

The use of qualitative measures in a quantitative framework results in a reasonable likelihood of triangulation; in contrast, the independent conduct of qualitative and quantitative evaluations is a greater challenge for triangulation, but it also holds promise for greater discovery.

Qualitative and Quantitative Methods: When Stories Converge

Louise H. Kidder, Michelle Fine

Questions about whether qualitative and quantitative research complement each other, whether they lead to compatible conclusions, and whether they stand up equally well under methodological scrutiny can themselves be answered with research evidence, and we will attempt to do this. The same questions, however, can also be understood in another way, as questions about two cultures. Members of these two cultures or research traditions are genuinely curious about one another but also harbor beliefs and judgments about their own relative superiority. Quantitative researchers who value numerical precision over "navel gazing" and qualitative researchers who prefer rich detail to "number crunching" betray not only a preference for one but also a distrust of the other. One way methodologists have attempted to allay the distrust is to call for synthesis, collaboration, and cooperation between the two cultures (see, for example, Cronbach and others, 1980; Goetz and LeCompte, 1984; Kidder and Judd, 1986; Reichardt and Cook, 1979). We share the call for "synthesis," but at the same time, we want to preserve the significant differences between the two cultures. Instead of homogenizing research methods and cultures, we would like to see researchers become bicultural. Rather than

“closing down the conversation” about quantitative-qualitative differences, we want to sustain it (compare Smith and Heshusius, 1986).

Triangulation: When Is It Possible and When Is It Probable?

Suppose two field workers independently set out to study the same parole system. Would they write the same story and arrive at the same conclusions? Would either of their conclusions agree with the results of a quantitative evaluation?

We have heard both answers to these questions. A psychohistorian observed ironically: When two quantitative researchers arrive at the same conclusions, we call it “reliability,” but when two biographers write the same story we call it “plagiarism” (Runyan, personal communication; Runyan and Seal, 1985). Perhaps the irony derives not so much from a fear of being charged with plagiarism as from a desire to discover something new. A somewhat confabulated memory of a hallway conversation between two eminent methodologists conveys the same message:

Quantitative: If two of your students went to the same street corner to study the same problem, would they come back with the same conclusion?

Qualitative: Probably not, why would they want to?

Quant.: Why, to demonstrate the reliability of the findings.

Qual.: But I'd rather have them find something new instead.

(Partially reconstructed conversation between Howard S. Becker and Donald T. Campbell.)

Sometimes the distinction between quantitative and qualitative research is portrayed as a difference in focus or scale. Quantitative work, when equated with laboratory research, is presumed to have a narrow or segmented focus, while qualitative work is portrayed as holistic, as in the following description by a primate anthropologist:

“The field worker and the laboratory [researcher] . . . tend to adopt different but compatible methods of achieving perspective. The methods are analogous to *zooming in* and *zooming out* with a lens. To the extent that they are reproduced objectively, wide-angle, telephoto, and microscopic views must be *simultaneously* valid, and *zooming* from different directions merely focuses attention on different facets of the same phenomenon” (E. W. Menzel, primate anthropologist, cited in Guba, 1978, p. 4; Willems and Rausch, 1969, pp. 82-83).

This interpretation implies that the differences between quantitative and qualitative work are like the differences between behavioral and physiological psychology or the differences between biology and biochemistry. The telephoto-microscope analogy suggests that qualitative and quantita-

tive methods reveal different levels of activity and create different levels of explanation that do not compete with each other because they address different questions. We want to add another element to the camera imagery—an element of movement versus fixed position. What is distinctive about quantitative research is the standardization of data collection—standardized instruments, uniformly administered, produce data that can be added and subtracted or at least tallied. The analogous photographic technique is the use of still camera shots with a set focus and shutter speed to produce passport or driver's license photos. By contrast, qualitative research methods are like a roving movie camera with variable exposures, a shifting focus, and nonroutinized selections of angles, durations, film speed, and so on. The roving camera and variable focus are not merely the consequence of studying people or animals moving about in their natural habitats. The movement and variation are deliberate aspects of the method itself—a method that can be used to study the fixed written records of psychohistory as well as living persons (compare Crosby and Crosby, 1981; Runyan, 1984; Stewart and Healy, 1986).

We have spoken of “qualitative” research as though it were all of a kind, and similarly for “quantitative.” We need to stop at this point to distinguish between two meanings of *qualitative*. (Someone else might want to make distinctions between two or more meanings of *quantitative*—certainly the distinction between laboratory and survey work is salient.)

Two Meanings of Qualitative: The Big *Q* and the Small *q*

Qualitative work with the big *Q* is field work, participant observation, or ethnography; it consists of a continually changing set of questions without a structured design. The big *Q* refers to unstructured research, inductive work, hypothesis generation, and the development of “grounded theory” (compare Glaser and Strauss, 1967).

Qualitative work with the small *q* consists of open-ended questions embedded in a survey or experiment that has a structure or design. Respondents' answers to the open-ended questions might vary greatly, and the purpose of including open-ended questions is usually to provide greater “richness” or “latitude” or to allow for the unexpected. The interviewer generally asks the questions in roughly the same order and manner, probing when the answers are too short, vague, or not forthcoming. Qualitative measurements (small *q*) embedded in structured research are usually selected to test a hypothesis stated at the outset. The hypothesis and questions do not change as the research progresses. The same questions are asked of everyone.

