characterize focus group discourse as "somewhere between a meeting and a conversation" (p. 80). This creates a far less homogeneous texture of discourse than one finds in individual interviews. In addition, within any focus group, the discourse depends on moderator skill as well as the participants' characteristics and their emotional stake in the topic(s). Less experienced moderators are apt to hew more rigidly to the interview guide (Merton & Kendall, 1946), increasing meeting-like discourse and decreasing opportunities for spontaneous conversation. Although Fern (1982) found little difference in the quality or number of ideas generated in focus groups relative to the degree of moderator involvement, his experimental model involved a topic in which the participants did not have much of an emotional stake.

In contrast to individual interviews, focus group participants relate their experiences and reactions among presumed peers with whom they likely share some common frame of reference. Focus group members comment on each other's point of view, often challenging each other's motives and actions in a pointed fashion. In our opinion, if an individual interview were that confrontational, it would very likely alienate the interviewee. With group interviews, agreements or disagreements are fundamental processes that influence the nature and content of responses as the group progresses. Individual and group interviews may provide evidence of ambivalence, inconsistency, conformance, or thinking out loud among informants. However, by definition, individual informants do not achieve or fail to achieve agreement, and they do not negotiate, confront, antagonize, or directly criticize or commiserate with one another.

During a session, focus group members may modify their opinions, or at least their statements about them, based on the give and take of discussion as the group progresses (Krueger, 1997a). Therefore, before one can make statements with any confidence about what a focus group or series of groups had to say on a given topic, one needs to assess the extent to which responses may have arisen from conformance or censoring (Carey & Smith, 1994), coercion, conflict avoidance, or just plain fickleness. For these reasons, Delphi or nominal group techniques may be more suitable than focus groups when formal assessment of consensus is a primary research aim (Sim, 1998). On the other hand, coercion, going along to get along, and the acquiescence of less committed individuals to positions more passionately

argued by others are all processes of social influence that may be critically relevant to the acceptance or rejection of a new or changed program or product (Schindler, 1992). Therefore, the fact that such processes occur in focus groups may raise legitimate concerns about the trustworthiness of findings, but in contrast to more self-contained forms of unreliability (e.g., inconsistency), it does not necessarily vitiate them completely. In any event, analyzing group discourse as though it was merely a concatenation of individual statements may only yield a laundry list of contradictory statements (Reed & Payton, 1997) or lead to erroneous conclusions such as the misattribution of idiosyncratic viewpoints to a group as a whole (Agar & MacDonald, 1995).

The focus group literature provides abundant and sound advice on process issues such as designing interview guides and structuring and moderating groups. Throughout this article, we assume that such advice has been followed conscientiously. However, merely following those "dos" and "don'ts" does not mean that the data analysis and conclusions are sound. The lengths (or depths) to which one goes in focus group analysis depend in good measure on the purposes of a given study and the requirements of a client, target audience, or sponsoring agency. Trusting in the virtues of focus groups and relying on the experience, reputation, and track record of those charged with conducting them may be necessary in some types of marketing and opinion research in which rapid turnaround is imperative. Satisfying peer reviewers as to the reliability of methods and data and the validity of conclusions is quite another matter.

Until recently, there was precious little guidance in the literature with respect to analytical methods, reliability, and validity when focus groups were used (Drayton et al., 1989; Morgan, 1986, 1989; Morgan & Spanish, 1985; Templeton, 1987; Tuck, 1976). In recent years, researchers have been more explicit about the analytical difficulties posed by focus group data and about their general approach to data analysis (Agar & MacDonald, 1995; Carey, 1995a; Carey & Smith, 1994; Catterall & Maclaran, 1997; Duncan & Morgan, 1994; Kitzinger, 1994; Knodel, 1993; Krueger, 1997a; McDaniel & Bach, 1996; Morgan, 1993a, 1996, 1997a; Morgan & March, 1992; Morgan & Zhao, 1993; Reed & Payton, 1997; Seals et al., 1995; Sim, 1998). However, too often there is still a frustrating vagueness with respect to specific analytical approaches that can be used to increase confidence in focus group findings. Before addressing specific approaches that we have used to enhance analytical rigor, we will briefly review some of the historical context surrounding the origins of focus group methods that is relevant to their purposes, their limitations, and how the data are analyzed. We will also briefly describe the context of our own work with focus groups to facilitate an understanding of our methods and examples.

HISTORICAL CONTEXT

Focus group methods evolved out of research methods designed by Paul Lazarsfeld, Robert Merton, and colleagues at the Bureau of Applied Social Research at Columbia University to gauge audience responses to propaganda and radio broadcasts during World War II. Lazarsfeld had devised an electronic system for the quantitative recording of positive and negative audience reactions. Merton was invited by Lazarsfeld to develop better debriefing methods (Merton, 1987; Merton,

Fiske, & Kendall, 1952, 1956/1990). Originally called "focused interviews" (Merton, 1987; Merton, Fiske, & Kendall, 1944, 1952; Merton & Kendall, 1946), the methods were somewhat indifferent to whether individual or group interviews were used, although group interviews were common. Merton and colleagues maintained that group interviews could either release or increase inhibitions among group members, depending on the topic and group composition. Advantages of group interviews were said to be a broader range of responses and elicitation of details that might otherwise be overlooked. Disadvantages included sidetracking of discussion to irrelevant issues and competition for dominance among group members (Merton et al., 1944, 1956/1990).

