TO

THE FACULTY OF
UNION GRADUATE SCHOOL,
THE UNION INSTITUTE,
CINCINNATI, OHIO

for their

Friendship and Scholarship,
and their commitments to
Methodological Eclecticism,
Interdisciplinary Inquiry,
Integration of Theory and Practice,
Valuing both Reflection and Action,
Scholarship that is Socially Relevant and Meaningful,
Individualized Professional and Personal Development,
Lifelong Learning,
Social Justice and Equity,
Human Diversity and Global Community;
a Scholarly Community governed by Principles and Processes
rather than rules and regulations,
and
Innovations in Learning-Centered Nontraditional
Doctoral Education,
including faculty meetings that are interesting and important—
an indication of innovation of the highest order.

QUALITATIVE EVALUATION AND RESEARCH METHODS
SECOND EDITION
Michael Quinn Patton

SAGE PUBLICATIONS
The International Professional Publishers
Newbury Park  London  New Delhi
Time sampling (sampling periods or units of time) can be an especially important approach because programs may function in different ways at different times during the year. Of course, with some programs there is never a good time to collect data. I have learned that this is the case with schools. Educators will tell you that you don't want to collect data in the schools before Halloween because the school year is just getting started and the kids aren't quite fixed in the patterns that will be maintained later in the year. The period between Halloween and Thanksgiving is really too short to do very much, and, then, of course, after Thanksgiving everybody's getting ready for Christmas, so that's not a typical or convenient period. It then takes students a few weeks after Christmas to get their attention focused back on school and then the winter malaise sets in and both teachers and students become deeply depressed with the endlessness of winter (at least in Minnesota). Then, of course, once spring hits, attention is focused on the close of school and the kids want to be outside, so that's not an effective time to gather data.

There are limits to how much one can apply logic and deduction in making sampling decisions, whether the decision is about which time periods to sample or which activities to observe. The trick is to keep coming back to the criterion of usefulness. What data collected during what time period describing what activities would make a difference? There are no perfect evaluation designs, only more and less useful ones.

The key issue in selecting and making decisions about the appropriate unit of analysis is to decide what it is you want to be able to say something about at the end of the study. Do you want to have findings about individuals, families, groups, or some other unit of analysis? At what level do decision makers really need information? Do they want information about the different experiences of individuals in programs or do they want to know about variations in program processes at different sites? These are differences in focus that are critical to the design but may not be easy to determine. The decision maker typically will be unable to say to the evaluator, "The unit of analysis we want to study is _____." The evaluator must be able to hear the real issues involved in the decision maker's questions and translate those issues into the appropriate unit of analysis, then check out that translation with the intended evaluation users.

Purposeful Sampling

Perhaps nothing better captures the difference between quantitative and qualitative methods than the different logics that undergird sampling approaches. Qualitative inquiry typically focuses in depth on relatively small samples, even single cases (n = 1), selected purposefully. Quantitative methods typically depend on larger samples selected randomly. Not only are the techniques for sampling different, but the very logic of each approach is unique because the purpose of each strategy is different.

The logic and power of probability sampling depends on selecting a truly random and statistically representative sample that will permit confident generalization from the sample to a larger population. The purpose is generalization.

The logic and power of purposive sampling lies in selecting information-rich cases for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research, thus the term purposive sampling. For example, if the purpose of an evaluation is to increase the effectiveness of a program in reaching lower-socioeconomic groups, one may learn a great deal more by focusing on depth on understanding the needs, interests, and incentives of a small number of carefully selected poor families than by gathering standardized information from a large, statistically representative sample of the whole program. The purpose of purposive sampling is to select information-rich cases whose study will illuminate the questions under study.

There are several different strategies for purposively selecting information-rich cases. The logic of each strategy serves a particular evaluation purpose.

1. Extreme or deviant case sampling. This approach focuses on cases that are rich in information because they are unusual or special in some way. Unusual or special cases may be particularly troublesome or especially enlightening, such as outstanding successes or notable failures. If, for example, the evaluation was aimed at gathering data to help a national program reach more clients, one might compare a few project sites that have long waiting lists with those that have short waiting lists. If staff morale was an issue, one might study and compare high-morale programs to low-morale programs.
The logic of extreme case sampling is that lessons may be learned about unusual conditions or extreme outcomes that are relevant to improving more typical programs. Let's suppose that we are interested in studying a national program with hundreds of local sites. We know that many programs are operating reasonably well, even quite well, and that other programs verge on being disasters. We also know that most programs are doing "okay." This information comes from knowledgeable sources who have made site visits to enough programs to have a basic idea about what the variation is. The question is this: How should programs be sampled for the study? If one wanted to precisely document the natural variation among programs, a random sample would be appropriate, preferably a random sample of sufficient size to be truly representative of and permit generalizations to the total population of programs. However, some information is already available on what program variation is like. This question of more immediate interest may concern extreme cases. With limited resources and limited time an evaluator might learn more by intensively studying one or more examples of really poor programs and one or more examples of really excellent programs. The evaluation focus, then, becomes a question of understanding under what conditions programs get into trouble and under what conditions programs exemplify excellence. It is not even necessary to randomly sample poor programs or excellent programs. The researchers and intended users involved in the study think through what cases they could learn the most from and those are the cases that are selected for study.

In a single program the same strategy may apply. Instead of studying some representative sample of people in the setting, the evaluator may focus on studying and understanding selected cases of special interest, for example, unexpected dropouts or outstanding successes. In many instances more can be learned from intensively studying extreme or unusual cases than can be learned from statistical depictions of what the average case is like. In other evaluations detailed information about special cases can be used to supplement statistical data about the normal distribution of participants.

Ethnomethodologists use a form of extreme case sampling when they do their field experiments. Ethnomethodologists are interested in everyday experiences of routine living that depend on deeply understood, shared understandings among people in a setting (see Chapter 3). One way of exposing these implicit assumptions and norms on which everyday life is based is to create disturbances that deviate from the norm. Observing the reactions to someone eating like a pig in a restaurant and then interviewing people about what they saw and how they felt would be an example of studying a deviant sample to illuminate the ordinary.

The Peters and Waterman (1982) best-selling study of "America's best run companies," *In Search of Excellence*, exemplifies the logic of purposeful, extreme group sampling. Their study was based on a sample of 62 companies "never intended to be perfectly representative of U.S. industry as a whole... [but] a list of companies considered to be innovative and excellent by an informed group of observers of the business scene" (Peters and Waterman, 1982: 19).

Another excellent example of extreme group sampling is Angela Browne's (1987) study, *When Battered Women Kill*. She conducted in-depth studies of the most extreme cases of domestic violence to elucidate the phenomenon of battering and abuse. The extreme nature of the cases presented are what render them so powerful. Browne's book is an exemplar of qualitative inquiry using purposeful sampling for applied research.

2. *Intensive sampling*. Intensive sampling involves the same logic as extreme case sampling but with less emphasis on the extremes. An intensity sample consists of information-rich cases that manifest the phenomenon of interest intensely (but not extremely). Extreme or deviant cases may be so unusual as to distort the manifestation of the phenomenon of interest. Using the logic of intensity sampling, one seeks excellent or rich examples of the phenomenon of interest, but not unusual cases.

Heuristic research uses intensity sampling. Heuristic research draws explicitly on the intense personal experiences of the researcher, for example, experiences with loneliness or jealousy. Coresearchers who have experienced these phenomena intensely also participate in the study (see Chapter 3). The heuristic researcher is not typically seeking pathological or extreme manifestations of loneliness, jealousy, or whatever phenomenon is of interest. Such extreme cases might not lend themselves to the reflective process of heuristic inquiry. On the other hand, if the experience of the heuristic researcher and his or her coresearchers is quite mild, there won't be much to study. Thus the researcher seeks a sample of sufficient intensity to elucidate the phenomenon of interest.

The same logic applies in a program evaluation. Extreme successes or unusual failures may be discredited as being too extreme or un-
usual for gaining information. Therefore, the evaluator may select cases that manifest sufficient intensity to illuminate the nature of success or failure, but not at the extreme.

Intensity sampling involves some prior information and considerable judgment. The researcher must do some exploratory work to determine the nature of the variation in the situation under study. One can then sample intense examples of the phenomenon of interest.

(3) Maximum variation sampling. This strategy for purposeful sampling aims at capturing and describing the central themes or principal outcomes that cut across a great deal of participant or program variation. For small samples a great deal of heterogeneity can be a problem because individual cases are so different from each other. The maximum variation sampling strategy turns that apparent weakness into a strength by applying the following logic: Any common patterns that emerge from great variation are of particular interest and value in capturing the core experiences and central, shared aspects or impacts of a program.

How does one maximize variation in a small sample? One begins by identifying diverse characteristics or criteria for constructing the sample. Suppose a statewide program has project sites spread around the state, some in rural areas, some in urban areas, and some in suburban areas. The evaluation lacks sufficient resources to randomly select enough project sites to generalize across the state. The evaluator can at least be sure that the geographical variation among sites is represented in the study.

When selecting a small sample of great diversity, the data collection and analysis will yield two kinds of findings: (1) high-quality, detailed descriptions of each case, which are useful for documenting uniqueness, and (2) important shared patterns that cut across cases and derive their significance from having emerged out of heterogeneity.

The same strategy can be used within a single program in selecting individuals for study. By including in the sample individuals the evaluator determines have had quite different experiences, it is possible to more thoroughly describe the variation in the group and to understand variations in experiences while also investigating core elements and shared outcomes. The evaluator using a maximum variation sampling strategy would not be attempting to generalize findings to all people or all groups but would be looking for information that elucidates programmatic variation and significant common patterns within that variation.

(4) Homogeneous samples. In direct contrast to maximum variation sampling is the strategy of picking a small homogeneous sample. The purpose here is to describe some particular subgroup in depth. A program that has many different kinds of participants may need in-depth information about a particular subgroup. For example, a parent education program that involves many different kinds of parents may focus a qualitative evaluation on the experiences of single-parent female heads of household because that is a particularly difficult group to reach and hold in the program.

Focus group interviews are typically based on homogeneous groups. Focus group interviews involve conducting open-ended interviews with groups of five to eight people on specially targeted or focused issues. The use of focus groups in evaluation will be discussed at greater length in the chapter on interviewing. The point here is that sampling for focus groups typically involves bringing together people of similar backgrounds and experiences to participate in a group interview about major program issues that affect them.

(5) Typical case sampling. In describing a program or its participants to people not familiar with the program it can be helpful to provide a qualitative profile of one or more “typical” cases. These cases are selected with the cooperation of key informants, such as program staff or knowledgeable participants, who can help identify what is typical. It is also possible to select typical cases from survey data, a demographic analysis of averages, or other programmatic data that provide a normal distribution of characteristics from which to identify “average” examples. Keep in mind that the purpose of a qualitative profile of one or more typical cases is to describe and illustrate what is typical to those unfamiliar with the program—not to make generalized statements about the experiences of all participants. The sample is illustrative not definitive.

When entire programs or communities are the unit of analysis, it is also possible to sample somewhat typical cases. Again, the study of such typical programs does not, of course, permit generalizations in any rigorous sense. It does, however, mean that the processes and effects described for the typical program need not be dismissed as peculiar to “poor” sites or “excellent” sites. When the typical site sampling strategy is used, the site is specifically selected because it is not in any major way atypical, extreme, deviant, or intensely unusual. This strategy is often appropriate in sampling villages for community development studies in Third World countries. A study of a typical
village illuminates key issues that must be considered in any development project aimed at this kind of village.

Decision makers may have made their peace with the fact that there will always be some poor programs and some excellent programs, but the programs they really want more information about are those run-of-the-mill programs that are "hard to get a handle on." It is important, when using this strategy, to attempt to get broad consensus about which programs are "typical." If a number of such programs are identified, only a few can be studied, and there is no other basis for selecting among them purposefully, then it is possible to randomly select from among all "typical" programs identified to select those few typical cases that actually will be included in the study.

(6) Stratified purposeful sampling. It is also clearly possible to combine a typical case sampling strategy with others, essentially taking a stratified purposeful sample of above average, average, and below average cases. This is less than a full maximum variation sample. The purpose of a stratified purposeful sample is to capture major variations rather than to identify a common core, although the latter may also emerge in the analysis. Each of the strata would constitute a fairly homogeneous sample. This strategy differs from stratified random sampling in that the sample sizes are likely to be too small for generalization or statistical representativeness.

(7) Critical case sampling. Another strategy for selecting purposeful samples is to look for critical cases. Critical cases are those that can make a point quite dramatically or are, for some reason, particularly important in the scheme of things. A clue to the existence of a critical case is a statement to the effect that "if it happens there, it will happen anywhere," or, vice versa, "if it doesn't happen there, it won't happen anywhere." The focus of the data gathering in this instance is on understanding what is happening in that critical case. Another clue to the existence of a critical case is a key informant observation to the effect that "if that group is having problems, then we can be sure all the groups are having problems."

Looking for the critical case is particularly important where resources may limit the evaluation to the study of only a single site. Under such conditions it makes strategic sense to pick the site that would yield the most information and have the greatest impact on the development of knowledge. While studying one or a few critical cases does not technically permit broad generalizations to all possible cases, logical generalizations can often be made from the weight of evidence produced in studying a single, critical case.

Physics provides a good example of such a critical case. In Galileo's study of gravity he wanted to find out if the weight of an object affected the rate of speed at which it would fall. Rather than randomly sampling objects of different weights in order to generalize to all objects in the world, he selected a critical case—the feather. If in a vacuum, as he demonstrated, a feather fell at the same rate as some heavier object (a coin), then he could logically generalize from this one critical case to all objects. His findings were enormously useful and credible.

There are many comparable critical cases in social science research—if one is creative in looking for them. For example, suppose national policymakers want to get local communities involved in making decisions about how their local program will be run, but they aren't sure that the communities will understand the complex regulations governing their involvement. The first critical case is to evaluate the regulations in a community of well-educated citizens; if they can't understand the regulations, then less-educated folks are sure to find the regulations incomprehensible. Or conversely, one might consider the critical case to be a community consisting of people with quite low levels of education: "If they can understand the regulations, anyone can."

Identification of critical cases depends on recognition of the key dimensions that make for a critical case. A critical case might be indicated by the financial state of a program; a program with particularly high or particularly low cost-per-client ratios might suggest a critical case. A critical case might come from a particularly difficult program location. If the funder of a new program are worried about recruiting clients or participants into a program, it may make sense to study the site where resistance to the program is expected to be greatest to provide the most rigorous test of the possibility of program recruitment. If the program works in that site, "It could work anywhere."

World-renowned medical hypnotist Milton H. Erickson became a critical case in the field of hypnosis. Erickson was so skillful that he became widely known for "his ability to succeed with 'impossibles'—people who have exhausted the traditional medical, dental, psychotherapeutic, hypnotic and religious avenues for assisting them in their
need, and have not been able to make the changes they desire” (Grinder et al., 1977: 109). If Milton Erickson couldn’t help, no one could help. He was able to demonstrate that anyone could be hypnotized.