The insertion of open-ended questions into an experiment or survey does not change the nature of the research from deductive to inductive.

Fifty-one open-ended questions interspersed with forty-nine structured-response items, or even ninety-nine interspersed with one, would not automatically change the nature of the research from deductive to inductive, neither would it change the mode from either quantitative or qualitative to Qualitative. The difference between deductive and inductive research lies not in the percentage of structured and unstructured questions but in how open-ended the research process itself is. Researchers who work deductively gather data to test specific hypotheses, not to generate new hypotheses, and serendipitous findings are considered interesting but unreliable. By contrast, researchers who work inductively continue to generate hypotheses and look for new questions even as they gather data.

We have overdrawn the distinction to make a point. Inductive and deductive research modes need not be mutually exclusive. These are better understood as end points on a continuum, with a variety of mixtures in between. Nonetheless, the distinction is heuristically useful to compare varieties of triangulation.

Looking for Answers Versus Looking for Questions. Qualitative (big *Q*) researchers begin their data collection with a hunch rather than an interview schedule of questions, and as they proceed, they continually discover new questions relevant to the people under study. Often fumbling at first, these researchers rely on the authentic participants, the people who belong there, to discover which questions make sense. Field workers begin with a vague hunch about "what's going on," and as they proceed, they revise both their questions and their hunches. A daily chore of a participant observer is deciding which questions to ask next of whom (Agar, 1980).

Experimenters and survey researchers, by contrast, have a set of questions prepared before they begin to gather data. Before deciding on a final design and instruments, they might conduct "pilot" work and revise the procedures numerous times, but once they begin to gather data, the revisions stop. The task of an experimenter or survey researcher is to follow a procedure or ask a set of questions as uniformly as possible, to treat successive respondents alike, and not to raise new questions with new respondents.

Field workers, however, do not structure their research in advance to study "the affect of X on Y" because they do not know from the outset whether X or Y will prove to be relevant as experienced by the people under study. They do not uniformly measure the X and Y attributes of every person or group because it is not obvious that the research is about X and Y. Instead they gather people's stories and observe their lives and in the process identify and define what the "variables" are, though this language is rarely used. With the collection of stories or events, the researcher develops a narrative, an account of what led to what, and each new story or event is used to confirm or revise the narrative.

Combining Measures. Quantitative measures have been used in field work, just as qualitative measures have been introduced into experiments or surveys. The introduction of alternative measures, however, does not automatically transform the overall nature of the work.

For example, Campbell (1955) describes research on the morale of submarine crews in which he combined self-report questionnaire measures of morale with the judgments of "informants." He chose as his informants men who performed secretarial duties in the headquarters and were not regularly aboard ship but who frequently shared meals and short cruises with the ships' crews and also received shore visits from crew members. The informants were privy to the "gossip" about morale and felt confident to speak about and rank the morale levels of ten ships. Campbell found an impressive level of agreement between the judgments of his informants and the morale questionnaires of the ships' crew members. Campbell's study demonstrates that it is possible to combine informants' judgments with subjects' questionnaires and find a high level of agreement. It also illustrates our contention, however, that this novel combination does not change the nature of the research. The introduction of the informants' judgments did not transform the study into field work or participant observation. Instead it demonstrated that independent judgments of morale, measured from a novel vantage point, converged with the judgments of the subjects—it provided convergent validity of measurement. The informants' judgments did not lead Campbell down unanticipated paths—they did not turn him into a participant observer.

Consider another example: McCleary (1978) describes an effort to introduce a quantitative index in his field work among parole officers. He tried to measure the amount of paperwork parole officers produce and chose "use of the Xerox machine" as a quantitative index. After recording the numbers of Xeroxed copies in several parole offices, he discovered the the Xerox machine was often appropriated for other purposes unrelated to parole work. Many of the parole officers were taking college courses and used the Xerox machines for classwork. The machine records, therefore, had too many other sources of variance to be reliable indicators of parole officers' paperwork.

McCleary might have tried to cleanse the records to obtain a more reliable quantitative index, but instead he concluded: "I admit to discredit this quantitative indicator, I would have been forced to discard a number of theories. The displeasure in this would have come about simply because I 'knew' the theories were correct" (1978, p. 41). He did not persist in his search for a new quantitative index because his field work did not require that each subject have a number. Therefore, the failed attempt and even a successful attempt at quantification would not have changed the underlying nature of his work—it remained field work rather than survey research.

McCleary's account illustrates several features of field work. He

decided to record use of the Xerox machine sometime after he had started his field work, and he dropped the measure before he completed the study. He describes his trials and errors and does not hide his glee at discrediting a measure that did not work. He reports that he "knew" what was happening despite contradictory quantitative evidence. He was more concerned with developing a coherent narrative than with applying uniform measures of variables. And finally, his research and writing locate him, the researcher, as a participant in the story.

The act of developing a narrative is not unique to field work. Experimenters and quasi-experimenters studying causal relationships between independent and dependent variables also construct a narrative about sequences of events, though theirs is usually not as lively as a field worker's because it is considerably more abstract. Their narratives differ in other ways, too, however. Where and when the narrative is developed differs dramatically between quantitative and Qualitative research. Field workers construct and revise the story continually throughout their data collection, whereas experimenters and survey researchers decide at the end, when the results are subjected to statistical testing, whether to accept or reject the narrative they planned at the outset.