The most salient aspect of focused interviews as opposed to more conventional research interviews was that interviewees were all known to have experienced some "particular concrete situation," which became the focus of the interview (Merton & Kendall, 1946, p. 541). In addition, the investigator(s) had conducted some kind of a priori analysis of that focal situation to develop an interview guide to elicit information relevant to research questions and hypotheses. This should not be construed as implying a highly structured interview. In fact, one of the most consistent threads in focus group literature is the vital importance of using nondirective questions to elicit spontaneous expression among participants. By focusing nondirective questions on responses to specific aspects of a program, product, or event, it became possible to go beyond the mere identification of what was well or ill received. Primary concerns of focussed interviews included (a) specifying which aspects of the focal situation had greater or lesser anticipated or unanticipated effects, (b) identifying the range of effective stimuli within the focal situation and the range of responses to them, and (c) developing interpretive or hypothetical accounts of processes underlying expected and unanticipated responses, including effective, ineffective, and off-putting aspects of the focal situation (Merton et al., 1952, 1956/1990; Merton & Kendall, 1946). In the context of attempts to reach and resonate with an audience in desired ways, the explanatory interpretations thus derived gave rise to hypotheses that could be tested by means of an iterative refinement of a program or product. Following World War II, focus group methods became mainstays of broadcasting, marketing, and public opinion research, but they were largely overlooked in more formal academic and evaluation research until the late 1970s (Merton, 1987).

Understanding this history is important for several reasons. First, focus group methods developed and matured outside of the major methodological traditions of qualitative research, and they are thus relatively agnostic in terms of the methodologies attending them. Consequently, there may well be tension between the commitment to a particular qualitative methodology and the choice of focus groups as a data collection method. For example, group interaction outside of a naturalistic setting is no substitute for the kinds of immersion in lived experience or culture desired in phenomenological or ethnographic research. Constraining discussion to a focal situation or experience may run counter to some of the assumptions underlying grounded theory. In addition, the interview guide in a focus group study may or may not evolve over a series of groups, whereas such evolution is intrinsic to the constant comparative method of grounded theory. More generally, focus group research may be driven by the specific needs of clients who are naive to the rhetoric and philosophical assumptions of qualitative methodologies and are too impatient to invest the time required to understand them.

In any event, one can see from this history that there is some separation between issues pertaining to the focus of an interview and issues pertaining to group dynamics. Both kinds of issues impact analytical approaches, reliability, and validity. An adequate analysis of focus group data should inform the audience credibly about both the focus and the group(s).

FOCUS GROUPS IN THE CONTEXT OF INTERVENTION DEVELOPMENT

Our work with focus groups has been concerned with developing occupational injury prevention interventions and evaluation of safety training programs. We are concerned primarily with employees and owner-operators in small, relatively unregulated operations (e.g., in the past, family farms, and in the present, small construction companies). Occupational injury rates are high in these sectors relative to others, but in terms of person-hours worked and the range of potentially hazardous situations encountered, they are still relatively infrequent. Taken together with the geographical dispersion of work sites, this makes participant observation impractical or prohibitively expensive. In addition, awareness of being observed would very likely alter much of the workplace behavior with which we are concerned.

Traditional approaches to injury prevention emphasize some combination of equipment or workplace engineering, worker training in the use of protective gear, and attempts to regulate behavior. These approaches are important, but they do not account adequately for attitudes, beliefs, and perceived pressures against or barriers to working safely. We depend heavily on focus group interviews to develop a theoretical understanding of cognitive, behavioral, situational, and environmental factors that, from the perspective of workers and owner-operators, contribute to or prevent injuries. We use focus group findings to develop an understanding of how those factors are situated in the social, psychological, and economic context of work. Based on those findings, we develop injury prevention interventions in the form of narrative simulation exercises that emphasize decision making and judgments in relation to characteristic demands and hazards on the job.

In addition to conducting multiple focus groups in different geographical locations, we collect limited demographic data from our focus group participants. We conduct individual interviews with informants who do not participate in the groups (e.g., health care providers) and use multiple external sources of quantitative data (e.g., workers' compensation claims data and hospital records) to find out about injury incidence, type, severity, and cost. However, of all these sources, the focus groups have pride of place. Without the candor and peer interaction from focus group interviews, we would stand little chance of discovering the underlying cognitive schemata, folk models, and narrative patterns. This knowledge is essential to developing credible and emotionally compelling interventions (Kitzinger, 1994).

We record all focus group interviews on audiotape and contract externally for transcription. Following each group, members of the research team conduct a debriefing to identify issues that may affect analysis (e.g., domineering or quiet members), discuss what went well and what did not, and suggest possible

modifications to the interview guide. In the course of analysis, we identify narrative themes and cognitive, behavioral, and motivational elements to embed in the simulations.