(9) Snowball or chain sampling. This is an approach for locating information-rich key informants or critical cases. The process begins by asking well-situated people: “Who knows a lot about ______? Who should I talk to?” By asking a number of people who else to talk with, the snowball gets bigger and bigger as you accumulate new information-rich cases. In most programs or systems, a few key names or incidents are mentioned repeatedly. Those people or events recommended as valuable by a number of different informants take on special importance. The chain of recommended informants will typically diverge initially as many possible sources are recommended, then converge as a few key names get mentioned over and over.

The Peters and Waterman (1982) study In Search of Excellence began with snowball sampling, asking a broad group of knowledgeable people to identify well-run companies. Another excellent and well-known example was Rosabeth Moss Kanter’s (1983) study of innovation reported in The Change Masters. Her book focused on ten core case studies. She began her search for the “best” or “most innovative” companies by getting the views of corporate experts in human resource fields. Nominations for cases to study snowballed from there and then converged into a small number of core cases nominated by a number of different informants.

(9) Criterion sampling. The logic of criterion sampling is to review and study all cases that meet some predetermined criterion of importance. This approach is common in quality assurance efforts. For example, the expected range of participation in a mental health outpatient program might be 4 to 26 weeks. All cases that exceed 28 weeks are reviewed and studied to find out what is happening and to make sure the case is being appropriately handled.

Critical incidents can be a source of criterion sampling. For example, all incidents of client abuse in a program may be objects of in-depth evaluation in a quality assurance effort. All former mental health clients who commit suicide within three months of release may constitute a sample for in-depth, qualitative study. In a school setting, all students who are absent more than half the time may merit the in-depth attention of a qualitative case study. The point of criterion sampling is to be sure to understand cases that are likely to be information-rich because they may reveal major system weaknesses that become targets of opportunity for program or system improvement.

Criterion sampling can add an important qualitative component to a management information system or an ongoing program monitoring system. All cases in the data system that exhibit certain predetermined criterion characteristics are routinely identified for in-depth, qualitative analysis. Criterion sampling also can be applied to identify cases from quantitative questionnaires or tests for in-depth follow-up.

(10) Theory-based or operational construct sampling. A more formal basic research version of criterion sampling is theory-based sampling. The researcher samples incidents, slices of life, time periods, or people on the basis of their potential manifestation or representation of important theoretical constructs. The sample becomes, by definition, representative of the phenomenon of interest. An ecological psychologist (see Chapter 3) is interested, for example, in studying the interaction between a person and the environment. Instances of such interaction must be defined based on theoretical premises in order to study examples that represent the phenomenon of interest.

This differs from the more practical sampling in program evaluation. The evaluator doesn’t need a theory-based definition of “program” because the entity to be studied is usually legally or financially defined. However, to sample social science phenomena that represent theoretical constructs of interest, one must define the construct to be sampled, such as person-environmental interactions or instances of social deviance, identity crisis, creativity, or power interactions in an organization.

When one is studying people, programs, organizations, or communities, the population of interest can be fairly readily determined. Constructs do not have as clear a frame of reference; neither does time.
correspondence to the construct of interest. Alternatively, we can use the same procedures to select multiple operational representations of each construct, chosen because they overlap in representing the critical theoretical components of the construct and because they differ from each other on irrelevant dimensions. This second form of sampling is called multiple operationalism, and it depends more heavily on individual judgment than does the random sampling of persons from a well-designated, target population. Yet such judgments, while inevitable, are less well understood than formal sampling methods and are largely ignored by sampling experts. (Cook et al., 1985: 163-64)

"Operational construct" sampling simply means that one samples for study real-world examples (i.e., operational examples) of the constructs in which one is interested. Studying a number of such examples is called "multiple operationalism" (Webb et al., 1966).

(11) Confirming and disconfirming cases. In the early part of qualitative fieldwork the evaluator is exploring—gathering data and beginning to allow patterns to emerge. Over time the exploratory process gives way to confirmatory fieldwork. This involves testing ideas, confirming the importance and meaning of possible patterns, and checking out the viability of emergent findings with new data and additional cases. This stage of fieldwork requires considerable rigor and integrity on the part of the evaluator in looking for and sampling confirming as well as disconfirming cases.

Confirmatory cases are additional examples that fit already emergent patterns; these cases confirm and elaborate the findings, adding richness, depth, and credibility. Disconfirming cases are no less important at this point. These are the examples that don't fit. They are a source of rival interpretations as well as a way of placing boundaries around confirmed findings. They may be "exceptions that prove the rule" or exceptions that disconfirm and alter what appeared to be primary patterns.

The source of questions or ideas to be confirmed or disconfirmed may be from stakeholders or previous scholarly literature rather than the evaluator's fieldwork. An evaluation may in part serve the purpose of confirming or disconfirming stakeholder's or scholars' preconceptions, these having been identified during early, conceptual evaluator-stakeholder design discussions or literature reviews.

Thinking about the challenge of finding confirming and disconfirming cases emphasizes the relationship between sampling and research conclusions. The sample determines what the evaluator will have something to say about—thus the importance of sampling carefully and thoughtfully.

(12) Opportunistic sampling. Fieldwork often involves on-the-spot decisions about sampling to take advantage of new opportunities during actual data collection. Unlike experimental designs, qualitative inquiry designs can include new sampling strategies to take advantage of unforeseen opportunities after fieldwork has begun. Being open to following wherever the data lead is a primary strength of qualitative strategies in research. This permits the sample to emerge during fieldwork.

When observing, it is not possible to capture everything. It is, therefore, necessary to make decisions about which activities to observe, which people to observe and interview, and what time periods will be selected to collect data. These decisions cannot all be made in advance. The purposeful sampling strategies discussed above often depend on some knowledge of the setting being studied. Opportunistic sampling takes advantage of whatever unfolds as it unfolds.

(13) Purposeful random sampling. The fact that a small sample size will be chosen for in-depth qualitative study does not automatically mean that the sampling strategy should not be random. For many audiences, random sampling, even of small samples, will substantially increase the credibility of the results. I recently worked with a program that annually appears before the state legislature and tells "war stories" about client successes, sometimes even including a few stories about failures to provide balance. They decided they wanted to begin collecting evaluation information. Because they were striving for individualized outcomes they rejected the notion of basing the evaluation entirely on a standardized pre-post instrument. They wanted to collect case histories and do in-depth case studies of clients, but they had very limited resources and time to devote to such data collection. In effect, staff at each program site, many of whom serve 200 to 300 families a year, felt that they could only do 10 or 15 detailed, in-depth clinical case histories each year. We systematized the kind of information that would be going into the case histories at each program site and then set up a random procedure for selecting those clients whose case histories would be recorded in depth. Essentially, this program thereby systematized and randomized their collection of "war stories." While they cannot generalize to the entire client
population on the basis of 10 cases from each program site, they will be able to tell legislators that the stories they are reporting were randomly selected in advance of knowledge of how the outcomes would appear and that the information collected was comprehensive. The credibility of systematic and randomly selected case examples is considerably greater than the personal, ad hoc selection of cases to report after the fact—that is, after outcomes are known.

It is critical to understand, however, that this is a purposeful random sample, not a representative random sample. The purpose of a small random sample is credibility, not representativeness. A small, purposeful random sample aims to reduce suspicion about why certain cases were selected for study, but such a sample still does not permit statistical generalizations.

(14) Sampling politically important cases. Evaluation is inherently and inevitably political to some extent (see Palumbo, 1987; Patton, 1986, 1987b; Turpin, 1989). A variation of the critical case sampling strategy involves selecting (or sometimes avoiding) a politically sensitive site or unit of analysis. For example, a statewide program may have a local site in the district of a state legislator who is particularly influential. By studying carefully the program in that district, evaluation data may be more likely to attract attention and get used. This does not mean that the evaluator then undertakes to make that site look either good or bad, depending on the politics of the moment. This is simply an additional sampling strategy for trying to increase the usefulness and utilization of information where resources permit the study of only a limited number of cases.

The same (broadly speaking) political perspective may inform case sampling in applied or even basic research studies. A political scientist or historian might select the Watergate or Iran-Contra scandals for study not only because of the insights they provide about the American system of government but because of the likely attention such a study would attract. A sociologist's study of a riot or a psychologist's study of a famous suicide would likely involve some attention during sampling to the political importance of the case.

(15) Convenience sampling. Finally, there is the strategy of sampling by convenience: doing what's fast and convenient. This is probably the most common sampling strategy—and the least desirable. Too often evaluators using qualitative methods think that, because the sample size they can study is too small to permit generalizations, it doesn’t matter how cases are picked, so they might as well pick ones that are easy to access and inexpensive to study. While convenience and cost are real considerations, they should be the last factors to be taken into account after strategically deliberating on how to get the most information of greatest utility from the limited number of cases to be sampled. Purposeful, strategic sampling can yield crucial information about critical cases. Convenience sampling is neither purposeful nor strategic.

Information-Rich Cases

Table 5.5 summarizes the 15 purposeful sampling strategies discussed above, plus a 16th approach—combination or mixed purposeful sampling. For example, an extreme group or maximum heterogeneity approach may yield an initial potential sample size that is still larger than the study can handle. The final selection, then, may be made randomly—a combination approach. Thus these approaches are not mutually exclusive. Each approach serves a somewhat different purpose. Because research and evaluations often serve multiple purposes, more than one qualitative sampling strategy may be necessary. In long-term fieldwork all of these strategies may be used at some point.

These are not the only ways of sampling qualitatively. The underlying principle that is common to all these strategies is selecting information-rich cases. These are cases from which one can learn a great deal about matters of importance. They are cases worthy of in-depth study.

In the process of developing the research design, the evaluator or researcher is trying to consider and anticipate the kinds of arguments that will lend credibility to the study as well as the kinds of arguments that might be used to attack the findings. Reasons for site selections or individual case sampling need to be carefully articulated and made explicit. Moreover, it is important to be open and clear about the study's limitations, including how any particular purposeful sampling strategy may lead to distortion in the findings—that is, to anticipate criticisms that will be made of a particular sampling strategy.

Having weighed the evidence and considered the alternatives, evaluators and primary stakeholders make their sampling decisions, sometimes painfully, but always with the recognition that there are no perfect designs. The sampling strategy must be selected to fit the purpose of the study, the resources available, the questions being
Table 5.5 Sampling Strategies

<table>
<thead>
<tr>
<th>Type</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Random probability sampling</td>
<td>Represents sample size as a function of population size and desired confidence level.</td>
</tr>
<tr>
<td>1. simple random sample</td>
<td>Permits generalization from sample to the population it represents.</td>
</tr>
<tr>
<td>2. stratified random and cluster samples</td>
<td>Increases confidence in making generalizations to particular subgroups or areas.</td>
</tr>
<tr>
<td>B. Purposeful sampling</td>
<td>Selects information-rich cases for in-depth study. Size and specific cases depend on study purpose.</td>
</tr>
</tbody>
</table>

• 1. extreme or deviant case sampling
  Learning from highly unusual manifestations of the phenomenon of interest, such as outstanding successes/notable failures, top of the class/dropouts, exotic events, crises.

• 2. intensity sampling
  Information-rich cases that manifest the phenomenon intensely, but not extremely, such as good students/poor students, above average/below average.

• 3. maximum variation sampling—purposefully picking a wide range of variation on dimensions of interest
  Documents unique or diverse variations that have emerged in adapting to different conditions. Identifies important common patterns that cut across variations.

• 4. homogeneous sampling
  Focuses, reduces variation, simplifies analysis, facilitates group interviewing.

• 5. typical case sampling
  Illustrates or highlights what is typical, normal, average.

• 6. stratified purposeful sampling
  Illustrates characteristics of particular subgroups of interest; facilitates comparisons.

• 7. critical case sampling
  Permits logical generalization and maximum application of information to other cases because if it's true of this one case it's likely to be true of all other cases.

• 8. snowball or chain sampling
  Identifies cases of interest from people who know people who know people who know what cases are information-rich, that is, good examples for study, good interview subjects.

Table 5.5 (continued)

<table>
<thead>
<tr>
<th>Type</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. criterion sampling</td>
<td>Picking all cases that meet some criterion, such as all children abused in a treatment facility. Quality assurance.</td>
</tr>
<tr>
<td>10. theory-based or operational construct sampling</td>
<td>Finding manifestations of a theoretical construct of interest so as to elaborate and examine the construct.</td>
</tr>
<tr>
<td>11. confirming and disconfirming cases</td>
<td>Elaborating and deepening initial analysis, seeking exceptions, testing variation.</td>
</tr>
</tbody>
</table>

⇒ 12. opportunistic sampling
  Following new leads during fieldwork, taking advantage of the unexpected, flexibility.

⇒ 13. random purposeful sampling
  (still small sample size)
  Adds credibility to sample when potential purposeful sample is larger than one can handle. Reduces judgment within a purposeful category. (Not for generalizations or representativeness.)

⇒ 14. sampling politically important cases
  Attracts attention to the study (or avoids attracting undesired attention by purposefully eliminating from the sample politically sensitive cases).

15. convenience sampling
  Saves time, money, and effort. Poorest rationale; lowest credibility. Yields information-poor cases.

⇒ 16. combination or mixed purposeful sampling
  Triangulation, flexibility, meets multiple interests and needs.

asked, and the constraints being faced. This holds true for sampling strategy as well as sample size.

SAMPLE SIZE

Qualitative inquiry is rife with ambiguities. There are purposeful strategies instead of methodological rules. There are inquiry approaches instead of statistical formulas. Qualitative inquiry seems to work best for people with a high tolerance for ambiguity. (And we’re still only discussing design. It gets worse when we get to analysis.)
Nowhere is this ambiguity clearer than in the matter of sample size. I get letters. I get calls. "Is 10 a large enough sample to achieve maximum variation?"

"I started out to interview 20 people for 2 hours each, but I've lost 2 people. Is 18 large enough, or do I have to find 2 more?"

"I want to study just one organization, but interview 20 people in the organization. Is my sample size 1 or 20 or both?"

My universal, certain, and confident reply to these questions is this: "it depends."

*There are no rules for sample size in qualitative inquiry. Sample size depends on what you want to know, the purpose of the inquiry, what's at stake, what will be useful, what will have credibility, and what can be done with available time and resources.

Earlier in this chapter, I discussed the trade-off between breadth and depth. With the same fixed resources and limited time, a researcher could study a specific set of experiences for a larger number of people (seeking breadth) or a more open range of experiences for a smaller number of people (seeking depth). *In-depth information from a small number of people can be very valuable, especially if the cases are information-rich.* Less depth from a larger number of people can be especially helpful in exploring a phenomenon and trying to document diversity or understand variation. I repeat, the size of the sample depends on what you want to find out, why you want to find it out, how the findings will be used, and what resources (including time) you have for the study.