Three Usages of Triangulation

Triangulation of Measurement. The most common meaning of triangulation refers to measurement. Triangulation of measurement means locating a point in space along one or more dimensions to describe a single person (or group or object—whatever the unit of analysis is). For instance, if we try to measure a person's "size," we can locate a point on a tape measure, say sixty-seven inches, which describes his or her size. We could also locate a point on a scale of weights, say 140 pounds, as another quantitative measure. And we could inquire about clothing size as a third quantitative index. For a qualitative index, we could ask observers or "informants" to describe the person in their own words. The more similar the focus of any two measures, the more highly correlated they will be, regardless of whether they are quantitative or qualitative. Therefore, in its common meaning, triangulation is presumably attainable across quantitative and qualitative (small *q*) measures. Campbell's (1955) study of the use of informants to study submarine-crew morale levels is a good example of such triangulation.

The imagery that Webb, Campbell, Schwartz, and Sechrest (1966) introduced in their discussion of triangulation is the imagery of measurement. Like surveyors, researchers try to fix a point in space by taking measurements from two or three locations. This use of triangulation also appears in geometry lessons, locating a point with two intersecting lines, and sometimes a third for good measure. Each of these is an effort to locate a single point in space to describe a person, group, or object.

Whenever we assign someone to a location on one or more dimensions, we know that any single effort to find that location will err. Therefore, measurements should include more than one method, such as a standardized questionnaire of psychological depression combined with a clinical interview. Scales with multiple items that yield roughly similar answers provide one form of triangulation of measurement known as reliability. Distinctly different methods aimed at measuring the same construct provide another form of triangulation of measurement known as validity.

Multiple methods are most likely to yield similar answers if they are designed to hone in on a single point along a single dimension. If researchers design an experiment or quasi-experiment to test the relationship of X and Y and devise three measures of Y, they are most likely to obtain similar results with each measure if the variable is narrowly defined. It does not matter whether the measurements are quantitative or qualitative. If we ask people to circle numbers to indicate whether they believe women should be free to choose abortions, or if we ask them to describe in their own words how they feel about reproductive choices, we are more likely to arrive at roughly the same point if we limit the conditions to first-trimester pregnancies. Even with uni-dimensional measures of a narrowly defined variable, however, experimenters and survey researchers often fail to achieve a high degree of agreement among multiple measures. Few researchers build multitrait-multimethod matrices into their research designs because the effort expended is often rewarded with low correlations (Campbell and O'Connell, 1982; Fiske, 1982).

This meaning of triangulation—agreement between measurements—is the most common and most clearly defined. It is not, however, what field workers or participant observers do. They are not primarily concerned with measurement; they do not routinely ask the same questions of all respondents, and for this reason they often do not arrive at the same conclusions as either quantitative researchers or other field workers.

Triangulation of Conclusions Within a Study. A second and less common usage of triangulation involves agreement about research *conclusions*. This does not require locating individuals along dimensions in space; it requires locating the same beginning and end points of a story and the same choice points or contingencies along the way. Triangulation of conclusions occurs *within* a study if different measurements produce the same pattern of results or if different informants give the same account of an experience or institution, such as "blue collar marriage," for instance (Komarovskiy, 1967; Rubin, 1976), or parole work (McCleary, 1978).

This is a minimal requirement for the validity of field work—that more than one informant provide evidence consistent with the researcher's analysis. The process of negative case analysis (Kidder and Judd, 1986) requires, moreover, that there be no disconfirming evidence. Any instances that contradict the emerging hypothesis are used to revise the hypothesis

until it incorporates all evidence. One meaning of triangulation across conclusions within a study, therefore, is that the final hypothesis (or conclusion) covers all cases. No incident or informant is discounted as "random error" or "an outlier." This all-inclusive character of negative case analysis produces conclusions that are very detailed and descriptive—more nearly a story than a conclusion by comparison with the more spare, bare-bones conclusions of experimental, quasi-experimental, or survey designs.

Triangulation of conclusions across different methods employed within a single study takes a different form when the methods include quantitative and Qualitative. For instance, in research among expatriates in India, Kidder (1977) was both a participant observer and a survey researcher. She wrote field notes of her own and other sojourners' experiences and also asked a sample of over one hundred respondents to answer questionnaires and scales. The risks in conducting research with such mixed methods are not so much that the methods will produce contradictory conclusions as that they will simply diverge—leading to noncomparable rather than incompatible ends.

A field worker follows leads much in the same way as an investigative reporter (compare Nelson, 1982). What begins as a simple hunch becomes more complex, the stories acquire detail, and the hunch becomes revised, reversed, or dropped altogether. Rather than identifying single dimensions on which to measure individuals, field workers trace the actors' histories and careers. They can achieve triangulation within a study by writing a common story, identifying regular routes with similar beginnings, choice points, and endings.