After the initial simulation development, we hold a series of authentication focus groups in which the exercises are worked through, discussed, and critiqued. As with the initial round of focus groups, there is still a common fund of experience and similarity in work culture, but the focal situation for the authentication groups is now the simulation. We revise the simulations based on that feedback, conduct further authentication groups, and make additional revisions. Essentially, our focus group data drive the theory and intervention development and initial validation. Only when wider dissemination of the simulations is undertaken do we collect extensive quantitative performance, psychometric, and evaluation data on the simulations. Based on those data, simulations are further refined. The overall sequence of qualitative exploration and theory development with subsequent use of quantitative data in a more confirmatory mode is fairly conventional. However, the use of focus group data for the intervention development and initial validation is less conventional. We feel an obligation to reassure ourselves and our funding agency as to the credibility of the findings before initiating a wider dissemination of the intervention. For the remainder of this article, we will discuss approaches that we use to substantiate claims that the focus group data are reliable and the conclusions are valid.

DATA ACQUISITION ISSUES AFFECTING RELIABILITY AND VALIDITY

A major difference between group and individual interviews is that, in the latter, it is always clear who is speaking. How this is dealt with depends to some degree on the recording method chosen. Audiotape is often easier for a transcriptionist to work with than videotape, but it leaves room for doubt about who said what and does not reproduce nonverbal behavior that may be important. Videotape makes it easier to know who said what, but it may be more cumbersome to transcribe; unless there are multiple cameras or a moving camera operator, which may be unacceptably obtrusive, the increase in nonverbal content is still only partial. Whatever mode of recording is used, some of what is said may be inaudible because it is harder to maintain a constant volume level with a fixed microphone or camera in a room large enough to seat a group comfortably. Regardless of the recording method used, if the discussion gets lively, several people may talk at once and obscure what each is saying. For these reasons, even the best recording and transcription will not reproduce a session completely.

To offset these problems, we recommend that no fewer than two members of the research team be present at every focus group interview. In addition to the moderator, an associate should take detailed notes of the order of speakers and any significant nonverbal behavior. This greatly aids the transcriptionist, and if a few words of each utterance are noted, it facilitates the checking of the transcripts against the tapes. If other research team members attend, which we recommend (usually, we have at least three present), it would be helpful if they take notes about issues to

bring up in debriefing (e.g., about productive and nonproductive parts of the interview).

Another difference between individual and focus group interviews is that reconvening a focus group at a subsequent point in time is often impractical, and even when it can be done, the group dynamics will not be the same. Therefore, most member checking (Denzin & Lincoln, 1994) must be done in real time while each focus group is conducted. At the end of each group, the moderator should present any tentatively identified issues to the members for confirmation or clarification. During the postsession debriefing, members of the research team should discuss these issues and their impressions of agreement, expressed or nonverbal dissent, or coercion.

ANALYSIS ISSUES

There is controversy about whether the individual or the group is the unit of analysis in focus group interviews (Carey, 1995a; Carey & Smith, 1994; Morgan, 1995, 1996). We suggest that neither one is *the unit* of analysis, whereas either or both might be *a focus* of analysis. The trick is to devise analytical approaches sufficiently flexible to identify any undue influence of the group on any individual participant(s), or vice versa, before drawing one's conclusions. The software application we are using, QSR-NUD*IST(r) (Version 4, Qualitative Solutions & Research, 1997), has a number of capabilities that we have found to be valuable in this regard. (The acronym NUD*IST, to be used henceforth, stands for Nonnumerical Unstructured Data Indexing Searching and Theorizing.) A discussion of specific commands and routines in NUD*IST is beyond the scope of this article; however, a few aspects of the software will be touched on to facilitate an understanding of some of the analytical approaches that we describe. We note in passing that several other qualitative data analysis applications appear to have capabilities whereby similar analytical ends might be achieved.

The use of software for qualitative data analysis is not without controversy (Catterall & Maclaran, 1997; Coffey, Holbrook, & Atkinson, 1996; Fielding & Lee, 1991; Lee & Fielding, 1996; Morison & Moir, 1998). The most serious worries are the extent to which the requirements of the software (e.g., formatting of transcripts to exploit software capabilities) and the routinization of certain analytical tasks might (a) distort the underlying context and meaning of remarks or (b) seduce the analyst away from a reflective engagement with the data. In the current version of NUD*IST, for any document imported into a project, any coded text unit is literally only a mouse click or two away from its location and full context in its source document. Indeed, working in NUD*IST does not even require any particular data source to be on the computer. Therefore, the limiting issue is the quality of the data source (and transcription, if any). This is an issue regardless of whether one uses analytical software; therefore, it is not a sufficient argument against its use. With respect to reflective engagement with data, our experience with NUD*IST (especially with Version 4) has been that this is enhanced and not inhibited (e.g., we feel that we find more questions to ask, more issues to ponder, and more vantage points from which to explore the data).