To understand the problem of small samples in qualitative inquiry, it's necessary to place these small samples in the context of probability sampling. A qualitative inquiry sample *only seems small* in comparison with the sample size needed for representativeness when the purpose is generalizing from a sample to the population of which it is a part. Suppose there are 100 people in a program to be evaluated. It would be necessary to randomly sample 80 of those people (80%) to make a generalization at the 95% confidence level. If there are 500 people in the program, 217 people must be sampled (43%) for the same level of confidence. If there are 1,000 people, 278 people must be sampled (28%); and if there are 5,000 people in the population of interest, 357 must be sampled (7%) to achieve a 95% confidence level in the generalization of findings. At the other extreme, if there are only 50 people in the program, 44 must be randomly sampled (88%) to achieve a 95% level of confidence. (See Fitzgibbon and Morris, 1987: 163, for a table on determining sample size from a given population.)

The logic of purposeful sampling is quite different from the logic of probability sampling. The problem is, however, that the utility and credibility of small purposeful samples are often judged on the basis of the logic, purpose, and recommended sample sizes of probability sampling. *What should happen is that purposeful samples be judged on the basis of the purpose and rationale of each study and the sampling strategy used to achieve the study's purpose. The sample, like all other aspects of qualitative inquiry, must be judged in context—the same principle that undergirds analysis and presentation of qualitative data. Random probability samples cannot accomplish what in-depth, purposeful samples accomplish, and vice versa.*

Piaget contributed a major breakthrough to our understanding of how children think by observing his own two children at length and in great depth. Freud established the field of psychoanalysis based on fewer than ten client cases. Bandura and Grinder (1975a, 1975b) founded neurolinguistic programming (NLP) by studying three renowned and highly effective therapists: Milton Erickson, Fritz Perls, and Virginia Satir. Peters and Waterman (1982) formulated their widely followed eight principles for organizational excellence by studying 62 companies, a very small sample of the thousands of companies one might study.

*The validity, meaningfulness, and insights generated from qualitative inquiry have more to do with the information-richness of the cases selected and the observational/analytical capabilities of the researcher than with sample size.*

This issue of sample size is a lot like the problem students have when they are assigned an essay to write.

Student: "How long does the paper have to be?"

Instructor: "Long enough to cover the assignment."

Student: "But how many pages?"

Instructor: "Enough pages to do justice to the subject—no more, no less."

Lincoln and Guba (1985: 202) recommend sample selection to the point of redundancy. . . . In purposeful sampling the size of the sample is determined by informational considerations. If the purpose
is to maximize information, the sampling is terminated when no new information is forthcoming from new sampled units; thus redundancy is the primary criterion. (emphasis in the original)

This strategy leaves the question of sample size open. There remains, however, the practical problems of how to negotiate an evaluation budget or how to get a dissertation committee to approve a design if you don’t have some idea of sample size. Sampling to the point of redundancy is an ideal, one that works best for basic research, unlimited time lines, and unconstrained resources.

The solution is judgment and negotiation. I recommended that qualitative sampling designs specify minimum samples based on expected reasonable coverage of the phenomenon given the purpose of the study and stakeholder interests. One may add to the sample as fieldwork unfolds. One may change the sample if information emerges that indicates the value of a change. The design should be understood to be flexible and emergent. Yet, at the beginning, for planning and budgetary purposes, one specifies a minimum expected sample size and builds a rationale for that minimum, as well as criteria that would alert the researcher to inadequacies in the original sampling approach and/or size.

In the end, sample size adequacy, like all aspects of research, is subject to peer review, consensus validation, and judgment. What is crucial is that the sampling procedures and decisions be fully described, explained, and justified so that information users and peer reviewers have the appropriate context for judging the sample. The researcher or evaluator is absolutely obligated to discuss how the sample affected the findings, the strengths and weaknesses of the sampling procedures, and any other design decisions that are relevant for interpreting and understanding the reported results. Exercising care not to overgeneralize from purposeful samples, while maximizing to the full the advantages of in-depth, purposeful sampling, will do much to alleviate concerns about small sample size.

METHODOLOGICAL MIXES

A study may employ more than one sampling strategy. It may also include multiple types of data. The chapters on interviewing, observation, and analysis will include information that will help in making design decisions. Before turning to those chapters, however, I want to briefly discuss the value of using multiple methods in research and evaluation.

Triangulation

One important way to strengthen a study design is through triangulation, or the combination of methodologies in the study of the same phenomena or programs. This can mean using several kinds of methods or data, including using both quantitative and qualitative approaches. Denzin (1978b) has identified four basic types of triangulation: (1) data triangulation—the use of a variety of data sources in a study; (2) investigator triangulation—the use of several different researchers or evaluators; (3) theory triangulation—the use of multiple perspectives to interpret a single set of data; and (4) methodological triangulation—the use of multiple methods to study a single problem or program.

The term triangulation is taken from land surveying. Knowing a single landmark only locates you somewhere along a line in a direction from the landmark, whereas with two landmarks you can take bearings in two directions and locate yourself at their intersection (Fielding and Fielding, 1986: 23). The term triangulation also works metaphorically to call to mind the world’s strongest geometric shape—the triangle (e.g., the form used to construct geodesic domes à la Buckminster Fuller). The logic of triangulation is based on the premise that no single method ever adequately solves the problem of rival causal factors. . . . Because each method reveals different aspects of empirical reality, multiple methods of observations must be employed. This is termed triangulation. I now offer as a final methodological rule the principle that multiple methods should be used in every investigation. (Denzin, 1978b: 26)

Triangulation is ideal. It can also be very expensive. An evaluation’s limited budget, short time frame, and political constraints will affect the amount of triangulation that is practical. Certainly, one important strategy for conducting evaluation research is to employ multiple methods, measures, researchers, and perspectives—but to do so reasonably and practically. Combinations of interviewing, ob-
mannerisms that will permit them to move easily among the people, sometimes secretly, sometimes openly, but always with the purpose of better understanding what the world of programs is really like. They are then able to report those understandings to our modern-day version of kings so that political wisdom can be enhanced and programmatic decisions enlightened.

That, at least, is one notion of how to think about evaluation through observation.

Chapter 7

Qualitative Interviewing

After much study of the evaluation masters, three youths came before Halcolm to ask how they might further increase their knowledge and wisdom. Halcolm sensed that they lacked experience in the real world, but he wanted to have them make the transition from the seclusion of their studies to the outside world in stages.

During the first stage he sent them forth under a six-month vow of silence. During those six months they wore the identifying garments of the muted truth-seekers so that people would know they were forbidden to speak. Each day, according to their instructions, they sat at the market in whatever village they entered, watching but never speaking. After six months in this fashion they returned to Halcolm.

"So," Halcolm began, "you have returned to us from your journey. Your period of silence is over. Your transition to the world beyond our walls of study has begun. What have you learned on this your first journey?"

The first youth answered, "In every village the patterns are the same. People come to the market. They buy the goods they need, talk with friends, and leave. I have learned that all markets are alike and the people in markets always the same. I have learned that all things are ultimately the same from place to place."

Then the second youth reported, "I too watched the people come and go in the markets. I have learned that all life is coming and going, people forever moving to and fro in search of food and basic material things. I understand now the simplicity of human life."

Halcolm looked at the third youth: "And what do you have to tell us?"

"I saw the same markets and the same people as my fellow-travelers, yet I know not what they know. My mind is filled with questions. I kept wondering where the people came from and where they went. I pondered what they might be thinking and feeling as they came and
went. I reflected on how they happened to be at this market on this day, who they left behind, and who came with them. I wondered how today was the same or different for them. I have failed, Master, for I am filled with questions rather than answers, questions for the people I saw. I do not know what I have learned."

Halcolm smiled. "You have learned most of all. You have learned the value of being able to ask questions. You have learned the importance of finding out what people have to say. You are ready now to return to the world, this time without the vow of silence.

"Go forth now. Go forth and question. Ask and listen. The world is just beginning to open up to you. Each person you question can take you into a new part of the world. For the person who is willing to ask and listen the world will always be new. The skilled questioner and attentive listener knows how to enter into another's experience."

—From Halcolm's Epistemological Parables

Inner Perspectives

The purpose of interviewing is to find out what is in and on someone else's mind. The purpose of open-ended interviewing is not to put things in someone's mind (for example, the interviewer's preconceived categories for organizing the world) but to access the perspective of the person being interviewed. We interview people to find out from them those things we cannot directly observe. The issue is not whether observational data is more desirable, valid, or meaningful than self-report data. The fact of the matter is that we cannot observe everything. We cannot observe feelings, thoughts, and intentions. We cannot observe behaviors that took place at some previous point in time. We cannot observe situations that preclude the presence of an observer. We cannot observe how people have organized the world and the meanings they attach to what goes on in the world. We have to ask people questions about those things. The purpose of interviewing, then, is to allow us to enter into the other person's perspective. Qualitative interviewing begins with the assumption that the perspective of others is meaningful, knowable, and able to be made explicit.

Interview data for program evaluation purposes allow the evaluator to capture the perspectives of program participants, staff, and others associated with the program. What does the program look and feel like to the people involved? What are the experiences of program participants? What thoughts do people knowledgeable about the program have concerning program operations, processes, and outcomes? What do people know about the program? What are their expectations? What features of the program are most salient to the people involved? What changes do participants perceive in themselves as a result of their involvement in the program? It is the responsibility of the evaluator to provide a framework within which people can respond comfortably, accurately, and honestly to these kinds of questions.

The task for the interviewer is to make it possible for the person being interviewed to bring the interviewer into his or her world. The quality of the information obtained during an interview is largely dependent on the interviewer. The purpose of this chapter is to discuss ways of obtaining high-quality evaluative information by talking with people who have that information.

Evaluators can enhance the utilization potential of the information they collect by making sure they take the necessary steps to increase the quality of their findings. As Hermann Sudermann said in Es Lebe das Leben I, "I know how to listen when clever men are talking. That is the secret of what you call my influence." Evaluators must learn how to listen when knowledgeable people are talking. That may be the secret of their influence.

This chapter begins by discussing three different types of interviews. Later sections consider the content of interviews: what to ask questions about and ways of phrasing interview questions. The chapter ends with a discussion of how to record the responses obtained during interviews. This chapter will emphasize skill and technique as ways of enhancing the quality of interview data, but no less important than skill and technique is a genuine interest in and caring about the perspectives of other people. If what people have to say about their world is generally boring to you, then you will never be a great interviewer. Unless you are fascinated by the rich variation in human experience, qualitative interviewing will become drudgery. On the other hand, a deep and genuine interest in learning about people will lead nowhere without disciplined and rigorous inquiry based on skill and technique.
Variations in Qualitative Interviewing

There are three basic approaches to collecting qualitative data through open-ended interviews. The three approaches involve different types of preparation, conceptualization, and instrumentation. Each approach has strengths and weaknesses, and each serves a somewhat different purpose. The three choices are these:

(1) the informal conversational interview,
(2) the general interview guide approach, and
(3) the standardized open-ended interview.

These three approaches to the design of the interview differ in the extent to which interview questions are determined and standardized before the interview occurs.

The informal conversational interview relies entirely on the spontaneous generation of questions in the natural flow of an interaction, typically an interview that occurs as part of ongoing participant observation fieldwork. During an informal conversational interview, the persons being talked with may not even realize they are being interviewed.

The general interview guide approach involves outlining a set of issues that are to be explored with each respondent before interviewing begins. The issues in the outline need not be taken in any particular order and the actual wording of questions to elicit responses about those issues is not determined in advance. The interview guide simply serves as a basic checklist during the interview to make sure that all relevant topics are covered. The interview guide presumes that there is common information that should be obtained from each person interviewed, but no set of standardized questions are written in advance. The interviewer is thus required to adapt both the wording and the sequence of questions to specific respondents in the context of the actual interview.

The standardized open-ended interview consists of a set of questions carefully worded and arranged with the intention of taking each respondent through the same sequence and asking each respondent the same questions with essentially the same words. Flexibility in probing is more or less limited, depending on the nature of the interview and the skills of interviewers. The standardized open-ended interview is used when it is important to minimize variation in the questions posed to interviewees. This reduces the possibility of bias that comes from having different interviews for different people, including the problem of obtaining more comprehensive data from certain persons while getting less systematic information from others. A standardized open-ended interview may be particularly appropriate when a large number of people are to conduct interviews on the same topic and the evaluator wishes to reduce the variation in responses due to the fact that, left to themselves, different interviewers will ask questions on a single topic in different ways. By controlling and standardizing the open-ended interview, the evaluator obtains data that are systematic and thorough for each respondent but the process reduces flexibility and spontaneity.

The Informal Conversational Interview

The informal conversational interview is the most open-ended approach to interviewing. The conversational interviewer wants to maintain maximum flexibility to be able to pursue information in whatever direction appears to be appropriate, depending on what emerges from observing a particular setting or from talking to one or more individuals in that setting. Most of the questions will flow from the immediate context. Thus the conversational interview is a major tool used in combination with participant observation to permit the evaluator who is participating in some programmatic activity to understand other participants’ reactions to what is happening. No predetermined set of questions is possible under such circumstances, because the evaluator does not know beforehand what is going to happen or what will be important to ask.

The data gathered from informal conversational interviews will be different for each person interviewed. In many cases, the same person may be interviewed on a number of different occasions using an informal, conversational approach. This approach is particularly useful where the researcher can stay in the setting for some period of time, so that he or she is not dependent on a single interview with a respondent. Interview questions will change over time, and each new interview builds on those already done, expanding information that
was picked up previously, moving in new directions, and seeking elucidations and elaborations from various participants.

The conversational interviewer must "go with the flow." Depending on how the interviewer's role has been defined, the people being interviewed may not know during any particular informal conversation that the purpose of the conversation is the collection of data. This means that in many cases participant observers do note take notes during the interview; rather, they write down what they learned after they have left the interview/observation situation. In other cases, it can be both appropriate and comfortable to take notes or even use a tape recorder.

The strength of the informal conversational approach is that it allows the interviewer/evaluator to be highly responsive to individual differences and situational changes. Questions can be individualized to establish in-depth communication with the person being interviewed and to make use of the immediate surroundings and situation to increase the concreteness and immediacy of the interview questions and responses.

The weakness of the informal conversational interview is that it requires a greater amount of time to collect systematic information because it may take several conversations with different people before a similar set of questions has been posed to each participant in the program. The informal conversational interview is also more open to interviewer effects in that it depends on the conversational skills of the interviewer/evaluator to a greater extent than do more formal, standardized formats. The conversational interviewer must be able to interact easily with people in a variety of settings, generate rapid insights, formulate questions quickly and smoothly, and guard against asking questions that impose interpretations on the situation by the structure of the questions.