In his study, *Dangerous Men*, McCleary (1978) analyzed the socialization of parole officers. He showed how officers learn to avoid trouble by classifying difficult parolees as "dangerous men." He did not regard each officer as unique and did not trace the idiosyncratic career paths of twenty individual parole officers. Instead, he identified the bureaucratic requirements that encourage most parole officers to classify some parolees as dangerous men and to write their reports in a way that will "avoid trouble." He used corroborating evidence from many parole officers to demonstrate the reliability of his conclusions (he did not say precisely how many parole officers he had observed or interviewed because the precise numbers never enter into a formula). The reader becomes convinced that McCleary's analysis presents a story of parole work, not the story of one or two individuals. This is an example of triangulation of conclusions *within* one person's field work.

The officers to whom McCleary showed his analysis agreed that he accurately described the work of parole officers. At the same time, each felt his own story was somewhat unique. One parole officer said, "I know who this PO [parole officer] you're talking about here is. It's Jerry. You know, what you could do is write another chapter comparing me to Jerry.

The way it is now, you're giving people the impression that all PO's are like Jerry" (1978, p. 35).

McCleary chose not to tell each individual's story, with its unique twists and turns. Instead, he developed a unified story from his observations of many parole officers in two district offices. The narrative that he developed is tantamount to the description of a main effect. Had he developed different narratives for different categories of parole officers (for example, men versus women) or for parole officers in different locations (for example, urban versus rural officers), his analysis would be comparable to an interaction effect (see Kidder, 1981). The unique variations that each parole officer thought applied to himself or herself are like "error variance" or individual differences in survey research and experimental designs. Statistical tests quantify the amount of error variance, and experimenters set alpha levels to specify how much uncertainty they will tolerate. For field workers there are no rules of thumb about how much error variance is tolerable. Their case rests more on how clear and persuasive the instances that support their conclusions are and on how well the conclusions coincide with other knowledge they and their audience possess about the issues.

Within a single study it is possible to hone in on a story that describes the experience of a number of people and to develop a narrative that is confirmed by multiple observations. Field workers can and do achieve triangulation at least within their own studies. We read with greater confidence the story about parole work if we know that McCleary observed and followed numerous officers, not just one who became his good friend. This does not mean there are no other stories about parole work that could be told from other points of view. The parolee has another version, and the official statements of the department of corrections contain a third version. Nonetheless, from studying many different parole officers in more than one district, McCleary was able to find common patterns, common forces, and repeating narratives. He achieved triangulation of his conclusions across people within his study.

If this is possible within the work of a single field researcher, it is presumably possible between field workers, too; possible but improbable. In the mixed-methods study of expatriate sojourners, Kidder (1977) concluded from her field work that expatriates generally became socialized into a "third culture" of other foreigners like themselves. Even those who came with the intention of becoming acquainted with Indian culture and society were tutored by other expatriates and developed a life-style that recreated their home culture. The results of several questionnaire measures, however, indicated the opposite—the longer the expatriate's stay in India, the more the person knew about Indian culture and customs. These apparently contradictory conclusions were consistent with the contradictions in the research literature on foreign visitors; some studies showed an increase

and others showed a decrease in liking for and involvement with host cultures. Being a researcher who stood between multiple methods and conclusions, Kidder was motivated to reconcile the apparent contradictions. (We suspect that any researcher is more motivated to find agreement across conclusions within his or her study than are two or more researchers working independently on separate studies.) This is probably a common experience for Ph.D. candidates who gather more data with more methods than they will ever again include in a single study. Kidder's reconciliation of results required further quantitative analyses showing that sojourners exhibit both positive and negative changes in their attitudes and acquaintance with the host culture. These two opposite trends were not simply the result of "ambivalence" but were actually different currents that appeared with different ways of measuring a sojourner's time away.

This process of reconciling the results of different methods within a single study is much like the process of negative case analysis. The conclusions become more complex as they incorporate more methods and more points of view.

Triangulation of Conclusions Across Studies. Finding agreement between research conclusions *across* studies that use different methods is the greatest challenge. These cases are also the ones we find most interesting. They represent maximal dissimilarity—different investigators with different perspectives using different methods—and therefore have the potential for producing conclusions with maximal construct validity, provided they agree.

One carefully documented case of the reconciliation of qualitative and quantitative analyses reveals the requirements and possibilities when the members of a research organization attempt to draw together divergent methods and conclusions (Trend, 1978). The research organization had contracted to evaluate the work of several administrative agencies that managed direct cash housing allowances to enable low-income families to obtain housing on the open market. In addition to analyzing the quantitative indicators (such as how many families obtained adequate housing at what cost to the administrative agency), the research organization had a field worker analyze the administrative agency's work using field notes from his participant observation.

On first analysis, the quantitative data and the Qualitative field work produced very different conclusions. It appeared as though the Qualitative analysis were "telling a different story" because it focused on discord and conflict within the administrative agency, which, the quantitative research said, was not relevant to the question of the program's costs. Moreover, the staff who worked on the cost analysis doubted the reliability of the field worker's notes—"He was suspected of having been caught up in office politics and of having lost his scientific objectivity" (Trend, 1978, p. 349).