A major aim of analysis with focus group data is to identify areas of agreement and controversy to better understand how perspectives arise and are modified in a group (Carey & Smith, 1994; Reed & Payton, 1997; Sim, 1998). Identifying issues on which disagreements have been voiced is relatively straightforward, but with respect to agreement, the analyst must evaluate whether apparent agreement resulted from coercion or self-censoring of members with alternative viewpoints (Carey & Smith, 1994; Sim, 1998). This implies a group perspective, and a meaningful discourse unit may well incorporate a chunk of consecutive (or even nonconsecutive) statements by several speakers representing multiple, and at times conflicting, viewpoints. Because of the importance of narrative in our work, we code what we have come to call narrative units (a story about an injury occurrence, including questions, comments, asides, and elaborations from any group members). Typically, these narrative units emerge during more conversational periods, after the group members have established rapport among themselves (e.g., after satisfying themselves that they do share some common fund of experience).

Depending on client need or how much prior analysis of the focal situation has taken place, there will probably be a few major issues covered in any focus group interview guide that give rise to sustained discussion. It may well be profitable to code those issues in large, discourse chunks as well. However, because such molar chunks often contain conflicting or contradictory statements, all such material should be subjected to a more detailed coding of substantive content. In essence, we use a combination of broad-brush coding for certain types of discourse (e.g., stories or exchanges between group members about the issues that constitute the focus for the group) and fine-grained (line-by-line) coding of substantive content (what they actually speak about). This kind of cross-coding between molar discourse chunks and more molecular units enhances the analyst's ability to discriminate, for example, between different levels of some category or different points of view in a disagreement. It also permits one to track individuals' attributions to see if or how they are influenced by group interaction, thus facilitating judgments about coerciveness versus spontaneity of agreement.

Using NUD*IST, the analyst can determine the length of text units that are indexed to coding categories. We use individual lines of transcript as the text unit to permit fine-grained coding but format the transcript so that each individual turn (i.e., everything said by one speaker until another speaks) is a section. By constraining line length, text units are more equal in length than turns or (presumed) sentences or paragraphs. Because some group members are inevitably more loquacious or taciturn than others, we believe that less investigator bias is injected if text units are defined arbitrarily by length rather than by value judgments about completeness. Treating each turn as a section in NUD*IST tags the text of an entire turn to the person speaking. Thus, the sorting, indexing, and retrieving functions of the application can be exploited to maximum advantage. For example, as long as each participant has a unique alias, speaker names can be used to sort and code all statements by categories such as sex, age, or role. At a later stage in analysis, patterns of substantive coding (e.g., themes) can be compared across such categories and across all focus groups.

In NUD*IST, there is neither a limit to the number or nature of codes that can be attached to any text unit nor a limit to the number of text units that can be coded en bloc to a category. Of critical importance, when examining the text to which a

particular code is attached, it is very simple to recover all the other codes attached to the same text unit or block of text or to jump to the exact location of any coded text unit in its transcript to recover the entire context.

NUD*IST also has capabilities that can be used or modified to (a) segregate moderator from group discourse, (b) test hypotheses qualitatively using intersections of thematic categories, (c) create matrices for comparing response patterns and frequencies, (d) generate output that can be used to assess similarities and differences between groups with descriptive or nonparametric statistics, and (e) evaluate coding structure in secondary analysis and in confirmatory research. Some of the techniques that we use involve the examination of coding frequencies to aid pattern detection (Morgan, 1993a), but in contrast to quantitative content analysis (Weber, 1990), interpretation is based on contextual analysis and not on frequencies. For example, coding frequencies are useful in identifying groups that have a lot more or a lot less discourse coded to a given category, but interpreting such differences still requires a careful examination of group composition and transcript content.

An important issue with focus groups is gauging whether an issue constitutes a theme for the group or merely a strongly held viewpoint of one or a few members. Assuming that reasonable steps have been taken to recompose who said what and that problems such as attempts to domineer have been identified in debriefing, appropriate adjustments may be taken during data analysis. Again, we have found certain software features to be extremely useful in this regard, specifically, the ability to separate out all statements by any individual or subgroup and calculate how much of any coding category came from only one member or type of member. If necessary, cut-off points for the coding of redundant material can be determined (e.g., coding only the first mention of a topic by a domineering member) to avoid overcoding idiosyncratic or obsessional concerns that are not shared by other members.

Several other issues can be handled in similar fashion: (a) whether an issue only comes up in one or a few groups; (b) whether issues are raised, or returned to, in spontaneous discussion by group members or only come up in response to questions posed by moderators; and (c) whether group members find an issue both important and interesting, merely one but not the other, or neither (Morgan, 1986, 1997a). With respect to the first issue, by assigning a unique code to each transcript, it is easy in NUD*IST to tell at a glance how many transcripts have been coded to any category. With respect to the second, by assigning a unique alias to each group member, it is reasonably simple to segregate all participant statements from all moderator statements in each group. One can conduct separate searches among participants and research team members for the occurrences of some term, phrase, or group of related terms to facilitate judgments about bias arising from moderator input.