Data obtained from informal conversational interviews are also difficult to pull together and analyze. Because different questions will generate different responses, the researcher has to spend a great deal of time sifting through responses to find patterns that have emerged at different points in different interviews with different people. By contrast, interviews that are more systematized and standardized facilitate analysis but provide less flexibility and are less sensitive to individual and situational differences.

Qualitative Interviewing

The Interview Guide

An interview guide is a list of questions or issues that are to be explored in the course of an interview. An interview guide is prepared in order to make sure that basically the same information is obtained from a number of people by covering the same material. The interview guide provides topics or subject areas within which the interviewer is free to explore, probe, and ask questions that will elucidate and illuminate that particular subject. Thus the interviewer remains free to build a conversation within a particular subject area, to word questions spontaneously, and to establish a conversational style—but with the focus on a particular subject that has been predetermined.

The advantage of an interview guide is that it makes sure that the interviewer/evaluator has carefully decided how best to use the limited time available in an interview situation. The interview guide helps make interviewing across a number of different people more systematic and comprehensive by delimiting in advance the issues to be explored. The interview guide approach is especially useful in conducting group interviews: It keeps the interactions focused but allows individual perspectives and experiences to emerge.

Interview guides can be developed in more or less detail, depending on the extent to which the researcher is able to specify important issues in advance and the extent to which it is felt that a particular sequence of questions is important to ask in the same way or in the same order for all respondents. Lofland (1971) provides a number of examples of interview guides that have been used in the conduct of sociological research. What follows is an example of an interview guide used with participants in a manpower training program.

Interview Guide for Manpower Program Evaluation

What has the trainee done in the program: activities? interactions? products? work performed?

What are the trainee's current work skills? What things can the trainee do that are marketable?

How has the trainee been affected by the program in areas other than job skills—feelings about self? attitudes toward work? aspirations? interpersonal skills? spin-offs?

What are the trainee's plans for the future—work plans? income expectations? life-style expectations/plans?
What does the trainee think of the program—strengths? weaknesses? things liked? things disliked? best components? poor components? things that should be changed?

This interview guide provides a framework within which the interviewer would develop questions, sequence those questions, and make decisions about which information to pursue in greater depth. The interviewer normally would not be expected to go into totally new subjects that are not covered within the framework of the interview guide. The interviewer does not ask questions, for example, about previous employment or education, how the person got into the program, how this program compares with other programs the trainee has experienced, and the trainee’s health. Other topics might still emerge during the interview, topics of importance to the respondent that are not listed explicitly on the guide and, therefore, would not normally be explored with each person interviewed. For example, respondents might comment on their reactions to staff, reactions to written materials, and reactions to specific program components. Comments on these concerns might emerge when, in accordance with the interview guide, the trainee is asked for reactions to program strengths, weaknesses, and so on, but if staff are not mentioned by the respondent, the interviewer would not raise that issue.

An additional, more detailed example of an interview guide approach is included as Appendix 7.1 at the end of this chapter (pp. 360-362). The appendix illustrates how it is possible to use a detailed outline guide to conduct a series of interviews with the same respondents during the course of a year. The guide in the appendix is the outline for a “descriptive interview” developed by the Educational Testing Service Collaborative Research Project on Reading.

The flexibility permitted by the interview guide approach will become clearer after reviewing the third strategy of qualitative interviewing in the next section.

The Standardized Open-Ended Interview

In many cases, when conducting a program evaluation, it is only possible to interview participants for a limited period of time. Sometimes it is only possible to interview each participant once. At other times it is possible and desirable to interview participants before they enter the program, when they leave the program, and again after some period of time (for example, six months) after they have left the program. Because of limited time, when it is desirable to have the same information from each person interviewed, a standardized open-ended format may be used in which each person is asked essentially the same questions. The interview questions are written out in advance exactly the way they are to be asked during the interview. Careful consideration is given to the wording of each question before the interview. Any clarifications or elaborations that are to be used are written into the interview itself. Probing questions are placed in the interview at appropriate places.

The basic purpose of the standardized open-ended interview is to minimize interviewer effects by asking the same question of each respondent. Moreover, the interview is systematic and the necessity for interviewer judgment during the interview is reduced. The standardized open-ended interview also makes data analysis easier because it is possible to locate each respondent’s answer to the same question rather quickly and to organize questions and answers that are similar.

There are three major reasons for using standardized open-ended interviews as part of an evaluation:

1. The exact instrument used in the evaluation is available for inspection by decision makers and information users;
2. variation among interviewers can be minimized where a number of different interviewers must be used; and
3. the interview is highly focused so that interviewee time is carefully used.

In many cases it is sufficient to make available a topical interview guide for evaluation users to inspect. However, the problems of legitimacy and credibility for qualitative data can make it politically wise to produce an exact interview form that one can show to decision makers and evaluation users, telling them with certainty that these are the exact questions that will be asked of clients or others who are interviewed. By generating a standardized form, evaluation users can participate more completely in writing the interview instrument. They will then know exactly what is going to be asked and what is not going to be asked. This reduces the likelihood of the data being attacked later because certain questions were missed or asked in the wrong way. By making it clear, in advance of data collection, exactly
what questions will be asked, the limitations of the data can be known and discussed beforehand.

In contrast with a standardized approach, asking different questions of different clients may reduce credibility. While a conversational approach, and even the interview guide approach, have the strengths of permitting greater flexibility and individualization, these approaches also open up the possibility that more information will be collected from some people than from others. When analyzing the data it becomes difficult to be certain how the findings are influenced by these qualitative differences in the depth and breadth of information received from different people.

For the conduct of basic research, when one is attempting to understand the holistic worldview of a group of people, it is not necessary to collect the same information from each person. The political credibility of the data collected is less an issue under basic research conditions. However, when using qualitative data-collection procedures for evaluation purposes, it can be helpful to minimize issues of legitimacy and credibility by carefully collecting the same information from everyone who is interviewed.

The standardized open-ended interview also reduces variation among interviewers. Some evaluations rely on volunteers to do interviewing; at other times program staff may be involved in doing some interviewing; and in still other instances interviewers may be novices, students, or others who are not professional social scientists or evaluators. When a number of different interviewers are used, variations in data created by differences among interviewers will become particularly apparent if an informal conversational approach to data gathering is used or even if each interviewer uses a basic guide. The best way to guard against variations among interviewers is to carefully word questions in advance and train the interviewers not to deviate from the prescribed forms. The data collected are still open-ended, in the sense that the respondent supplies his or her own words, thoughts, and insights in answering the questions, but the precise wording of the questions is determined ahead of time.

The weakness of the standardized approach is that it does not permit the interviewer to pursue topics or issues that were not anticipated when the interview was written. Constraints are also placed on the use of different lines of questioning with different people based on their unique experiences. Therefore, a standardized open-ended interview approach will reduce the extent to which individual differences and circumstances can be taken into account; on the other hand, this approach can reduce individual interviewer effects and facilitate data analysis.

Just as it is possible to some extent to combine a conversational approach with an interview guide approach, it is also possible to combine an interview guide approach with a standardized open-ended approach. Thus a number of basic questions may be worded precisely in a predetermined fashion, while permitting the interviewer more flexibility in probing and more decision-making flexibility in determining when it is appropriate to explore certain subjects in greater depth or even to undertake whole new areas of inquiry that were not originally included in the interview instrument. It is even possible to adopt a standardized open-ended interview format in the early part of an interview and then leave the interviewer free to pursue any subjects of interest during the latter parts of the interview. Another combination would include using the informal conversational interview early in an evaluation project, followed midway through by an interview guide, and then closing the program evaluation with a standardized open-ended interview to get systematic information from a sample of participants at the end of the program or when conducting follow-up studies of participants.

To illustrate the standardized open-ended interview, three interviews have been reproduced in Appendix 7.2 at the end of this chapter (pp. 363-368). These interviews were used to gather information from participants in an Outward Bound wilderness program for disabled persons. The first interview was conducted at the beginning of the program; the second interview was used at the end of the ten-day experience; and the third interview took place six months after the program.

**Summary of Interviewing Strategies**

The common characteristic of all three qualitative approaches to interviewing is that the persons being interviewed respond in their own words to express their own personal perspectives. While there are variations in strategy concerning the extent to which the wording and sequencing of questions ought to be predetermined, there is no variation in the principle that the response format should be open-
<table>
<thead>
<tr>
<th>Type of Interview</th>
<th>Characteristics</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Informal conversational</td>
<td>Questions emerge from the immediate context and are asked in the natural course</td>
<td>Increases the salience and relevance of questions; interviews are built on and emerge from observations; the interview can be matched to</td>
<td>Different information collected from different people with different questions. Less systematic and comprehensive if certain</td>
</tr>
<tr>
<td>interview</td>
<td>of things; there is no predetermination of question topics or wording.</td>
<td>individuals and circumstances.</td>
<td>questions do not arise &quot;naturally.&quot; Data organization and analysis can be quite difficult.</td>
</tr>
<tr>
<td>(2) Interview guide approach</td>
<td>Topics and issues to be covered are specified in advance, in outline form;</td>
<td>The outline increases the comprehensiveness of the data and makes data collection somewhat systematic for each respondent. Logical gaps in</td>
<td>Important and salient topics may be inadvertently omitted. Interviewer flexibility in sequencing and wording questions can result in</td>
</tr>
<tr>
<td></td>
<td>interviewer decides sequence and wording of questions in the course of the</td>
<td>data can be anticipated and closed. Interviews remain fairly conversational and situational.</td>
<td>substantially different responses from different perspectives, thus reducing the comparability of responses.</td>
</tr>
<tr>
<td></td>
<td>interview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Standardized open-ended</td>
<td>The exact wording and sequence of questions are determined in advance. All</td>
<td>Respondents answer the same questions, thus increasing comparability of responses; data are complete for each person on the topics</td>
<td>Little flexibility in relating the interview to particular individuals and circumstances; standardized wording of questions may</td>
</tr>
<tr>
<td>interview</td>
<td>interviewees are asked the same basic questions in the same order. Questions</td>
<td>addressed in the interview. Reduces interviewer effects and bias when several interviewers are used. Permits evaluation users to see and</td>
<td>constrain and limit naturalness and relevance of questions and answers.</td>
</tr>
<tr>
<td></td>
<td>are worded in a completely open-ended format.</td>
<td>review the instrumentation used in the evaluation. Facilitates organization and analysis of the data.</td>
<td></td>
</tr>
<tr>
<td>(4) Closed, fixed response</td>
<td>Questions and response categories are determined in advance. Responses are</td>
<td>Data analysis is simple; responses can be directly compared and easily aggregated; many questions can be asked in a short time.</td>
<td>Respondents must fit their experiences and feelings into the researcher's categories; may be perceived as impersonal, irrelevant,</td>
</tr>
<tr>
<td>interview</td>
<td>fixed; respondent chooses from among these fixed responses.</td>
<td></td>
<td>and mechanistic. Can distort what respondents really mean or experienced by so completely limiting their response choices.</td>
</tr>
</tbody>
</table>


ended. The interviewer never supplies and predetermines the phrases or categories that must be used by respondents to express themselves. The purpose of qualitative interviewing in evaluation is to understand how program staff and participants view the program, to learn their terminology and judgments, and to capture the complexities of their individual perceptions and experiences. This is what distinguishes qualitative interviewing from the closed interview, questionnaire, or test typically used in quantitative evaluations. Such closed instruments force program participants to fit their knowledge, experiences, and feelings into the evaluator’s categories. The fundamental principle of qualitative interviewing is to provide a framework within which respondents can express their own understandings in their own terms.

Table 7.1 summarizes variations in interview instrumentation. In reviewing this summary table, it is important to keep in mind that these are presented as pure types. In practice, any particular evaluation may employ several of these strategies together.

THE CONTENT OF INTERVIEWS: WHAT QUESTIONS TO ASK

A number of decisions must be made in planning an interview, whether the interview takes place spontaneously in the field or is carefully prepared as a standardized open-ended instrument. The evaluator must decide what questions to ask, how to sequence questions, how much detail to solicit, how long to make the interview, and how to word the actual questions. These are all measurement questions that will affect the quality of interview responses. There are basically six kinds of questions that can be asked of people. On any given topic it is possible to ask any of these questions.

Experience/Behavior Questions

These are questions about what a person does or has done. These questions are aimed at eliciting descriptions of experiences, behaviors, actions, and activities that would have been observable had the observer been present. “If I had been in the program with you, what would I have seen you doing?” “If I followed you through a typical day, what would I see you doing? What experiences would I observe you having?”

Opinion/Values Questions

These are questions aimed at understanding the cognitive and interpretive processes of people. Answers to these questions tell us what people think about some issue. They tell us about people’s goals, intentions, desires, and values. These questions typically carry an implication of respondent rationality and decision making. “What do you believe?” “What do you think about ______?” “What would you like to see happen?” “What is your opinion of ______?”

Feeling Questions

These are questions aimed at understanding the emotional responses of people to their experiences and thoughts. There is an implicit assumption of spontaneity about the origin of emotional responses. Feelings occur inside people; they are their natural, emotional responses to what happens around them or to them. Feelings tap the affective dimension of human life. In asking feeling questions, the interviewer is looking for adjective responses, for example, “To what extent do you feel anxious, happy, afraid, intimidated, confident, . . . ?”

Opinions and feelings are often confused. It is critical that interviewers understand the distinction between the two in order to know when they have the kind of answer they want to the question they are asking. Suppose an interviewer asks: “How do you feel about that?” The response: “I think it’s probably the best that we can do under the circumstances.” The question about feelings has not really been answered. Analytical, interpretive, and opinion statements are not answers to questions about feelings.

This confusion sometimes occurs because interviewers give the wrong cues when asking questions—for example, by asking opinion questions using the format “How do you feel about that?” instead of “What is your opinion about that?” or “What do you think about it?” When one wants to understand the respondents’ emotional reactions it is appropriate to ask about feelings. When one wants to understand what they think about something, the question should explicitly ask about opinions, beliefs, and considered judgments—not about feelings.
Knowledge questions are asked to find out what factual information the respondent has. The assumption here is that certain things are considered to be known—these things are not opinions and they are not feelings; rather, they are the things that one knows, the facts of the case. Knowledge about a program may consist of reporting on what services are available, who is eligible, the characteristics of clients, who the program serves, how long people spend in the program, what the rules and regulations of the program are, how one enrolls in the program, and so on. While from a philosophical point of view it is possible to argue that all knowledge is merely a set of beliefs rather than facts, the issue here is to find out what the person being interviewed considers to be factual. It is the respondent’s understanding of “factual” knowledge that is being elicited.

Sensory Questions

These are questions about what is seen, heard, touched, tasted, and smelled. The purpose of these questions is to allow the interviewer to enter into the sensory apparatus of the respondent. “When you walk through the doors of the program, what do you see? Describe to me what I would see if I walked through the doors into the program.” Or again: “What does the counselor ask you when you meet with him? What does he actually say?” Sensory questions attempt to have interviewees describe the stimuli to which they are subject. Technically, sensory data are a type of behavioral or experiential data—they capture the experience of the senses. However, the types of questions asked to gather sensory data are sufficiently distinct to merit a separate category.