When a second analyst was enlisted to re-examine the field work, the two Qualitative workers revised those conclusions. Their new report suggested that the administrative personnel at the site they had been studying (one of eight sites where the housing allowances were being evaluated) did indeed have serious conflicts because the personnel had been concerned with the *quality* of the living conditions they were providing, whereas the pressures from the contracting agency put a premium on the *quantity* of applicants processed. These tensions had lowered the morale and efficiency of the administrative agency. Just as the Qualitative report was being revised, the quantitative researchers produced new analyses showing that the recipients at this site had received the second highest quality of housing of the eight sites. Tension mounted, and the participant observer who had originally been hired to conduct the field work was dismissed.

When Trend (1978) assumed responsibility for producing yet another Qualitative analysis and reconciliation of the data, he was struck by the fact that "neither side seriously doubted 'the facts' uncovered by either method of inquiry" (p. 349). His task, therefore, was to explain how a program could "produce such admirable results in so many of its aspects, when all of the observational data indicated that the program would be a failure" (p. 349).

The synthesis that Trend created maintained that the disputed site was efficient (cost-effective) in the quantitative analysis in ways that were consistent with the field observations of staff conflict and inefficiency. The field worker's reports of conflict pertained to one of three offices within that site—the urban office that served low-income black families. The other two offices were in rural districts, served white families that did not have the same level of financial need, and had access to better-quality housing. Several factors made this site paradoxically "efficient" or "cost-effective." The rural offices were able to save money because the families they served were smaller and in less financial need, so they received smaller subsidies. The urban office saved money in several ways. The staff decided to limit the number of black enrollees, which meant the office became more efficient on two counts: It required less money for subsidies and a higher percentage of applicants eventually became recipients. At the same time, a number of the urban staff members resigned ahead of schedule because they were discouraged both by the pressure to meet the contracting agency's quotas and by their inability to counsel applicants, given the urgency to serve large numbers of families. The unused portions of those staff member's salaries could be used for other purposes, adding to the apparent cost-effectiveness of the agency. Therefore, Trend's conclusions illustrated how a site with alienated staff members could still appear efficient by other criteria.

Finding agreement between conclusions of different studies, one

quantitative and the other Qualitative, is undoubtedly the greatest challenge and most notable accomplishment in the area of mixed or multiple methods. Finding agreement between conclusions of two Qualitative studies is less notable simply because the methods are not different. Such agreement is nonetheless not commonplace, however, primarily because field workers are more interested in discovering something new rather than in replicating someone else's work.

The desire to discover something new does not preclude triangulation of another sort, however. The convergence we find in field work usually occurs at a more abstract or conceptual level; for example, instead of showing how the socialization of prisoners in one city resembles the socialization of prisoners in another city, two field workers might show how socialization in a mental hospital is similar to socialization in prison, or how aptitude testing of school children mimics psychiatric typing. The triangulation in these instances occurs across settings. The research reveals similarities not between two schools or between two mental hospitals but between presumably distinct institutions, a school and a mental hospital. Much field work, therefore, arrives at conclusions that sound quite like analogies. For instance, in an ethnographic study of a religious community that calls itself the Community of Joy, Zablocki (1971) shows how this voluntary intentional community resembles other total institutions that are neither voluntary nor intentional communities. He does not say, "The Community of Joy is a prison," but he identifies the characteristics of total institutions that characterize prisons, mental hospitals, summer camps, and the Community of Joy.

Field workers and ethnographers are interested in showing similarities across settings, even when the settings they choose for comparisons are likely to be quite distinct. For instance, rather than demonstrating that the process of becoming hypnotized is replicated in many different hypnosis workshops, Kidder has shown what hypnosis has in common with dissonance reduction (Kidder, 1972). Drawing comparisons from disparate rather than similar settings is part of the process of analysis in Qualitative work. Analogies further our understanding of the case under study. If we read that becoming hypnotized resembles dissonance reduction, we learn more about hypnosis than if we read that being hypnotized by Dr. Z is similar to being hypnotized by Dr. B. The appeal of Qualitative research often lies in its ability to reveal parallels between settings that are assumed to be structurally or functionally unrelated.

Research as Story Telling

Research reports never begin, "Once upon a time, in a laboratory far away . . ." When we ask about experiments, we ask, "What were the results?" not "What was the setting and who were the characters?" or,

"Tell me the plot." Yet even quantitative research tells a story. Experiments tell us what we can expect if people find themselves, usually alone or with strangers, in a particular setting with certain events scheduled to take place. The experimental paradigm, "if X, then Y," is the start of a story, albeit a bare-bones story.

In the preceding discussion we described Qualitative researchers as story gatherers and referred to the analysis of field work as the construction of a narrative pertaining to more than one actor. All research is a form of story telling, some more obvious than others. Randomized experiments are the least obvious because their procedures are modeled on the conduct of physical science, not on the creation of narratives. Experimental characters are anonymous, and the authors of experimental reports do not say "I . . ." They say, "It was hypothesized that . . ." or, "The subjects were randomly assigned to one of three treatments. . . ." The results are reports of statistical tests such as, "A t-test revealed significant differences between the treatment and control conditions . . ." instead of characters' actions. Nonetheless, beneath the technical language is a story about how people behave under various conditions.