Judgments about the importance and interest (Morgan, 1986) or complexity of an issue are more subjective, and a major caveat is that the number of text units coded to a category cannot discriminate between those possibilities. For example, in our construction focus groups, there was generally more discussion of lapses in attention or judgment in injury events than there was of fatigue. It would be foolish to argue on that basis that fatigue was unimportant as a risk factor for injury. A careful examination of transcripts indicated that the notion of fatigue as a risk factor for injury was readily accepted in all groups. However, most members felt that there was little one could do to avoid fatigue, given the physical demands of their work.

Agreement on those two points left little need for further discussion. On the other hand, lapses in attention and judgment were more complicated and interesting to talk about than fatigue, possibly because they were potentially more avoidable. These lapses were not entirely independent of fatigue, but they figured more prominently than fatigue in participants' injury stories. This suggested that such lapses were important risk factors, although they were not necessarily more risky than fatigue.

We have found limited use of descriptive and nonparametric statistics helpful in analyzing differences between groups. For example, examining the number or percentage of text units coded to a given theme across groups makes it easy to identify groups that may be anomalous in terms of how much or how little discussion was devoted to an issue. We have used nonparametric statistics to assess interrater reliability and have, on occasion, used rank or concordance correlations as an index of relative distance or proximity between groups (e.g., the ranking of a series of related codes in terms of the percentage of text units coded to each). When used in this fashion, the statistical significance of correlations is less material than their magnitude. Greater distance between groups warrants another look at the transcripts to determine if there were inconsistencies in coding or if substantive differences between groups account for the dissimilarity. Whether this is necessary or desirable depends on study purposes and client needs.

RELIABILITY AND VALIDITY

Throughout this section, we assume that multiple focus groups have been conducted and that the usual qualitative research procedures have been followed (e.g., maintenance of an audit trail throughout the analysis). We prefer to discuss issues pertaining to reliability and validity with fairly conventional terminology. This facilitates communication between members of our research team, consultants, and funding agency personnel, who represent different disciplines and have varying degrees of familiarity with more recondite terminologies. In essence, whatever terms are used to characterize the various kinds of reproducibility amount to concerns about reliability, whereas concerns about the scope and meaning of content and the generalizability of findings amount to concerns about validity (Brink, 1991; Dreher, 1994; LeCompte & Goetz, 1982).

Reliability

Without wishing to force the analogy, we find it useful to consider reliability in the conventional terms of stability, equivalence, and internal consistency. With focus groups, stability is an issue when the same group is convened on more than one occasion, especially if some members present on one occasion are absent on another. If it is known that a group will be convened on more than one occasion, it may be useful to ask the participants to rank a group of related issues in order of importance on each occasion as a means of testing for stability. Alternatively, it may suffice simply to demonstrate that a similar range of concerns is expressed at different times.

Equivalence is primarily an issue when multiple moderators or coders are used. Differences in moderator experience and interviewing style may affect the flow,

texture, and content of focus group interviews. We have found that with our most experienced moderator a lesser percentage of lines of transcript are uttered by research team members than when a less experienced moderator presides. As one would expect, with the more experienced moderator, group interaction tends to be more spontaneous and conversational, but we found little difference in the range of concerns elicited. When there are multiple coders, the equivalence of coding can be addressed by selecting 10% of the text of a transcript and comparing across two or more raters for consistency using a kappa coefficient. Given the complexity and inhomogeneity of group discourse, it is not reasonable to assume that one rater can independently reproduce all coding for a given statement or chunk of discourse. It is best to constrain testing of agreement to core concepts or themes of substantive importance. The level of agreement should reflect a basic agreement between raters that a given statement or chunk is at least about one or several specific topics, whatever other codes may be attached to it.

The internal consistency of coding is enhanced if one team member has the primary responsibility for conducting the analysis, participates in as many groups and debriefings as possible, and communicates regularly with other team members as the analysis proceeds (e.g., as coding of each transcript is completed). Although this is true of individual interviews as well, it may be more feasible with focus groups because, for a given number of participants, there will be fewer transcripts. However, the heterogeneity of viewpoints and types of discourse in a focus group make consistency of coding across groups an ongoing challenge. The combination of line-by-line coding of substantive content with molar coding of narrative units and discussion about focal issues amounts to a set of coding rules that are internally consistent and potentially reproducible while remaining true to the heterogeneity of discourse in focus groups.

Validity

We advocate a stepwise process for validating findings that begins with the composition of the focus groups. Although we have a mix of work experience in our groups, the vast majority of participants are highly experienced—in our current study, 80% of the focus group participants ($n = 60$) had at least 7 years of experience in construction ($M = 19$ years, $SD = 11$ years). These focus groups amount to panels of experts in the kinds of work that they do. As interviews proceed, we modify the interview guide based on their responses to make sure that we are tapping the range of experiences relevant to the focal situation. As analysis proceeds, members of the research team and external consultants review the coding scheme and the contents and definitions of coded categories. As simulations are developed, we go to additional, similarly constituted focus groups that judge the credibility of what we derived from earlier groups. We maintain that this entire process is one of progressive, iterative content validation.