Background/Demographic Questions

These questions concern the identifying characteristics of the person being interviewed. Answers to these questions help the interviewer locate the respondent in relation to other people. Questions concerning age, education, occupation, residence/mobility, and the like are standard background questions. They are distinguishable from knowledge questions primarily because of their routine nature.

<table>
<thead>
<tr>
<th>Qualitative Interviewing</th>
<th>Past</th>
<th>Present</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior/Experience Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opinion/Value Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeling Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic/Background Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7.1. A Matrix of Question Options

Questions concerning behaviors, opinions, feelings, knowledge, sensory data, and demographics—these are all the kinds of questions that it is possible to ask in an interview. Any kind of question one might want to ask can be subsumed in one of these categories. Keeping these types of questions in mind can be particularly helpful when it comes to planning the comprehensiveness of the interview and ordering the questions in some sequence. Before considering the sequence of questions, however, it is important to consider how the time dimension intersects with the different kinds of questions.

The Time Frame of Questions

Any of the questions described above can be asked in the present tense, past tense, or future tense. For example, it is possible to ask a person what they are doing now, what they have done in the past, and what they plan to do in the future. Likewise, one might be interested in present attitudes, past attitudes, or future attitudes. By combining the time frame of questions with the different types of questions it is possible to construct a matrix that generates 18 different types of questions. Figure 7.1 shows that matrix.

Asking all 18 questions about any particular situation, event, or programmatic activity may become somewhat tedious, especially if the sequence is repeated over and over throughout the interview for different program elements. The matrix constitutes a set of options from which one can select which pieces of information are most important to obtain. In order to understand how these options are
applied in an actual interview situation it may be helpful to review an actual interview. The Outward Bound standardized interview in Appendix 7.2 at the end of this chapter can be used for this purpose (pp. 363-368). Try identifying which cell in the matrix (Figure 7.1) is represented by each question in the Outward Bound interviews.

The Sequencing of Questions

There are no fixed rules of sequence in organizing an interview. Informal conversational interviewing is flexible and responsive so that a fixed sequence is seldom possible. However, standardized open-ended interviews must establish a fixed sequence of questions due to their structured format. I offer, then, some suggestions about sequencing.

I prefer to begin the interview with questions about noncontroversial present behaviors, activities, and experiences. Such questions ask for relatively straightforward descriptions, they require minimal recall and interpretation. Such questions are, therefore, fairly easy to answer. They encourage the respondent to talk descriptively. Probes should focus on eliciting greater detail—filling out the descriptive picture.

Once some experience or activity has been described, it is appropriate to ask about interpretations, opinions, and feelings. Opinions and feelings are likely to be more accurate and meaningful once the respondent has just verbally relived the experience. Thus a context is established for expressing feelings and opinions, that is, grounding feelings and opinions in relation to experiences.

Knowledge and skill questions also typically need a context. These questions can be quite threatening. It is helpful to ask them in conjunction with specific questions about program activities and experiences that have a bearing on knowledge and skills. Finding out from people what they know and what skills they possess works best once some rapport and trust have been established in the interview. Relating knowledge and skills to descriptions of program activity can help provide a concrete context for these kinds of questions.

Questions about the present tend to be easier for respondents than questions about the past. Future-oriented questions involve considerable speculation, and responses to questions about future actions or attitudes are typically less reliable than questions about the present or past. I generally prefer to begin by asking questions about the present, then, using the present as a baseline, ask questions about the same activity or attitude in the past. Only then will I broach questions about the future.

Background and demographic questions are basically boring; they epitomize what people don't like about interviews. They can also be somewhat uncomfortable for the respondent, depending on how personal they are. I keep such questions to an absolute minimum and prefer to space them strategically and unobtrusively throughout the interview. I advise never beginning an interview with a long list of routine demographic questions. In qualitative interviewing, the interviewee needs to become actively involved in providing descriptive information as soon as possible instead of becoming conditioned to providing short-answer, routine responses to uninteresting categorical questions. Some background information may be necessary at the beginning to make sense out of the rest of the interview, but such questions should be tied to descriptive information about present program experience as much as possible. Otherwise, save the sociodemographic inquiries (age, socioeconomic status, birth order, and the like) for the end.

THE WORDING OF QUESTIONS

An interview question is a stimulus that is aimed at creating or generating a response from the person being interviewed. The way a question is worded is one of the most important elements determining how the interviewee will respond. As Payne (1951) observed in his classic book on questioning, asking questions is an art. For purposes of qualitative inquiry, good questions should, at a minimum, be open-ended, neutral, singular, and clear. Each of these criteria will be discussed in some detail.

Asking Truly Open-Ended Questions

The basic thrust of qualitative interviewing is to minimize the imposition of predetermined responses when gathering data. When using qualitative interviewing strategies for data collection it is critical that questions be asked in a truly open-ended fashion. This means that the questions should permit respondents to respond in their own terms.
The standard questionnaire item in quantitative measurement provides the respondent with a categorical list of response possibilities: "How do you feel about the program? Would you say that you are (a) very satisfied, (b) somewhat satisfied, (c) not too satisfied, (d) not at all satisfied." It is clear in this instance that the question is closed and that the respondent has been provided with a limited and predetermined set of alternatives. The response possibilities are clearly stated and made explicit in the way in which the question is asked. Many interviewers think that the way to make a question open-ended is simply to leave out the structured response categories. Such an approach does not, however, make a question truly open-ended. It merely makes the predetermined response categories implicit and disguised.

Consider the following "open-ended" question: "How satisfied are you with this program?" On the surface this appears to be an open-ended question. On close inspection, however, it is clear that the dimension along which the respondent can answer the question has already been identified. The respondent is being asked for some degree of satisfaction. It is true that the interviewee can use a variety of modifiers for the word satisfaction—for example, "pretty satisfied," "kind of satisfied," "mostly satisfied," and so on. But, in effect, the possible response set has been narrowly limited by the wording of the question. The desired dimension of response is identified in the wording of the question such that the typical answers are only slightly different from those that would have been obtained had the categories been made explicit from the start.

The truly open-ended question does not presuppose which dimension of feeling or thought will be salient for the interviewee. The truly open-ended question allows the person being interviewed to select from among that person's full repertoire of possible responses. Indeed, in qualitative inquiry one of the things the evaluator is trying to determine is what dimensions, themes, and images/words people use among themselves to describe their feelings, thoughts, and experiences. Examples, then, of truly open-ended questions would take the following format:

How do you feel about the program?
What is your opinion of the program?
What do you think of the program?

The truly open-ended question permits persons being interviewed to take whatever direction and use whatever words they want in order to represent what they have to say. Moreover, to be truly open-ended a question cannot be phrased as a dichotomy. The next section discusses the problem of dichotomous questions in interviews.

The Horns of a Dichotomy

Dichotomous response questions provide the interviewee with a grammatical structure suggesting a "yes" or "no" answer. Are you satisfied with the program? Have you changed as a result of your participation in this program? Was this an important experience for you? Do you know the procedures for enrolling in the program? Have you interacted much with the staff in the program?

The object of an in-depth interview is to get the person being interviewed to talk—to talk about experiences, feelings, opinions, and knowledge. Far from encouraging the respondent to talk, dichotomous response questions create a dilemma for the respondent because they frequently are not sure whether they are being asked a simple yes-no question or, indeed, the interviewer expects a more elaborate response. I have found in many cases that interviewers who report that they have difficulty getting respondents to talk are using a string of dichotomous response questions to guide the interview and thereby have programmed the respondent to be entirely reactive in a binary way, allowing the interviewer to supply the content to the interview.

Perhaps the classic example of a series of dichotomous questions is a conversation between a parent and a teenager.

(Teenager returns home from a date.)
Do you know that you're late?
Yeah.
Did you have a good time?
Yeah.
Did you go to a movie?
Yeah.
Was it a good movie?
Yeah, it was okay.
So, it was worth seeing?
Yeah, it was worth seeing.
I've heard a lot about it. Do you think I would like it?
I don't know. Maybe.
Anything else you'd like to tell me about your evening?
No, I guess that's it.
(Teenager goes upstairs to bed. One parent turns to the other and says:
"It sure is hard to get him to talk to us, I guess he's at the age where kids
just don't want to tell their parents anything.")

Dichotomous questions can turn an interview into an interrogation
or a quiz rather than an in-depth conversation. In everyday convers-
ation our interactions with each other are filled with dichotomous
questions that we unconsciously ignore and treat as if they were
open-ended questions. In a more formal interview setting, however,
the interviewee will become more conscious of the grammatical struc-
ture of questions and will be less likely to ignore questions that pose
dichotomous alternatives. Indeed, the more intense and concentrated
the interview situation, the more likely the respondent is to pay close
attention to the structure of questions and to take questions literally.

In training interviewers, I like to play a game where I will only
respond literally to the questions asked without volunteering any
information that is not clearly demanded in the question. I do this
before explaining the difficulties involved in asking dichotomous
questions. I have played this game hundreds of times, and the reaction
is typically the same. When getting dichotomous responses to general
questions, the interviewer will begin to rely on more and more specific
dichotomous response questions, thereby digging a deeper and deeper
hole, which makes it difficult to pull the interview out of the
dichotomous response pattern. Transcribed below is an actual inter-
view from a training workshop. In the left column I have recorded the
interview that took place; the right column records truly open-ended
alternatives to the dichotomous questions that were asked.

INTERVIEW DEMONSTRATION

Instruction: Okay, now we're going to play an interviewing game.
I want you to take turns asking me questions about an evaluation I
just completed. The program being evaluated was a staff develop-
ment demonstration project that involved taking professionals into a
wilderness setting for a week. That's all I'm going to tell you at this

Qualitative interviewing

point. I'll answer your questions as precisely as I can, but I'll only
answer what you ask. I won't volunteer any information that isn't
directly asked for by your questions.

<table>
<thead>
<tr>
<th>Actual interview</th>
<th>What the interviewer really wanted to know: open-ended question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question</strong>: Were you the evaluator of this program?</td>
<td><strong>What was your role in this program?</strong></td>
</tr>
<tr>
<td><strong>Answer</strong>: Yes.</td>
<td></td>
</tr>
<tr>
<td><strong>Q</strong>: Were you doing a formative evaluation?</td>
<td><strong>What was the purpose of the evaluation?</strong></td>
</tr>
<tr>
<td><strong>A</strong>: Mostly.</td>
<td></td>
</tr>
<tr>
<td><strong>Q</strong>: Were you trying to find out if the people changed from being in the wilderness?</td>
<td><strong>What were you trying to find out in doing the evaluation?</strong></td>
</tr>
<tr>
<td><strong>A</strong>: That was part of it.</td>
<td></td>
</tr>
<tr>
<td><strong>Q</strong>: Did they change?</td>
<td><strong>How did participation in the program affect participants?</strong></td>
</tr>
<tr>
<td><strong>A</strong>: Some of them did.</td>
<td></td>
</tr>
<tr>
<td><strong>Q</strong>: Did you interview people both before and after the program?</td>
<td><strong>What kinds of information did you collect for the evaluation?</strong></td>
</tr>
<tr>
<td><strong>A</strong>: Yes.</td>
<td></td>
</tr>
<tr>
<td><strong>Q</strong>: Did you also go along as a participant in the program?</td>
<td><strong>How were you personally involved in the program?</strong></td>
</tr>
<tr>
<td><strong>A</strong>: Yes.</td>
<td></td>
</tr>
<tr>
<td><strong>Q</strong>: Did you find that being in the program affected what happened?</td>
<td><strong>How do you think your participation in the program affected what happened?</strong></td>
</tr>
<tr>
<td><strong>A</strong>: Yes.</td>
<td></td>
</tr>
<tr>
<td><strong>Q</strong>: Did you have a good time?</td>
<td><strong>What was the wilderness experience like for you?</strong></td>
</tr>
<tr>
<td><strong>A</strong>: Yes.</td>
<td></td>
</tr>
</tbody>
</table>
Q: Are you reluctant to tell us about the program?

I'd like to find out more about the program. What would be the best way for me to learn more from you about it?

This is clearly an extreme example of using dichotomous questions in an interview. It should be clear, however, that the truly open-ended questions would have generated quite different information than was being generated, and was likely to be generated, by the dichotomous questions. In addition, dichotomous questions can easily become leading questions. Once the interviewer begins to deal with what appears to be a reluctant or timid interviewee, by asking more and more detailed dichotomous response questions he or she can easily begin guessing at possible responses and actually impose those responses on the person being interviewed. One sure sign that this is happening is when the interviewer is doing more talking than the person being interviewed. Consider the following excerpt from an actual interview. The interviewee was a teenager who was participating in a chemical dependency program. The interview took place during the time the teenager was resident in the program.

Interview

Q: Hello, John. It's nice to see you again. I'm anxious to find out what's been happening with you. Can I ask you some questions about your experience?

A: Okay.

Q: I'd like you to think about some of the really important experiences you've had here. Can you think of something that stands out in your mind?

A: Yeah, ... the hot seat.

Q: The hot seat is when one person is the focus of attention for the whole group, right?

A: Right.

Comments

The opening is dominated by the interviewer. No informal give-and-take. The interviewee is set up to take a passive/reactive role.

Introductory cue sentence is immediately followed by a dichotomous response question.

John goes beyond the dichotomous response.

The interviewer has provided the definition, rather than getting John's own definition of the hot seat.

Dichotomous question.

Multiple questions. Unclear connections. Ambiguous, multiple-choice format at the end.

Dichotomous questions.

Leading question, setting up an easy acquiescence response.

Same as previous question.
A: It was pretty heavy.

John doesn't actually respond to the question. Ambiguous response.

Q: Okay, I want to ask you some about the lecture part of the program. Anything else you want to say about the hot seat?

Transition. John is cued that the hot seat questions are over. No response really expected.

(John doesn't answer verbally. Sits and waits for the next questions.)

The person conducting this interview said she wanted to find out two things in this portion of the interview: what experiences were most salient for John and how personally involved John was becoming in the program experience. She has learned that the "hot seat" was highly salient for John, but she really knows very little about the reasons for that salience. With regard to the second question of his personal involvement, the only data she has come from his acquiescence to leading questions. In fact, if one lists the data from the interview, there is very little there:

Okay.
Yeah, ... the hot seat.
Right.
One person does it every day.
Yeah, it depends.
Okay, let's see, hmmm ... there was this guy yesterday who really got nailed. I mean he really caught a lot of crap from the group. It was really heavy.
No, it was them others.
Yeah, right, and it really got to him.
He started crying and got mad and one guy really came down on him and afterwards they were talking, and it seemed to be okay for him.
Yeah, it really was.
It was pretty heavy.