Field workers or participant observers describe their work in a form and language that more nearly resemble a story. Characters have names, even if they are pseudonyms; authors are likely to identify themselves by saying, "I"; and the findings are often presented in narrative language with descriptions of place, people, and their actions. In McCleary's study he locates himself as both actor and observer in the parole offices: "I first tried measuring the amount of time parole officers spent on paperwork in the office. I discovered what I already knew, however: That PO's spend little or no time in the office writing reports. Most of this work is done at home. Office time is spent answering phones, interviewing clients, and socializing with office-mates and supervisors" (1978, p. 40).

If we regard the construction of social scientific results as a form of story telling, is it possible to tell the story from a neutral perspective? The explicitly narrative quality of field work, the act of constructing theory by analyzing many individual stories, brings the matter of multiple points of view to the forefront. Field workers who observe events in existing social groups are liable to come across conflicting points of view. For this reason alone they might be less likely to develop the same narrative and arrive at the same conclusion as another researcher, even if they set out to do the same study. Unstructured field work or participant observation has a life of its own that unfolds almost despite the researcher's initial plans or intentions. As it unfolds, the field worker adopts the perspective of particular actors in the field, and their perspectives shape the story. For instance, McCleary could have begun to study parole officers and found subsequently that he could gain better access to parolees, and he might therefore written about "dangerous men" from a different perspective.

Experimenters and survey researchers do not allow the research sample and procedures to “unfold” once they have begun gathering data. Samples and procedures are determined in advance, and it is feasible, therefore, to select subjects and questions that will most likely reproduce the same point of view as previous research. Such predetermination makes it more likely that quantitative researchers can find triangulation of conclusions across different studies, provided they use the same selection procedures.

Can There Be a Narrator with No Point of View?

Playwrights and novelists are accustomed to selecting points of view and sometimes portraying events from multiple perspectives. Social research that tells a story can also be told from one or more perspectives. Social programs and program evaluations have various “stakeholders” whose perspectives differ and sometimes conflict (Mark and Shotland, 1985). School desegregation programs, for instance, have multiple stakeholders. Lukas’s (1985) recent account of school desegregation in Boston tells the story from three families’ perspectives—one, white working class, with a long family history of residence in the community; one, white middle class, initially committed to the city but eventually moving to the suburbs; and one, black working class, with a history of civil right activities. Lukas does not suggest which perspective should prevail; he attempts to present all three with an equal voice.

McCleary’s analysis of parole presents the composite voice of one set of stakeholders, the parole officers. It does not tell the story of parole from the parolees’ or the department of corrections officers’ perspectives. Fine’s (1986) field work among high school drop-outs contains more information from the students’ position than from the principals.’

Whose Story Shall Prevail?

Actors and observers who occupy different positions in relation to an organization or an event can disagree in at least three respects: the description of a process, the identification of salient effects, and the selection of causes (Kidder, 1982).

Different Accounts of the Process: Parole Work. McCleary did not study parole work from the official perspective of the Department of Corrections. Therefore, when he showed his manuscript to a high-ranking representative in the department, his description of a department practice was called into question. McCleary describes his response:

“In my defense, I quoted two official Department of Corrections documents. The official replied: ‘You have to realize that the code book was written by lawyers and was meant to be read by lawyers. You don’t have a law degree, so you’re not qualified to read it. I can have a legal

opinion written up for you. Meanwhile, I suggest that you delete this passage' " (McCleary, 1978, p. 34).

McCleary neither received the legal opinion nor did he delete the passage. He adds, "I confess, however, that my descriptions of 'official' or 'formal' policy are the descriptions understood by lower-level actors and by myself from readings of official documents" (1978, p. 34).

Different Accounts of Effects: School Desegregation. A political scientist at the University of Chicago, Gary Orfield (1985), resigned from the Civil Rights Commission's evaluation of school desegregation because he was concerned that the evaluation almost exclusively addressed white parents' flight rather than black parents' wishes or their children's achievements. If white flight is the sole criterion, school desegregation has probably failed. If academic performance of black children constitutes the criterion, desegregation has succeeded (Stephan and Brigham, 1985). If children's self-esteem is the measure of success, the results are mixed (Gerard and Miller, 1975; Weber, Cook and Campbell, 1971). And if reducing white prejudice is important, desegregation looks promising (Brewer, 1985).

Using multiple criteria to judge the success of a program is like telling the story from multiple points of view. White flight is one possibly important criterion—if school busing were to drive all white families from a neighborhood, it would recreate the problem by requiring an ever wider radius to achieve desegregation. However, using white flight as the sole indicator is, as Orfield contends, prejudicing the evaluation unfavorably.

Different Accounts of Causes: Foster Placements. In deciding whether to remove a particular child from the custody of his mother, officials in one child welfare agency had to make a diagnosis: Was the mother a pyromaniac or a beleaguered tenant? Records in a state agency identified her as a pyromaniac who had set fire to her apartment and endangered her child. By the child's account his mother had repeatedly asked that the heating be repaired, but to no avail. One winter night, when his mother lit a fire in a wastebasket to provide heat, the basket tipped over and the fire spread (Bush and Gordon, 1978).

Actors and observers are particularly likely to differ in their identifications of causes (Jones and Nisbett, 1971; Kidder, 1982). Observers are likely to locate causes within the actor (pyromania) and actors are likely to locate causes in their surroundings (insufficient heat). The co-occurrence of multiple plausible explanations exemplifies the problems raised at the beginning of this chapter about the likelihood of reaching agreement among different methods of inquiry.