In other contexts, focus groups may not have this degree of experience (e.g., if one were using focus groups to study adjustment to some new situation). However, because focus group participants can talk it over directly (whatever it may be), the emergence of a substantively similar viewpoint on some issue in multiple focus groups, especially if they are geographically dispersed, will tend to support content validity. Other methods that can be used for assessing content validity include

secondary analysis (e.g., independent coding), content validity indexes (Atwood & Hinds, 1986; Lynn, 1986), or comparison with themes or theory in extant literature (Morgan, 1986; Morgan & Spanish, 1985; Torn & McNichol, 1998).

Beyond content validity, the combination of broad and fine coding that we advocate facilitates (a) the identification of salient aspects of persons, setting, and actions from members' points of view and (b) the contrasting of members' attributions about how things are in general with how things play out when reconstructed as a coherent narrative (cf. Morgan, 1986; Morgan & Spanish, 1985). In our work in occupational injury prevention, we have come to view the content of stories about injury events as a subjective criterion against which the members' attributions in general about injury risk can be evaluated from within their own frame of reference in the light of their experiences (i.e., a subjective analog of criterion-referenced validity). We value first-person or eyewitness accounts above hearsay, and we value risk factors that appear in injury stories across several groups above those found in a story in only one group.

By way of example, consider a variable such as experience. We code all statements pertaining to experience to a category (Experience) and create distinct subcategories for statements pertaining to greater and lesser poles of experience. We code parsimoniously, line by line to these subcategories, leaving any ambiguous or ambivalent statements about experience in the overall category only. If one were to think of this as a qualitative analog to a hypothesis test, the null hypothesis would be that, in the aggregate, the injury stories cross-index as much coding (or coding from as many groups) from the high as from the low experience subcategories. A one-tailed, alternative hypothesis (to which, we hasten to add, many of our participants subscribed) would be that workers with a lot of experience are less likely to get injured than are those with little experience. In NUD*IST, it is a fairly simple matter to explore the coding to facilitate a judgment as to which hypothesis is more credible. Across groups, we found about as much evidence of experienced as inexperienced workers getting injured in the stories. In contrast, we found that, regardless of experience, many injuries occurred when someone was in a hurry or not paying attention to the task at hand, whereas there were few, if any, instances in the stories of injuries occurring when an individual had been working at a reasonable pace and paying attention to the immediate demands of the job.

Construct validity is more a matter of accumulating a pattern of expected convergent and discriminant associations between themes over a series of studies, at least some of which, preferably, will have used other methods. Similar findings derived from multiple sources increase confidence in the validity of constructs and the theoretical generality of relations between them. For example, through focus groups of family farm owners, spouses, and children, we found that economic stressors were associated with agricultural injuries; similar findings have been identified through national surveys performed independently and in injury investigation reports. Returning to the earlier example pertaining to experience, among farmers and construction workers, we have found a similar ambiguity concerning the role of experience as a protective or risk factor for injury. We have found substantial congruity between farmers and construction workers about the degree to which being in a hurry and lapses of attention or judgment are associated unambiguously with injury. Ultimately, by using focus group data to generate interventions that are faithful to their qualitative origins, and yet yield quantitative psychometric and

evaluative data, we have a further check for the validity of risk factors identified through our analysis of focus group data.

CONCLUSION

Several authors have argued that focus group interviews are not adequate as a stand-alone method for social science and need to be augmented by participant observation or other ethnographic methods (Agar & MacDonald, 1995; Reed & Payton, 1997). The history of focus groups suggests that they were not originally conceived as a stand-alone method. Rather, as they were originally used, what have come to be called focus group methods were an early instance of triangulating qualitative and quantitative data from the same participants. We believe that, for purposes of peer-reviewed social and health research, confidence in focus group findings almost always can be enhanced by conducting multiple groups (ideally from multiple sites) and by including other data sources. How much data should be collected from other sources, whether it should be quantitative or qualitative (or both), and whether the focus group data should play first-, second-, or nth-fiddle in relation to other data sources are judgments that depend on the research purposes and the needs of the audience and empirical judgments about the range and saturation of data.

We acknowledge that, in some contexts, some of the issues we have addressed may not be essential to an adequate analysis (e.g., when results are disseminated only to a client for whom it would be too time-consuming). For example, Henderson (1995) and Templeton (1987) describe highly systematic processes that marketing research firms use when analyses have to be completed and placed in the hands of clients in a matter of days to weeks. Although such time pressures argue against performing the kind of analysis that we describe, procedural clarity and consistency give reasonable assurance that results are reproducible and credible. When focus groups are used in more formal research contexts, greater attention needs to be paid to analytical rigor than usually has been evident in focus group literature. Some of the approaches that we describe are in the arsenal of most qualitative researchers. However, little attention has been paid to how they might require modification to adjust for differences between focus group discourse and individual interviews. There are no cookbook solutions to the issues we raise. Although the general lines of approach that we describe are not exhaustive, we do believe that they help us to get both the focus and the group in a manner that is defensible and potentially reproducible. We hope that this article will encourage an increasing dialogue between those who use focus groups about the analytical difficulties posed by their data and the approaches that they use to resolve them.