In looking over the transcript of this portion of the interview, it is clear that the interviewer is talking more than the interviewee. The questions put the interviewee in a passive stance, able to confirm or deny the substance provided by the interviewer but not really given the opportunity to provide in-depth, descriptive detail.

Qualitative Interviewing

Presupposition Questions

Presuppositions are a major focus of study for many linguists (Karttunen, 1973; Bandler and Grinder, 1975a). Natural language is filled with presuppositions. In the course of communicating as we go about our day-to-day activities, it would be unusual to interact with other people without relying heavily on presuppositions. The dominance of presupposition structures in language has important implications for interviewing. By becoming aware of the effects of presupposition structures in interviewing situations, it is possible for the skillful interviewer to use presuppositions to increase the richness and depth of responses and data obtained.

What, then, are presuppositions? Linguists Grinder and Bandler define presuppositions as follows:

When each of us uses a natural language system to communicate, we assume that the listener can decode complex sound structures into meanings, i.e., the listener has the ability to derive the Deep-Structure meaning from the Surface-Structure we present to him auditorily. . . . [W]e also assume the complex skill of listeners to derive extra meaning from some Surface-Structures by the nature of their form. Even though neither the speaker nor the listener may be aware of this process, it goes on all the time. For example, if someone says:

I want to watch Kung Fu tonight on TV we must understand that Kung Fu is on TV tonight in order to process the sentence I want to watch . . .
to make any sense. These processes are called presuppositions of natural language. (Bandler and Grinder, 1975a: 241)

Presuppositions are particularly useful in interviewing because the interviewer presupposes that the respondent has something to say. Such a presupposition increases the likelihood that the person being interviewed will, indeed, have something to say. Consider the following question: "What is the most important experience you had in the program?" This question presupposes that the respondent has had an important experience. The person of whom the question is asked, of course, has the option of responding, "I haven't had any important experiences." However, it is more likely that the interviewee will go directly to the issue of which experience to report as important, rather than dealing first with the question of whether or not an important experience has occurred.
Contrast the presupposition format of the open-ended question to the format of the following dichotomous response question: "Have you had any experiences in the program so far that you would call really important?" This dichotomous question requires the person to make a decision about what an important experience is and whether or not an important experience has occurred. By raising the question at all, the interviewer focuses on the decision about whether or not something important has occurred rather than finding out what has occurred. The presupposition format, then, bypasses this initial step by asking directly for description rather than asking for an affirmation of the existence of the phenomenon in question. Listed below, on the left, are typical dichotomous response questions that are used to introduce a longer series of questions. On the right are presuppositions asked in a truly open-ended format that bypass the dichotomous response questions.

**Alternative Question Formats**

**Dichotomous lead-in question**

Do you feel like you know enough about the program to assess its effectiveness?  

Have you learned anything from this program?  

Do you do anything now in your work that you didn’t do before the program began?  

Is there any misuse of funds in this program?  

Are there any conflicts among the staff?  

**Presupposition lead-in question**

How effective do you think the program is? (Presupposes that a judgment can be made)  

What have you learned from this program? (Presupposes learning)  

What do you do now that you didn’t do before the program began? (Presupposes change)  

What kinds of misuse of funds have occurred in this program? (Presupposes at least some misuse of funds)  

What kinds of staff conflicts have occurred here? (Presupposes conflicts)  

There is often a naturalness about the use of presuppositions that makes more comfortable what might otherwise be embarrassing questions. The presupposition includes the implication that what is presupposed is the natural way things occur. It is natural for there to be conflict in programs; it is natural for there to be some misuse of funds in programs; and it is natural for people to have learned something from participation in a program. The presupposition provides a stimulus that asks the respondent to mentally access the answer to the question directly without making a decision about whether or not something has actually occurred.

I first learned about interview presuppositions from a friend who worked with the agency in New York City that had responsibility for interviewing carriers of venereal disease. The purpose of the interview was to find out about the carrier’s previous sexual contacts so that those persons could be informed that they might have venereal disease. He had learned from experience that there was all the difference in the world between asking a man, "Have you had any sexual relationships with other men?" and asking him, "How many sexual contacts with other men have you had?" The dichotomous question requires a decision about some admission of homosexuality. The presupposition form of the open-ended question suggests that some sexual contacts with other men might be quite natural and focuses on the frequency of occurrence rather than whether or not homosexual contacts have occurred at all. The venereal disease interviewers found that they were much more likely to generate responses with the presupposition format than with the dichotomous response format.

The real point here is that the purpose of in-depth interviews is to find out what someone has to say. By presupposing that the person being interviewed does, indeed, have something to say, the quality of the descriptions received may be enhanced. However, a note of warning: Presuppositions, like any single form of questioning, can be overused. Presuppositions are one option. There are many times when it is more comfortable and appropriate to check out the relevance of a question with a dichotomous inquiry ("Did you go to the lecture?") before asking further questions ("What did you think of the lecture?")

**Asking Singular Questions**

One of the basic rules of questionnaire writing is that each item must be singular—that is, no more than one idea should be contained in any given question. Consider this example:
An opportunity presents itself to talk with someone and the interview is under way. More structured and scheduled interviewing takes place as part of formal evaluation site visits. Staff and program participants know that the interview is to take place. Appointments are made and a specific place is set aside for the interview. Yet the best laid plans for scheduled interviews can go awry. The evaluator arrives at the appointed time and place only to find that the person to be interviewed is unwilling to cooperate or needs to run off to take care of some unexpected problem. When faced with such a situation it is helpful to have a single, one-shot question in mind to salvage at least something. This is the question you ask if you are only going to get one shot at the interviewee.

In evaluating an agricultural extension program I was interviewing farmers 150 miles north of the capital city. The farmers in the area were economically distressed and many felt alienated from politicians and professionals. I arrived at a farm for a scheduled interview, but the farmer refused to cooperate. At first he refused to even come out of the barn to call off the dogs surrounding my truck. Finally, he appeared and said,

I don’t want to talk to you tonight. I know I said I would, but we’ve got a family problem and I’m tired and upset. I’ve always helped with your government surveys. I fill out all the forms the government sends. But I’m tired of it. No more. I don’t want to talk.

I had driven a long way to get this interview. The fieldwork was tightly scheduled, and I knew that I would not get another shot at this farmer, even if he later had a change of heart. To try and salvage the situation, I asked my one-shot question, a question stimulated by his demeanor and overt hostility.

I’m sorry I caught you at a bad time. But as long as I’m here, let me ask you just one quick question. Is there anything you want to tell the bastards in St. Paul?

He hesitated for just a moment and then launched into a tirade that turned into a full, two-hour interview. I never got out of the truck, but I was able to cover the entire interview (although without ever referring to or taking out the written interview schedule). At the end of this conversational interview, which had fully satisfied my data-collection needs, he said, “Well, I’ve enjoyed talkin’ with you, and I’m sorry about refusin’ to fill out your form. I just don’t want to do a survey tonight.”

I told him I understood and thanked him for the conversation. My scheduled, structured interview had become an informal, conversational interview developed from a last ditch, one-shot question.

FOCUS GROUP INTERVIEWS

A focus group interview is an interview with a small group of people on a specific topic. Groups are typically six to eight people who participate in the interview for one-half to two hours.

Focus group interviewing was developed in recognition that many of the consumer decisions that people make are made in a social context, often growing out of discussions with other people. Thus market researchers began using focus groups in the 1950s as a way of simulating the consumer group process of decision making in order to gather more accurate information about consumer product preferences. The classic work on focus group interviews, *The Focused Interview*, was written by Robert K. Merton and associates in 1956.

The focus group interview is, indeed, an interview. It is not a discussion. It is not a problem-solving session. It is not a decision-making group. It is an interview.

The participants are typically a relatively homogeneous group of people who are asked to reflect on the questions asked by the interviewer. Participants get to hear each other’s responses and to make additional comments beyond their original responses as they hear what other people have to say. It is not necessary for the group to reach any kind of consensus. Nor is it necessary for people to disagree. The object is to get high-quality data in a social context where people can consider their own views in the context of the views of others.

Focus group interviews have several advantages when used for program evaluation purposes. It is a highly efficient qualitative data-collection technique. In one hour the evaluator can gather information from eight people instead of only one person. Thus the sample size can be increased significantly in an evaluation using qualitative methods through focus group interviewing. Focus group interviews also provide some quality controls on data collection in that participants
tend to provide checks and balances on each other that weed out false or extreme views. The group’s dynamics typically contribute to focusing on the most important topics and issues in the program, and it is fairly easy to assess the extent to which there is a relatively consistent, shared view of the program among participants. Finally, focus groups tend to be highly enjoyable to participants.

There are also some weaknesses of focus groups. Because the amount of response time to any given question is increased considerably by having a number of people respond, the number of questions that can be asked is limited. With eight people in an hour, it is typically possible to ask no more than ten major questions. Facilitating and conducting a focus group interview requires considerable group process skill. It is important to know how to manage the interview so that it is not dominated by one or two people, and so that those participants who tend not to be highly verbal are able to share their views.

It can be difficult to take notes during a focus group interview while also facilitating the discussion, so many groups are conducted by pairs of interviewers with one person focusing on taking notes and the other focusing on facilitation. Good notes help in sorting out what was said when the tape recording is transcribed later.

It is always possible that unexpected diversions will occur in a focus group, particularly in an evaluation setting where participants know each other. Conflicts may arise, power struggles may be played out, and status differences may become a factor. In market research, focus groups are typically conducted with people who do not know each other. Of course, when program participants in a focus group do know each other, it is not possible to guarantee confidentiality.

Focus group interviews can be used at any point in the evaluation process. Focus groups can be conducted as part of a needs assessment process with both potential client groups and professionals who know the needs of client groups. Focus groups can be conducted with client groups during a program to identify strengths, weaknesses, and needed improvements. Focus groups can be used at the end of a program, or even months after program completion, to gather perceptions about outcomes and impacts. Key community people can be interviewed in groups when their views of a program may be of interest for evaluation purposes. Focus group interviews can also be used with staff to identify key elements in a program’s implementation and treatment. In short, focus groups can be used for a full range of evaluation purposes.

Qualitative Interviewing

Focus groups are now widely used in market research with quite credible and useful results. This technique is only beginning to be used in evaluation. Focus group interviews, when conducted carefully and used appropriately, promise to provide a rich, new way of gathering qualitative evaluation information (AED, 1989; Krueger, 1988; Morgan, 1988; Higinbotham and Cox, 1979).

CROSS-CULTURAL INTERVIEWING

Evaluation has become an international activity. The 1989 meeting of the American Evaluation Association had as its theme: “International and Cross-Cultural Evaluation.” International development agencies have begun making extensive use of site visit teams to conduct evaluations. These teams typically conduct a great many interviews.

Intercultural interactions are always subject to misunderstandings. A great deal of attention has been devoted to developing materials and training approaches for cross-cultural sensitization (e.g., Brislin et al., 1986; Stewart, 1985; Casse and Deol, 1985; Harris and Moran, 1979).

International and cross-cultural short-term evaluation site visits are much more subject to misinterpretations and miscommunications than traditional, long-term anthropological fieldwork. There is only space here to mention a few of the potential problems by way of sensitizing evaluation researchers to the precariousness of cross-cultural interviewing.

Language Differences

The data from interviews are words. It is tricky enough to be sure what a person means when using a common language, but words can take on a very different meaning in other cultures. In Sweden I participated in an international conference discussing policy evaluations. The conference was conducted in English, but I was there two days, much of the time confused, before I came to understand that their use of the term policy corresponded to my American use of the term program. I interpreted policies, from an American context, to be fairly general directives, often very difficult to evaluate because of
their vagueness. In Sweden, however, policies are very specific programs.

The situation becomes more precarious when a translator or interpreter must be used because of language differences. Using an interpreter for conducting interviews is fraught with difficulty. Special and very precise training of translators is critical. It is important that questions be asked precisely as you want them asked, and that full and complete answers be translated. Interpreters often want to be helpful by summarizing and explaining responses. This contaminates the interviewee's actual response with the interpreter's explanation to such an extent that you can no longer be sure whose perceptions you have—the interpreter or the interviewee.

There are also words and ideas that simply can't be translated. People who regularly use the language come to know the unique cultural meaning of special terms, but they don't translate well. One of my favorites from the Caribbean is liming. It means something like hanging out, just being, doing nothing—guilt free. In conducting interviews for a program evaluation, a number of participants said they were just "liming" in the program. But that was not meant as a criticism. Liming is a highly desirable state of being, at least to participants. Funders might view the situation differently.

Rheingold (1988) has published a whole book on untranslatable words with special meanings in other cultures. Below are four examples that are especially relevant to evaluators.

* animater (French)—a word of respect for a person who can communicate difficult concepts to general audiences.
* Schlimmesserung (German)—a so-called improvement that makes things worse.
* ts (Chinese)—to understand things and thus take them lightly.
* biga peula (Kiriwina, New Guinea)—potentially disruptive, unredeemable true statements.

All in all, cross-cultural interviewing is intriguing, challenging, rewarding, and not a little precarious. In addition to the possibility of misunderstandings, there may be the danger of contracting disease, including for some what the Chinese call koro—"the hysterical belief that one's penis is shrinking" (Rheingold, 1988: 59).

Qualitative Interviewing

Differing Norms and Values

The high esteem in which science is held has made it culturally acceptable in Western countries to conduct interviews on virtually any subject in the name of science. Such is not the case worldwide. Evaluation researchers cannot simply presume that they have the right to ask intrusive questions. Many topics may be taboo. I have experienced cultures where it was simply inappropriate to ask questions of a subordinate about a superordinate. Any number of topics may be taboo, or at least delicate, for strangers—family matters, political views, who owns what, how people come to be in certain positions, and sources of income.

There are also different norms governing interactions. I remember with great embarrassment going to an African village to interview the Chief. The whole village was assembled. Following a brief welcoming ceremony, I asked if we could begin the interview. I expected a private, one-on-one interview. He expected to perform in front of and involve the whole village. It took me a while to understand this, during which time I kept asking to go somewhere else so we could begin the interview. He did not share my concern about and preference for privacy. What I expected to be an individual interview soon became a whole village focus group interview!

In many cultures it is a breach of etiquette for an unknown man to ask to meet alone with a woman. Even a female interviewer may need the permission of a husband, brother, or parent to interview a village woman. A female colleague created a great commotion, and placed a woman in jeopardy, by pursuing a personal interview without permission from the male head man.

As difficult as cross-cultural interviewing may be, it is still far superior to standardized questionnaires for collecting data from non-literate villagers. Salmen (1987) describes a major water project undertaken by the World Bank based on a needs assessment survey. The project was a failure because the local people ended up opposing the approach used. His reflection on the project's failure includes a comparison of survey and qualitative methods.