Rival Descriptions: The Perils of Qualitative Methods and Phenomenological Analyses

Trained as self-respecting experimentalists, we have elsewhere espoused the benefits of randomized experiments and quasi-experiments

for ruling out rival explanations (Saxe and Fine, 1981; Kidder and Judd, 1986). We still extoll the virtues of experimental design and analysis of selecting among rival causal hypotheses. There is another form of "rival explanation," however, that resists the solutions offered by experimental and quantitative analyses. The multiple points of view and competing stories constructed by actors who occupy different positions and hold different stakes cannot be reconciled or ruled implausible by quantitative analyses. It is this form of competition that qualitative research is most likely to unearth, precisely because it employs the language of the actors.

Field work differs from structured surveys and experiments not only in the extent to which hypotheses and procedures are predetermined but also in the extent to which causes, effects, and experiences are understood from the perspective of the participants. The data of field work contain the language of the actors, not only the language of the observers; the phenomena are selected, defined, and described by the participants. The analysis, therefore, is phenomenological, and it portends the co-existence of competing perspectives. Unlike the rival causal explanations of experiments or quasi-experiments (selection, history, maturation, testing), the competing explanations of phenomenological work cannot be ruled "implausible." They co-exist just as rival philosophies and political factions co-exist, without the means for declaring one or the other invalid.

Conclusion

Given the various meanings of triangulation, the question we posed at the outset has multiple answers. If we seek triangulation across measures, the likelihood of achieving agreement between quantitative and qualitative measures (small q) is as great as the likelihood of achieving agreement between two quantitative measures. The logic of the multitrait-multimethod matrix is applicable in such cases. The degree of correlation depends not only on the similarity of the methods but on the similarity of the presumed traits they are measuring. Two quantitative measures are not necessarily more alike than a quantitative and qualitative measure. Triangulation across measures is attainable and is not mysterious.

The more interesting and troubling problems arise when we seek triangulation of conclusions across studies. The same logic applies here as in the case of measurement: The likelihood of achieving agreement decreases as both the traits and the methods diverge. The problem is compounded, however, because the more divergent the methods are, the greater the likelihood that the traits or concepts under study will also diverge. Field work and other Qualitative methods that permit the researcher to generate and revise hypotheses en route also permit the concepts or traits to evolve as the research progresses. Therefore, although two Qualitative researchers might begin with the intention of investigat-

ing similar concepts and hypotheses, they are less likely to arrive at similar conclusions than two experimenters or survey researchers who begin with similar intentions.

Shotland and Mark (Chapter Five of this volume) have called for a systematic theory or set of guidelines for triangulation with multiple methods. To answer their call, we find it useful to distinguish among three different meanings of *triangulation* and two meanings of *qualitative*. The most ambitious version of triangulation arises when researchers try to reach similar conclusions across two or more studies that include at least one Qualitative (big Q) approach. Qualitative methods that permit the researcher to generate and revise hypotheses throughout the course of the study provide the greatest latitude for arriving at different rather than identical conclusions. Conclusions that differ from one another need not be contradictory or even rival explanations, however. The narratives and hypotheses that field workers generate represent the divergent perspectives of competing actors or stakeholders; they do not have the same status as rival explanations in causal models. Two or more groups of actors, stakeholders, or participant observers tell different stories in much the same way that two or more evaluation researchers reach different conclusions if they disagree about the criteria for defining "success" or "failure" (compare Guba, 1978.)

If any two studies focus on the same stakeholders, their stories are more likely to converge (see Mark and Shotland, 1985). The more structured the research methods are and the more predetermined the hypotheses and procedures, the greater the feasibility of arriving at the same conclusions. Field work and other methods with the Q writ large do not preclude triangulation but increase the chance of obtaining multiple rather than identical stories.

References

- Agar, M. *The Professional Stranger*. Orlando, Fla.: Academic Press, 1980.
- Brewer, M. B. "Experimental Research and Social Policy: Must It Be Rigor Versus Relevance?" *Journal of Social Issues*, 1985, 41, 159-176.
- Bush, M., and Gordon, A. C. "Client Choice and Bureaucratic Accountability: Possibilities for Responsiveness in a Social Welfare Bureaucracy." *Journal of Social Issues*, 1978, 34, 22-43.
- Campbell, D. T. "The Informant in Quantitative Research." *American Journal of Sociology*, 1955, 60, 339-342.
- Campbell, D. T., and O'Connell, E. J. "Methods as Diluting Trait Relationships Rather Than Adding Irrelevant Systematic Variance." In D. Brinberg and L. H. Kidder (eds.), *Forms of Validity in Research*. New Directions for Methodology of Social and Behavioral Science, no. 12. San Francisco: Jossey-Bass, 1982.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., and Weiner, S. S. *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*. San Francisco: Jossey-Bass, 1980.