REFERENCES

Agar, M., & MacDonald, J. (1995). Focus groups and ethnography. *Human Organization, 54*, 78-86.

Atwood, J., & Hinds, P. (1986). Heuristic heresy: Application of reliability and validity criteria to products of grounded theory. *Western Journal of Nursing Research, 6*, 135-147.

Basch, C. (1987). Focus group interview: An underutilized research technique for improving theory and practice in health education. *Health Education Quarterly, 14,* 411-448.

Bertrand, J. T., Brown, J. E., & Ward, V. M. (1992). Techniques for analyzing focus group data. *Evaluation Review, 16,* 198-209.

Brink, P. J. (1991). Issues of reliability and validity. In J. M. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (pp. 164-186). Newbury Park, CA: Sage.

Bristol, T., & Fern, E. F. (1996). Exploring the atmosphere created by focus group interviews: Comparing consumers' feelings across qualitative techniques. *Journal of the Market Research Society, 38,* 185-195.

Carey, M. A. (1995a). Concerns in the analysis of focus group data. *Qualitative Health Research, 5,* 487-495.

Carey, M. A. (Ed.). (1995b). Issues and applications of focus groups [Special issue]. *Qualitative Health Research, 5,* 413-530.

Carey, M. A., & Smith, M. W. (1994). Capturing the group effect in focus groups: A special concern in analysis. *Qualitative Health Research, 4,* 123-127.

Catterall, M., & Maclaran, P. (1997). Focus group data and qualitative analysis programs: Coding the moving picture as well as the snapshots. *Sociological Research Online* [Online], 2. Available: http://www.socresonline.org.uk/socresoline/2/1/6.html

Coffey, A., Holbrook, B., & Atkinson, P. (1996). Qualitative data analysis: Technologies and representations. *Sociological Research Online* [Online], *1.* Available: http://www.socresonline.org.uk/socresoline/1/1/4.html

Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research.* Thousand Oaks, CA: Sage.

DesRosier, M., & Zellars, K. (1989). Focus groups: A program planning technique. *Journal of Nursing Administration, 19,* 20-25.

Drayton, J. L., Fahad, G. A., & Tynan, A. C. (1989). The focus group: A controversial research technique. *Graduate Management Research, 4* (3), 34-51.

Dreher, M. (1994). Qualitative research methods from the reviewer's perspective. In J. M. Morse (Ed.), *Critical issues in qualitative research methodology* (pp. 281-297). Thousand Oaks, CA: Sage.

Duncan, M. T., & Morgan, D. L. (1994). Sharing the caring: Family caregivers' views of their relationships with nursing home staff. *The Gerontologist, 34,* 235-244.

Fern, E. F. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research, 19,* 1-13.

Festervand, T. A. (1985). An introduction and application of focus group research to the health care industry: Marketing ambulatory care services. *Health Marketing Quarterly, 2,* 199-209.

Fielding, N. G., & Lee, R. M. (Eds.). (1991). *Using computers in qualitative research.* Newbury Park, CA: Sage.

Frey, J. H., & Fontana, A. (1991). The group interview in social research. *Social Science Journal, 28,* 175-187.

Henderson, N. R. (1995). A practical approach to analysis and reporting focus group studies: Lessons from qualitative market research. *Qualitative Health Research, 5,* 463-477.

Kaufman, L. (1997, August 18). Enough talk. *Newsweek, 130,* 48-49.

Kingry, M., Tiedje, L., & Friedman, L. (1990). Focus groups: A research technique for nursing. *Nursing Research, 90,* 124-125.

Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health & Illness, 16,* 103-121.

Knodel, J. (1993). The design and analysis of focus group studies: A practical approach. In D. L. Morgan (Ed.), *Successful focus groups: Advancing the state of the art* (pp. 35-50). Newbury Park, CA: Sage.

Krueger, R. A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). Thousand Oaks, CA: Sage.

Krueger, R. A. (1997a). *Analyzing and reporting focus group results.* Thousand Oaks, CA: Sage.

Krueger, R. A. (1997b). *Developing questions for focus groups.* Thousand Oaks, CA: Sage.

Krueger, R. A. (1997c). *Moderating focus groups.* Thousand Oaks, CA: Sage.

Krueger, R. A., & King, J. A. (1997). *Involving community members in focus groups.* Thousand Oaks, CA: Sage.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research, 52,* 31-60.

Lederman, L. C. (1990). Assessing educational effectiveness: The focus group interview as a technique for data collection. *Communication Education, 38,* 117-127.