Although it is difficult to reconstruct the events and motivation that led to the rejection there is little question that a failure of adequate communication between project officials and potential beneficiaries was at least
partly responsible. The municipality's project preparation team had conducted a house-to-house survey in Guasmo Norte before the outset of the project, primarily to gather basic socioeconomic data such as family size, employment and income. The project itself, however, was not mentioned at this early stage. On the basis of this survey, World Bank and local officials had decided that standpipes would be more affordable to the people than household connections. It now appears, from hindsight, that the questionnaire survey method failed to elicit the people's negative attitude toward standpipes, their own criterion of affordability, or the opposition of their leaders who may have played on the negative feelings of the people to undermine acceptance of the project. Qualitative interviews and open discussions would very likely have revealed people's preferences and the political climate far better than did the preconstructed questionnaire. (Salmen, 1987: 37)

Appropriately, Salmen's book, published by the World Bank, is called *Listen to the People* and advocates qualitative methods for international project evaluation of development efforts.

Interviewers are not in the field to judge or change values and norms. Researchers are there to understand the perspectives of others. Getting valid, reliable, meaningful, and usable information in cross-cultural environments requires special sensitivity to and respect for differences. For additional discussion of cross-cultural research and evaluation, see Patton (1985) and Lonner and Berry (1986). For examples of doctoral dissertations based entirely on cross-cultural qualitative interviewing, see McClure (1989) and Sandmann (1989).

One final observation on international and cross-cultural evaluations may help emphasize the value of such experiences. Connor (1985) found that doing international evaluations may well make us more effective domestic evaluators. The heightened sensitivity we expect to need in exotic, cross-cultural settings can serve us well in our own cultures. Sensitivity to and respect for other people's values, norms, and worldviews is as needed at home as abroad.

**CREATIVE INTERVIEWING**

Thus far, the discussion of interviewing has focused on asking questions. There are, however, many other ways to elicit responses from people.

*Projection techniques* are widely used in psychological assessment to gather information from people. The best-known projective test is probably the Rorschach. The general principle involved is to have people react to something other than a question—an inkblot, a picture, a drawing, an abstract, a film, a story, a cartoon, or whatever is relevant. This approach is especially effective in interviewing children, but it can be helpful with people of any age. I have found, for example, that in doing follow-up interviews two years after completion of a program, some photographs of the program site and a few program activities greatly enhanced recall.

Another way of stimulating people to talk about their experiences and feelings is to have them react to program products or materials. If the program had materials of some kind (books, handouts, forms, even participant questionnaires), these may be brought to the interview and reviewed to stimulate recall and reactions. This is even more effective if the materials were produced by participants. Students can be interviewed about the work they have produced. In the wilderness program evaluation we interviewed participants about entries in their journals for certain days. (We didn't actually see the journals, but they referred to their journals during the interview and shared whatever they wanted to.)

Robert Kegan and colleagues have had success basing interviews on reactions to ten words in what they call the "subject-object interview . . . . In order to understand how the interviewee organizes interpersonal and intrapersonal experiencing, real-life situations are elicited from a series of ten uniform probes" (Lahay et al., n.d.). The interviewee responds to ten index cards, each listing an idea, concept, or emotion:

1. ANGRY
2. ANXIOUS, NERVOUS
3. SUCCESS
4. STRONG STAND, CONVICTION
5. SAD
6. TORN
7. MOVED, TOUCHED
8. LOST SOMETHING
9. CHANGE
10. IMPORTANT TO ME
Reactions to these words provide data for the interviewer to further explore for the purpose of elucidating the interviewee’s underlying epistemology or “principle of meaning-coherence” based on Kegan’s work *The Evolving Self* (1982). The subject-object interview is a complex and sophisticated methodology that requires extensive training for proper application and theoretical interpretation. For my purposes, the point is that a lengthy and comprehensive interview interaction can be based around reaction to ten deceptively simple ideas presented on index cards. This is a creative interviewing technique.

The subject-object interview methodology illustrates another basis for interviewing: *including writing as part of the interview*. Prior to interviewing the research participants about the ten ideas, they are given 15-20 minutes to jot down things on the index cards. They subsequently choose which cards to talk about and can use their jottings to facilitate their verbal responses. Such an approach gives interviewees a chance to think through some things before responding verbally.

Another substitute for straight questions in interviewing is to ask for explanations of or reactions to critical incidents. These often come from field observations or previous interviews. The evaluation researcher describes the critical incident to get the interviewee’s perspective on what it means and how it relates to other experiences (McClure, 1989).

There are many creative ways to conduct interviews. The evaluation researcher has considerable freedom to adapt interview methods to specific situations and purposes. Not only are there many variations in what stimuli to use and how to elicit responses, there are also creative possibilities for who conducts interviews.

**Participant Interview Chain**

As a participant observer in the wilderness training program for adult educators, I was involved in (1) documenting the kinds of experiences program participants were having and (2) collecting information about the effects of those experiences on their regular work situations. In short, the purpose of the evaluation was to provide formative insights that could be used to help understand the personal, professional, and institutional outcomes of intense wilderness experiences for these adult educators.

Sufficient time and resources were not available to permit much data collection by the two of us doing the evaluation. Therefore, we began discussing with the program staff ways in which the participants might become involved in the data-collection effort to meet both program and evaluation needs. The staff liked the idea of involving participants, thereby introducing them to observation and interviewing as ways of expanding their own horizons and deepening their perceptions.

The participants’ backpacking field experience was organized in two groups of ten participants each. We used this fact to design a data-collection approach that would fit with the programmatic needs for sharing of information between the two groups. Participants were paired for interviewing each other. At the very beginning of the first trip, before people knew each other, all of the participants were given a short, open-ended interview of ten questions. They were told that each of them, as part of their project participation, was to have responsibility for documenting the experiences of their pair-mate throughout the year. They were given a little bit of interview training, a lot of encouragement about probing, and told to record responses fully thereby taking responsibility for helping to build this community record of individual experiences. They were then sent off in pairs and given two hours to complete the interviews with each other, recording the responses by hand.

At the end of the ten-day experience, when the separate groups came back together, the same pairs of participants, consisting of one person from each group, were again given an interview outline and sent off to interview each other about their respective experiences. This served the program need for sharing of information and an evaluation need for the collection of information. The trade-off, of course, was that with the minimal interview training given the participants and the impossibility of carefully supervising, controlling, and standardizing the data collection, the resulting information was of variable quality. On the other hand, there were not sufficient evaluation resources for the two evaluators to conduct 20 in-depth interviews with all of the participants, nor was there time at either the beginning or the end of the experiences to allow such data collection to occur. This mode of data collection also meant that confidentiality was minimal and certain kinds of information might not be shared.
There are limitations to how far one can push client involvement in data collection and analysis. But before those limits are reached, there is a great deal of useful information that can be collected by involving program participants in the actual data-collection process. I have since used similar participant interview pairs in a number of training program evaluations with good results. The trick is to integrate the data collection into the program.

**Data Collection by Program Staff**

Another resource for data collection that is often overlooked is the program staff. Raising the possibility of involving program staff in data collection immediately raises objections about staff subjectivity, data contamination, losses of confidentiality, the vested interests of staff in particular kinds of outcomes, and the threat that staff can pose to clients or students from whom they are collecting the data. Balancing these objections are the things that can be gained from staff involvement in data collection: greater staff commitment to the evaluation, greater staff understanding of the data-collection process, training staff in data-collection procedures, increased understanding by staff of program participants’ perceptions, increased data validity because of staff rapport with participants, and cost savings in data collection.

One of my first evaluation experiences was studying a program to train teachers in open education at the University of North Dakota. Faculty were interested in evaluating that program, but there were almost no resources available for a formal evaluation. There certainly was not enough money available to bring in an external evaluation team to design the study, collect data, and analyze the results. The main data collection consisted of in-depth interviews with student teachers in 24 different schools and classrooms throughout North Dakota and structured interviews with 300 parents who had children in those classrooms. The only evaluation monies available would barely pay for the transportation and the actual mechanical costs of data collection. The interviewers were the staff and students at the university. Structured interview forms were developed for both the teacher and the parent interviews; a full day of training was given to all of the interviewers; and a highly structured system of assigning interviewers to geographical areas was worked out so that no staff were collecting data from their own student teachers. The in-depth interviews with student teachers were tape-recorded and transcribed. The parent interviews involved a precoded, structured instrument. I did follow-up interviews with a 5% sample of the parents as a check on the validity and reliability of the student and staff data.

After data collection, seminars were organized for staff and students to share their personal perceptions based on their interview experiences. It was clear that the interviewing had had an enormous impact on both staff and students. One major outcome was the increased respect both staff and students had for the parents. They found the parents to be perceptive, caring, and deeply interested in the education of their children. Prior to the interviewing many of the interviewers had held quite negative and derogatory images of North Dakota parents. The systematic interviewing had put them in a situation where they were forced to listen to what parents had to say, rather than tell parents what they (as educators) thought about things, and in that listening it was clear that they had learned a great deal. The formal analysis of the data yielded some interesting findings about the program; the evaluation was used to make some changes in the program; and the data provided a source of case materials used in training future program participants. But it is very likely that the major and most lasting impact of the evaluation was the actual experiences of students and staff, who learned a great deal by participating in the data collection. That experiential impact was more powerful than the formal findings of the study.

By the way, had the interviewers been paid at the going commercial rate, the data collection could have cost at least $30,000 just in personnel expenses. As it was, there were no personnel costs in data collection and a considerable human contribution was made to the university program by both students and staff.

**Interactive Group Interviewing**

The involvement of program staff or clients as colleagues in program evaluation changes the relationship between evaluators and staff. The relationship becomes interactive and cooperative rather than one-sided and antagonistic. William Titchenoff (1980) used an “interactive research” approach in educational research and development projects. He found that putting teachers, researchers, and trainers/developers together as a team increased both the meaningfulness and the validity of the findings because teacher cooperation with and
understanding of the research made the research less intrusive, thus reducing rather than increasing reactivity. Their discussions were a form of group interview in which they all asked each other questions.

The problem of how research subjects or program clients will react to staff involvement in an evaluation, particularly involvement in data collection, needs careful scrutiny and consideration in each situation in which it is attempted. Reactivity is a potential problem in both conventional and nonconventional designs. Breaches of confidence and/or reactivity-biased data cannot be justified in the name of creativity. On the other hand, as Tikunoff's experiences indicate, interactive designs may increase the validity of data and reduce reactivity by making evaluation less intrusive and making subjects or clients less resistant or suspicious.

Creativity and Data Quality

There can be no definitive list of creative interview or evaluation approaches. Such a list would be a contradiction in terms. Creative approaches are those that are situationally responsive, appropriate, credible, and useful.

Some of the creative interviewing approaches presented in this chapter are complete deviations from standard practice. Each idea is subject to misuse and abuse if applied without regard for ways in which the quality of the data collected can be affected because of threats to reliability and validity. I have not discussed such threats and possible errors because I believe it is impossible to identify in the abstract and in advance all the trade-offs involved in balancing concerns for accuracy, utility, feasibility, and propriety. For example, having program staff do client interviews in an outcomes evaluation could (1) seriously reduce the validity and reliability of the data, (2) substantially increase the validity and reliability of the data, or (3) have no measurable effect on data quality. The nature and degree of effect would depend on staff relationships with clients, how staff were assigned to clients for interviewing, the kinds of questions being asked, the training of the interviewers, attitudes of clients toward the program, and the like. Program staff might make better or worse interviewers than external evaluation researchers depending on these and other factors. An evaluator must grapple with these kinds of data-quality questions for all designs, particularly nontraditional approaches.

Practical, but creative, data collection consists of using whatever resources are available to do the best job possible. There are many constraints. Our ability to think of alternatives is limited. Resources are always limited. This means that data collection will be imperfect, so dissenters from evaluation findings, who want to attack a study's methods, can always find some grounds for doing so. A major reason for actively involving evaluation users in methods decisions is to deal with weaknesses and consider trade-off threats to data quality before data are collected. By strategically calculating threats to utility, as well as threats to validity and reliability, it is possible to make practical decisions about the strengths of creative and nonconventional data-collection procedures (Patton, 1987a).

Creativity begins with being open to new data-collection possibilities. Naturalistic inquiry calls for ongoing openness to whatever emerges in the field, including during interviews. This openness means avoiding forcing new possibilities into old molds. The admonition to be open and creative applies throughout naturalistic inquiry, from design through data collection and into analysis. Failure to remain open and creative can lead to the error made by a traveler who came across a peacock for the first time, a story told by Halcolm.

A traveler to a new land came across a peacock. Having never seen this kind of bird before, he took it for a genetic freak. Taking pity on the poor bird, which he was sure could not survive for long in such deviant form, he set about to correct nature's error. He trimmed the long, colorful feathers, cut back the beak, and dyed the bird black. "There now," he said, with pride in a job well done, "you now look more like a standard guinea hen."

RECORDING THE DATA

No matter what style of interviewing is used, and no matter how carefully one words interview questions, it all comes to naught if the interviewer fails to capture the actual words of the person being interviewed. The raw data of interviews are the actual quotations spoken by interviewees. There is no substitute for these data.

Data interpretation and analysis involve making sense out of what people have said, looking for patterns, putting together what is said in one place with what is said in another place, and integrating what different people have said. These are processes that belong primarily
to the analysis phase of qualitative evaluations after the data are collected. During the interviewing process itself—that is, during the data-collection phase of evaluation—the purpose of each interview is to record as fully and fairly as possible that particular interviewee’s perspective. Some method for recording the verbatim responses of people being interviewed is, therefore, essential.

Tape-Recording Interviews

A tape recorder is part of the indispensable equipment of researchers using qualitative methods. Tape recorders do not “tune out” conversations, change what has been said because of interpretation (either conscious or unconscious), or record words more slowly than they are spoken. (Tape recorders, do, however, break down and malfunction—a point addressed in the next section.) In addition to increasing the accuracy of data collection, the use of a tape recorder permits the interviewer to be more attentive to the interviewee. The interviewer who tries to write down every word will have a difficult time responding appropriately to interviewee needs and cues. The pace of the interview can become decidedly nonconversational. In brief, the interactive nature of in-depth interviewing can be seriously affected by the attempt to take verbatim notes during the interview.

How interview notes are taken depends on the nature of the fieldwork. Obviously a researcher doing conversational interviews as part of covert fieldwork does not walk around with a tape recorder. However, most evaluation interviews are arranged in such a way that tape recorders are appropriate if properly explained to the interviewee:

I’d like to tape record what you have to say so that I don’t miss any of it. I don’t want to take the chance of relying on notes and thereby miss something that you say or inadvertently change your words somehow. So, if you don’t mind, I’d very much like to use the recorder. If at any time during the interview you would like to turn the tape recorder off, all you have to do is press this button on the microphone, and the recorder will stop.