- Crosby, F., and Crosby, T. L. "Psychohistory and Psychobiography." In S. Long (ed.), *Handbook of Political Behavior*. Vol. 1. New York: Plenum, 1981.
- Fine, M. "Why Urban Adolescents Drop into and out of Public High School." *Teachers' College Record*, 1986, 87, 393-409.
- Fiske, D. W. "Convergent-Discriminant Validation in Measurements and Research Strategies." In D. Brinberg and L. H. Kidder (eds.), *Forms of Validity in Research*. New Directions for Methodology of Social and Behavioral Science, no. 12. San Francisco: Jossey-Bass, 1982.
- Gerard, H. B., and Miller, N. *School Desegregation*. New York: Plenum, 1975.
- Glaser, B. G., and Strauss, A. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine, 1967.
- Goetz, J., and LeCompte, M. *Ethnography and Qualitative Design in Educational Research*. Orlando, Fla.: Academic Press, 1984.
- Guba, E. G. *Toward a Methodology of Naturalistic Inquiry in Educational Evaluation*. Center for the Study of Evaluation, Monograph Series in Evaluation. Los Angeles: Graduate School of Education, University of California, 1978.
- Jones, E. E., and Nisbett, R. E. *The Actor and the Observer: Divergent Perceptions of the Courses of Behavior*. Morristown, N.J.: General Learning Press, 1971.
- Kidder, L. H. "Becoming Hypnotized: How Skeptics Become Convinced." *Journal of Abnormal Psychology*, 1972, 80, 317-322.
- Kidder, L. H. "The Inadvertent Creation of a Neocolonial Culture: A Study of Western Sojourners in India." *International Journal of Intercultural Relations*, 1977, 1, 48-60.
- Kidder, L. H. "Qualitative Research and Quasi-Experimental Frameworks." In M. B. Brewer and B. E. Collins (eds.), *Scientific Inquiry and the Social Sciences: A Volume in Honor of Donald T. Campbell*. San Francisco: Jossey-Bass, 1981.
- Kidder, L. H. "Face Validity from Multiple Perspectives." In D. Brinberg and L. H. Kidder (eds.), *Forms of Validity in Research*. New Directions for Methodology of Social and Behavioral Science, no. 12. San Francisco: Jossey-Bass, 1982.
- Kidder, L. H., and Judd, C. M. *Research Methods in Social Relations*. New York: Holt, Rinehart & Winston, 1986.
- Komarovsky, M. *Blue Collar Marriage*. New York: Random House, 1967.
- Lukas, J. A. *Common Ground*. New York: Knopf, 1985.
- McCleary, R. *Dangerous Men*. Newbury Park, Calif.: Sage, 1978.
- Mark, M. M., and Shotland, R. L. "Stakeholder-Based Evaluation and Value Judgments." *Evaluation Review*, 1985, 9, 605-626.
- Nelson, D. E. "Investigative Journalism Methods in Educational Evaluation." In N. L. Smith (ed.), *Field Assessments of Innovative Evaluation Methods*. New Directions for Program Evaluation, no. 13. San Francisco: Jossey-Bass, 1982.
- Orfield, G. Office memorandum, University of Chicago, October 30, 1985.
- Reichardt, C. S., and Cook, T. D. "Beyond Qualitative Versus Quantitative Methods." In T. D. Cook and C. S. Reichardt (eds.), *Qualitative and Quantitative Methods in Evaluation Research*. Newbury Park, Calif.: Sage, 1979.
- Rubin, L. *Worlds of Pain*. New York: Basic Books, 1976.
- Runyan, S. E., and Seal, B. C. "A Comparison of Supervisors' Ratings While Observing a Language Remediation Session." *Clinical Supervisor*, 1985, 3, 61-75.
- Runyan, W. M. *Life Histories and Psychobiography: Explorations in Theory and Method*. New York: Oxford University Press, 1984.
- Saxe, L., and Fine, M. *Social Experiments: Methods for Design and Evaluation*. Newbury Park, Calif.: Sage, 1981.
- Smith, J. K., and Heshusius, L. "Closing Down the Conversation: The End of the Quantitative-Qualitative Debate Among Educational Inquiries." *Educational Researcher*, 1986, 15, 4-12.

- Stephan, W. G., and Brigham, J. C. (eds.). *Journal of Social Issues*, 1985, 41 (3).
- Stewart, A. J., and Healy, J. M., Jr. "The Role of Personality Development and Experience in Shaping Political Commitment: An Illustrative Case." *Journal of Social Issues*, 1986, 42, 11-32.
- Trend, M. G. "On the Reconciliation of Qualitative and Quantitative Analysis: A Case Study." *Human Organism*, 1978, 37, 345-354.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. *Unobtrusive Measures*. Skokie, Ill.: Rand McNally, 1966.
- Weber, S. J., Cook, T. D., and Campbell, D. T. "The Effects of School Integration on the Academic Self-Concept of Public School Children." Paper presented at the meeting of the Midwestern Psychological Association, Detroit, Mich.: 1971.
- Willems, E. P., and Rausch, H. L. *Naturalistic Viewpoints in Psychological Research*. New York: Holt, Rinehart & Winston, 1969.
- Zablocki, B. D. *The Joyful Community*. Baltimore, Md.: Penguin, 1971.

Louise H. Kidder is professor of psychology at Temple University. She is coauthor of Research Methods in Social Relations, Forms of Validity in Research, and "Making Sense of Injustice," in Redefining Social Problems. Her research interests include conceptions of social justice and the social psychology of gender.

Michelle Fine is associate professor of psychology in education at the University of Pennsylvania. Her research and social activism focus on issues of gender, race, and class equity in public schools.