Lee, R. M., & Fielding, N. G. (1996). Qualitative data analysis: Representations of a technology: A comment on Coffey, Holbrook, and Atkinson. *Sociological Research Online, 1* [Online]. Available: http://www.socresonline.org.uk/socresoline/1/1/4/lf.html

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 382-385.
Madriz, I. (1998). Using focus groups with lower socioeconomic status Latina women. *Qualitative Inquiry, 4,* 114-128.
McDaniel, R. W., & Bach, C. A. (1996). Focus group research: The question of scientific rigor. *Rehabilitation Nursing Research, 5,* 53-59.
Merton, R. K. (1987). The focused interview and focus groups. *Public Opinion Quarterly, 51,* 550-566.
Merton, R. K., Fiske, M., & Kendall, P. L. (1944). *The focused interview.* New York: Bureau of Applied Social Research, Columbia University.
Merton, R. K., Fiske, M., & Kendall, P. L. (1952). *The focused interview* (2nd ed.). New York: Bureau of Applied Social Research, Columbia University.
Merton, R. K., Fiske, M., & Kendall, P. L. (1956/1990). *The focused interview* (2nd ed.). New York: Free Press.
Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology, 51,* 541-557.
Morgan, D. L. (1986). Personal relationships as an interface between social networks and social cognitions. *Journal of Social and Personal Relationships, 3,* 403-422.
Morgan, D. L. (1989). Adjusting to widowhood: Do social networks really make it easier? *The Gerontologist, 29,* 101-107.
Morgan, D. L. (1993a). Qualitative content analysis: A guide to paths not taken. *Qualitative Health Research, 3,* 112-121.
Morgan, D. L. (1993b). *Successful focus groups: Advancing the state of the art.* Newbury Park, CA: Sage.
Morgan, D. L. (1995). Why things (sometimes) go wrong in focus groups. *Qualitative Health Research, 5,* 516-522.
Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology,* 22, 129-152.
Morgan, D. L. (1997a). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.
Morgan, D. L. (1997b). *The focus group guidebook.* Thousand Oaks, CA: Sage.
Morgan, D. L. (1997c). *Planning focus groups.* Thousand Oaks, CA: Sage.
Morgan, D. L., & March, S. J. (1992). The impact of life events on networks of personal relationships: A comparison of widowhood and caring for a spouse with Alzheimer's disease. *Journal of Social and Personal Relationships, 9,* 563-584.
Morgan, D. L., & Spanish, M. T. (1985). Social interaction and the cognitive organisation of health-relevant knowledge. *Sociology of Health and Illness, 7,* 401-422.
Morgan, D. L., & Zhao, P. Z. (1993). The doctor-caregiver relationship: Managing the care of family members with Alzheimer's disease. *Qualitative Health Research, 3,* 133-164.
Morison, M., & Moir, J. (1998). The role of computer software in the analysis of qualitative data: Efficient clerk, research assistant or Trojan horse? *Journal of Advanced Nursing, 28,* 106-116.
Qualitative Solutions & Research Pty. Ltd. (1997). QSR-NUD*IST(r) (Version 4) [Computer software]. Melbourne, Victoria, Australia: Author.
Ramirez, A. G., & Shepperd, J. (1988). The use of focus groups in health research. *Scandinavian Journal of Primary Health Care, 6* (Suppl. 1), 81-90.
Reed, J., & Payton, V. R. (1997). Focus group: Issues of analysis and interpretation. *Journal of Advanced Nursing, 26,* 765-771.
Schindler, R. M. (1992, December). The real lesson of new coke: The value of focus groups for predicting the effects of social influence. *Marketing Research: A Magazine of Management and Applications, 4,* 22-27.
Seals, B. F., Sowell, R. L., Demi, A. S., Moneyham, L., Cohen, L., & Guillory, J. (1995). Falling through the cracks: Social service concerns of women infected with HIV. *Qualitative Health Research, 5,* 496-515.
Sim, J. (1998). Collecting and analysing qualitative data: Issues raised by the focus group. *Journal of Advanced Nursing, 28,* 345-352.
Stewart, D., & Shamdasani, P. (1990). *Focus groups: Theory and practice.* Newbury Park, CA: Sage.
Straw, M. K., & Marks, K. (1995). Use of focus groups in program development. *Qualitative Health Research, 5,* 428-443.
Straw, R. B., & Smith, M. W. (1995). Potential uses of focus groups in federal policy and evaluation studies. *Qualitative Health Research, 5,* 421-427.
Templeton, J. F. (1987). *Focus groups: A guide for marketing and advertising professionals.* Chicago: Probus.
Torn, A., & McNichol, E. (1998). A qualitative study utilizing a focus group to explore the role and concept of the nurse practitioner. *Journal of Advanced Nursing, 27,* 1202-1211.
Tuck, M. (1976). *How do we choose?* London: Methuen.

Weber, R. P. (1990). *Basic content analysis* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-049). Newbury Park, CA: Sage.

Pamela S. Kidd is an associate professor of nursing at the University of Kentucky College of Nursing and the director of the Kentucky Injury Prevention and Research Center in Lexington, Kentucky.

Mark B. Parshall is an assistant professor of nursing at the University of New Mexico College of Nursing, and was formerly a doctoral candidate in nursing at the University of Kentucky College of Nursing and a research associate at the Kentucky Injury Prevention and Research Center, Lexington, Kentucky.