The use of the tape recorder does not eliminate the need for taking notes. Notes can serve at least two purposes: (1) Notes taken during the interview can help the interviewer formulate new questions as the interview moves along, particularly where it may be appropriate to check out something that was said earlier; and (2) taking notes about what is said will facilitate later analysis, including locating important quotations from the tape itself. In addition, note taking is one of the nonverbal behaviors that helps pace the interview. Note taking becomes a kind of nonverbal feedback to the interviewee about when something was sufficiently important to have been written down; conversely, the failure to take notes will often indicate to the respondent that nothing of particular importance is being said.

I want to emphasize again that the use of a tape recorder does not mean the interviewer can become less attentive to the respondent. This is important regardless of whether a standardized, open-ended interview format is used or the more informal conversational approach is the basis for data collection.

One’s full attention must be focused upon the interviewee. One must be thinking about probing for further explication or clarification of what he is now saying: formulating probes linking up current talk with what he has already said; thinking ahead to putting in a new question that has now arisen and was not taken account of in the standing guide (plus making a note at that moment so one will not forget the question); and attending to the interviewee in a manner that communicates to him that you are indeed listening. All of this is hard enough simply in itself. Add to that the problem of writing it down—even if one takes shorthand in an expert fashion—and one can see that the process of note-taking in the interview decreases one’s interviewing capacity. Therefore, if conceivably possible, tape record; then one can interview. (Lofland, 1971:89)

Transcribing Interviews

Because the raw data of interviews are quotations, the most desirable data to obtain would be full transcription of interviews. Unfortunately, transcribing is enormously expensive. At the Minnesota Center for Social Research, we found that the ratio of transcribing time to tape time was typically 4:1—on the average, it took four hours to transcribe one hour of tape. Despite these costs, full transcriptions are the most desirable data to obtain. Transcripts can be enormously useful in data analysis and later in replications or independent analyses of the data.
Where resources are not sufficient to permit full transcriptions, the interviewer can work back and forth between interview notes and sections of the tape; only those quotations that are particularly important for data analysis and reporting need be transcribed. In either case, whether the full tape is transcribed or only parts of the tape are used to preserve exact quotations, it is critical that the tape recording be of high technical quality. Few things are more distressing in collecting qualitative data than finding that the tape is blank or that background noise is so severe that the tape is virtually worthless. In the first large-scale interviewing project with which I was involved, nearly 20% of the data were lost because of poor-quality recordings. Transcribers are particularly sensitive to the quality of tapes, and costs vary directly with tape-recording quality. Because of the continuing problem of poor-quality recording, transcribers at the Minnesota Center for Social Research put together the following suggestions for interviewers using tape recorders.

How to Keep Transcribers Sane:

High Quality Tape Recording

I. Equipment
   a. Use electrical outlet and outside mike whenever possible.
   b. If you use batteries check them.
   c. Recorder should be clean and in good condition—check before going to an interview.
   d. Take along extra tape cassettes.

II. Before Interview
   a. Choose a place that’s quiet and free from interruptions.
   b. Place microphone close to respondent, then speak loud enough so we hear what you’re saying; most important, we want to hear the answer.
   c. Set recorder on stable surface.
   d. Test the recording system.

III. During Interview
   a. Speak clearly and not too fast—respondent is likely to do the same.
   b. Ask respondent to speak clearly.
   c. Make test with respondent: Then rewind and listen so respondent can hear whether she/he is speaking distinctly; if not, say, “The recorder does not seem to be picking up well. Could you speak up a little?” Whether the problem is mechanical or personal, correct it before continuing.
   d. Don’t rustle papers, cups, bottles, and so on near the mike.
   e. Turn off recorder during irrelevant discussion.
   f. Watch for tape breakage and tangling.
   g. Follow all cassette recorder instructions.
   h. Repeat test if tape change is necessary.
   i. At end of interview, say “This is the end of interview with.”

IV. After Interview
   a. Listen to tape—make notes and erase irrelevant discussion (make note of this for transcribers); list proper names and unfamiliar terminology.
   b. Label tapes and return them to appropriate containers.
   c. Keep tapes and recorder in good condition—do not touch tape or expose it to extreme temperatures.

Taking Notes During Interviews

When a tape recorder is being used during the interview, notes will consist primarily of key phrases, lists of major points made by the respondent, and key terms or words shown in quotation marks that
capture the interviewee’s own language. While most interviewers will not know how to take secretarial shorthand, it is enormously useful to develop some system of abbreviations and informal shorthand to facilitate taking notes. Some important conventions along this line include the following: (1) use quotation marks only to indicate full and actual quotations; (2) develop some mechanism for indicating interpretations, thoughts, or ideas that may come to mind during the interview—for example, the use of brackets to set off one’s own ideas from those of the interviewee; and (3) keep track of questions asked as well as answers received. Questions provide the context necessary for interpreting answers.

When it is not possible to use a tape recorder because of some sensitive situation, interviewee request, or tape recorder malfunction, notes must become much more thorough and comprehensive. Again, it is critical to gather actual quotations as often as possible. When the interviewee has said something that seems particularly important or insightful, it may be necessary to say: “I’m afraid I need to stop you at this point so that I can get down exactly what you said, because I don’t want to lose that particular quote. Let me read back to you what I have and make sure it is exactly what you said.” This point emphasizes once again the importance of capturing what people say in their own words.

After the Interview

The period after an interview or observation is critical to the rigor and validity of qualitative inquiry. This is a time for guaranteeing the quality of the data.

The first thing to be done after a recorded interview is to check the tape to make sure it was functioning properly. If, for some reason, a malfunction occurred, the interviewer should immediately make extensive notes of everything that can be remembered. Even if the tape functioned properly, the interviewer should go over the interview notes to make certain that they make sense, to uncover areas of ambiguity or uncertainty, and to review the quality of information received from the respondent. Did you find out what you really wanted to find out in the interview? If not, what was the problem? Poorly worded questions? Wrong topics? Poor rapport?

On occasion this process of immediately reviewing and elaborating the interview will reveal areas of ambiguity or uncertainty where you are not really sure what the person said or meant. As soon as these areas of vagueness are found, you should check back with the respondent for clarification. This can often be done over the telephone. In my experience, people who are interviewed appreciate such a follow-up because it indicates the seriousness with which the interviewer is taking their responses. Guessing the meaning of a response is unacceptable; if there is no way of following up the comments with the respondent, then those areas of vagueness and uncertainty simply become missing data.

Immediately following an interview is also the time when observations should be written down about the interview itself. The interviewer should note where the interview occurred, who was present, observations about how the interviewee reacted to the interview, observations about the interviewer’s own role in the interview, and any additional information that would help establish a context for interpreting and making sense out of the interview. If a tape recorder is available, the interviewer may simply want to record these observations. This period after an interview or observation is a critical time of reflection and elaboration. It is a time of quality control to guarantee that the data obtained will be useful, reliable, and valid.

This requires great discipline. Interviewing or observing can be exhausting, and it is easy to forgo this time of reflection and elaboration, put it off, or neglect it altogether. To do so is to seriously undermine the rigor of qualitative methods. Interviews and observations should be scheduled so that sufficient time is available for data clarification, elaboration, and evaluation. When a team is working together, the whole team needs to meet regularly to share observations and debrief together. This is the beginning of analysis, because, while the situation and data are fresh, insights can occur that might otherwise have been lost. Thus ideas and interpretations that emerge following an interview or observation should be written down and clearly marked as such.

THE POWER OF INTERVIEWS: IMPACTS AND ETHICS

Interviews are interventions. They affect people. A good interview lays open thoughts, feelings, knowledge, and experience not only to the interviewer but also to the interviewee. The process of being taken
through a directed, reflective process affects the persons being interviewed and leaves them knowing things about themselves that they didn’t know—or least were not aware of—before the interview. Two hours or more of thoughtfully reflecting on an experience, a program, or one’s life can be change-inducing.

Yet, the purpose of a research interview is first and foremost to gather data, not change people. Earlier, in the section on neutrality, I asserted that an interviewer is not a judge. Neither is a research interviewer a therapist. Staying focused on the purpose of the interview is critical to gathering high-quality data. Still, there will be many temptations to stray from that purpose. It is common for interviewees to ask for advice, approval, or confirmation. Yielding to these temptations, the interviewer may become the interviewee—answering more questions than are asked.

On the other hand, the interviewer, in establishing rapport, is not a cold slab of granite—unresponsive to the human issues, including great suffering and pain, that may unfold during an interview. In a major farming systems needs assessment project to develop agricultural extension programs for distressed farm families during the farm crisis of the mid-1980s, I was part of a team of ten interviewers (working in pairs) who interviewed 50 farm families. Many of these families were in great pain. They were losing their farms. Their children had left for the city. Their marriages were under stress. The two-hour interviews traced their family history, their farm situation, their community relationships, and their hopes for the future. Sometimes questions would lead to husband-wife conflict. The interviews would open old wounds, lead to second-guessing decisions made long ago, or bring forth painful memories of dreams never fulfilled. People often asked for advice—what to do about their crops, their children, government subsidy programs, even their marriages. But we were not there to give advice. We were there to get information that might, or might not, lead to new programs of assistance. Could we do more than just ask our questions and leave? Yet, as researchers, could we justify in any way intervening? Yet again, our interviews were already an intervention. Such are the ethical dilemmas that derive from the power of interviews.

What we decided to do in the farm family interviews was leave each family a packet of information about resources and programs of assistance, everything from agricultural referrals to financial and family counseling. To avoid having to decide which families really needed such assistance, we left the information with all families. When interviewees asked for advice during the interview, we could tell them that we would leave them referral information at the end of the interview.

While interviews may be intrusive in reopening old wounds, they can also be healing. In doing follow-up interviews with families who had experienced child sexual abuse, we found that most mothers appreciated the opportunity to tell their stories, vent their rage against the system, and share their feelings with a neutral, but interested, listener. Our interviews with elderly residents participating in a program to help them stay in their homes and avoid nursing home institutionalization typically lasted much longer than planned because the elderly interviewees longed to have company and talk. When interviewees are open and willing to talk, the power of interviewing poses new risks. People will tell you things they never intended to tell you. This can be true even with reluctant or hostile interviewees, a fact depended on by journalists. Indeed, it seems at times that the very thing someone is determined not to say is the first thing they tell, just to release the psychological pressure of secrecy or deceit.

I repeat, people in interviews will tell you things they never intended to tell. Interviews can become confessions, particularly under the promise of confidentiality. But beware that promise. Social scientists can be summoned to testify in court. We do not have the legal protection that clergy and lawyers have. In addition, some information must be reported to the police—for example, evidence of child abuse. Thus the power of interviewing can put the interviewees at risk. The interviewer needs to have an ethical framework for dealing with such issues.

There are also direct impacts on interviewers. We learned that the interviewers in the family sex abuse project needed to be extensively debriefed, sometimes in focus groups, to help them process and deal with the things they heard. They could only take in so much without having some release, some safety valve for their own building anger and grief. Middle-class interviewers going into poor areas may be shocked and depressed by what they hear and see. It is not enough to do preparatory training before such interviewing. Interviewers may need debriefing—and their observations and feelings can become part of the data on team projects.
Qualitative Interviewing

(7) Advice. Who will be the researcher’s confidant and counselor on matters of ethics during a study? (Not all issues can be anticipated in advance. Knowing who you will go to in the event of difficulties can save time and bring comfort.)

PERSONAL REFLECTIONS ON INTERVIEWING

This chapter has offered suggestions about how to interview effectively. There is no single right way of interviewing, no single correct format that is appropriate for all situations, and no single way of wording questions that will always work. The particular evaluation situation, the needs of the interviewee, and the personal style of the interviewer all come together to create a unique situation for each interview. Therein lies the challenge of qualitative interviewing.

I find that interviewing people can be invigorating and stimulating. It is a chance for a short period of time to try to get inside another person’s world. If participant observation means “walk a mile in my shoes,” then in-depth interviewing means “walk a mile in my head.”

I’m personally convinced that to be a good interviewer you must like doing it. This means taking an interest in what people have to say. You must yourself believe that the thoughts and experiences of the people being interviewed are worth knowing. In short, you must have the utmost respect for these persons who are willing to share with you some of their time to help you understand their world. There is a Sufi story that describes what happens when the interviewer loses this basic sensitivity to and respect for the person being interviewed.

An Interview with the King of the Monkeys

A man once spent years of his life learning the language of monkeys so that he could personally interview the King of monkeys. Having completed his studies he made careful inquiries to find the King of the monkeys. In the course of searching for the King of the monkeys he had to talk to a number of monkey underlings. He found that the monkeys he spoke to were generally, to his mind, neither very interesting, nor very clever. He began to doubt whether he could learn very much from the King of the monkeys either.

Finally he located the King of the monkeys and arranged for an interview. Because of his doubts, however, he decided to begin with a few
basic questions before moving on to the deeper questions in which he 
was really interested. "What is a tree?" he asked. 
"It is what it is," said the King of the monkeys. "We use trees to swing 
on." 
"And what is the purpose of the banana?" 
"They are to eat." 
"How do animals find pleasure?"
"By doing things they enjoy."
At this point the man decided that the King's responses were rather 
shallow and uninteresting, and went on his way, severely disappointed. 
Soon afterwards, an owl flew into the tree next to the King of the 
monkeys. "What was that man doing here?" the owl asked. 
"Oh, he was only another silly human," said the King of the monkeys. 
"He asked a bunch of simple and meaningless questions, so I gave him 
simple and meaningless answers."

Not all interviews are interesting and not all interviews go well. 
Certainly there are uncooperative respondents, people who are para-
noid, respondents who seem overly sensitive and easily embarrassed, 
aggressive and hostile interviewees, timid people, and the endlessly 
verbose who go on at great length about very little. When an interview 
is going badly it is easy to call forth one of these stereotypes to explain 
how the interviewee is ruining the interview. Such blaming of the 
victim (the interviewee), however, does little to improve the quality 
of the data. Nor does it improve interviewing skills. 

A different approach is to believe that there is a way to unlock the 
internal perspectives of every interviewee. It is the task and responsi-
bility of the interviewer to find which interviewing style and which 
question format will work with a particular respondent. It is the 
responsibility of the interviewer to establish an interview climate that 
facilitates open responses. When the interview goes badly, it is the 
responsibility of the interviewer, not the fault of the interviewee.

Halcolm's Evaluation Interviewing Beatitudes

Ask.

Listen and record.

Ask.

Listen and record.

Qualitative Interviewing

It is a grave responsibility to ask.
It is a privilege to listen.

Evaluators, listen. Do you not know that you shall be evaluated by 
your questions?
To ask is to seek entry into another's world. Therefore, ask respect-
fully and with sincerity. Do not waste questions on trivia and tricks, 
for the value of the answering gift you receive will be a reflection of 
the value of your question.
Blessed are the skilled questioners, for they shall be given moun-
tains of words to ascend.
Blessed are the wise questioners, for they shall unlock hidden 
corridors of knowledge.
Blessed are the listening questioners, for they shall gain perspective.