

Chapter from: 7

Scientific
Inquiry and the
Social Sciences

*A Volume in Honor of
Donald T. Campbell*

Marilynn B. Brewer

Barry E. Collins

Editors



Jossey-Bass Publishers

San Francisco • Washington • London • 1981

1. The first part of the document discusses the importance of maintaining accurate records of all transactions. This is essential for ensuring the integrity of the financial statements and for providing a clear audit trail. The records should be kept up-to-date and should be easily accessible to all relevant parties.

2. The second part of the document outlines the procedures for handling cash and other assets. It is important to ensure that all cash receipts are properly recorded and that all disbursements are supported by valid documentation.

3. The third part of the document discusses the process of reconciling bank statements with the company's records. This process is crucial for identifying any discrepancies and ensuring that the company's books are in balance.

4. The fourth part of the document addresses the issue of budgeting and cost control. It is important to establish a realistic budget and to monitor actual performance against that budget on a regular basis.

5. The fifth part of the document discusses the process of preparing financial statements. These statements should be prepared accurately and on a timely basis, and should be reviewed by management before being distributed to the board of directors.

6. The sixth part of the document outlines the procedures for handling payroll and other employee-related matters. It is important to ensure that all payroll transactions are properly recorded and that all employees are paid accurately and on time.

7. The seventh part of the document discusses the process of handling taxes. It is important to ensure that all tax obligations are properly calculated and paid on time, and that all tax records are properly maintained.

- Silverman, M. G. "Ambiguities and Disambiguities in Field Work." In S. T. Kimball and J. B. Watson (Eds.), *Crossing Cultural Boundaries: The Anthropological Experience*. San Francisco: Chandler, 1972.
- Trope, Y. "Inferences of Personal Characteristics on the Basis of Information Retrieved from One's Memory." *Journal of Personality and Social Psychology*, 1978, 36, 93-106.
- Tversky, A., and Kahneman, D. "Belief in the Law of Small Numbers." *Psychological Bulletin*, 1971, 76, 105-110.
- Uchendu, V. A. "A Navajo Community." In R. Naroll and R. Cohen (Eds.), *A Handbook of Method in Cultural Anthropology*. New York: Natural History Press/Doubleday, 1970.
- van Velsen, J. "The Extended-Case Method and Situational Analysis." In A. L. Epstein (Ed.), *The Craft of Social Anthropology*. London: Tavistock, 1967.
- Winch, P. "Understanding a Primitive Society." *American Philosophical Quarterly*, 1964, 1, 307-324.
- Witkowski, S. "Ethnographic Fieldwork: Optimal Versus Non-optimal Conditions." *Behavior Science Research*, 1978, 13, 245-253.

In M.B. Brewer & B.E. Collins (Eds.)
Scientific Inquiry and the Social Sciences: A
Festschrift Tribute to Donald T. Campbell. New
York: Jossey-Bass, 1981.

9

Louise H. Kidder

Qualitative Research and Quasi-Experimental Frameworks

Qualitative and quantitative research occupy different worlds, and Donald Campbell has educated us in both. His writings about degrees of freedom in the case study (Campbell, 1975), about the mutual methodological relevance of psychology and anthropology (Campbell, 1961), and about qualitative knowing in action research (Campbell, 1974) analyze qualitative research as a legitimate form of scientific knowing. His writing about experiments and quasi-experiments (Campbell and Stanley, 1966; Cook and Campbell, 1979) demonstrates his high standards for what constitutes science. While Campbell has informed us about both worlds, the two remain separate in his writings. This chapter is an attempt to bridge the gap between the two.

This is an immodest venture, because the division between qualitative and quantitative research is entrenched in several disciplines. In psychology it appears as a tension between experimental and clinical methods. In sociology it appears in the separation of fieldwork and statistical work. In the logic of scientific inquiry, it appears as a difference between hypothetico-deduction and analytic induction (see, for example, Becker, 1963; Glaser and Strauss, 1967; Lindesmith, 1968; McCall and Simmons, 1969; Znaniecki, 1934). The differences between qualitative and quantitative researchers appear also in caricatures; the former are considered "soft," and the latter "hard," or the former called "navel gazers" and the latter "number crunchers" (see Sherif, 1979). Without denying some of the real differences, this chapter explores the similarities. The exploration is one-sided; it applies the quantitative researcher's criteria for reliability and validity to qualitative work, using the typology that Campbell and his colleagues have developed (Campbell and Stanley, 1966; Cook and Campbell, 1979).

Not all qualitative research is as different from quantitative work as appears on the surface. Although field workers or participant observers do not use the language of quasi-experimental designs, we can discern designs in their studies. By studying naturally occurring phenomena, participant observers never become experimenters in the literal sense. They can seldom control what happens to whom, but they can *observe* what happens to whom when, and their work has an implicit quasi-experimental design. ("Participant observation" is here used in the broad sense, in which actual participation need not be involved.)

In the pages that follow, I will make explicit the quasi-experimental designs of several examples of qualitative research and assess their internal, external, construct, and statistical conclusion validity. I will apply the rules of one to the other, to answer the question posed by Zelditch (1962, p. 566): "Quantitative data are often thought of as 'hard,' and qualitative as 'real and deep'; thus, if you prefer 'hard' data you are for quantification and if you prefer 'real, deep' data you are for qualitative participant observation. What to do if you prefer data that are real, deep, *and* hard is not immediately apparent."

Causal Analysis in Qualitative Work

Internal validity as defined by Campbell (Campbell and Stanley, 1966; Cook and Campbell, 1979) is an issue only for research that reports causal relationships. Some qualitative work is purely descriptive and therefore not concerned with internal validity. The descriptive character of some fieldwork may contribute to its image as "soft," for it is not engaged in the difficult enterprise of causal analysis. Some qualitative researchers *do* make causal assertions, however, handicapped though they are without experimental designs.

Whenever qualitative research presents a career analysis—the career of a marijuana user, an embezzler, a parole officer—it includes causal assertions. Careers are socialization patterns or paths, and the steps necessary to become socialized are links in a causal chain. Participant observation rarely contains only one or two causal statements of the "A causes B" variety. Instead, it contains a set of propositions about the causal chain. The following are examples of causal statements from participant observation research:

Instead of the deviant motives leading to the deviant behavior, it is the other way around; the deviant behavior in time produces the deviant motivation. Vague impulses and desires—in this case, probably most frequently a curiosity about the kind of experience the drug will produce—are transformed into definite patterns of action through the social interpretation of a physical experience [Becker, 1963, p. 42].

Routine career contingencies reward parole officers for underreporting deviant behavior. *Ceteris paribus*, parole officers will ignore most of the crimes, incidents, and violations they observe in their case-loads. The exceptions to this rule are situations where the parole officer realizes some benefit from reporting an incident. . . . Ordinarily, the parole officer will report an incident and thereby create a record only when he can use that record to enhance his work environment [McCleary, 1977, p. 576].

Many research students and professionals are reluctant to say "this caused that" when they report the results of their work. Instead, they claim that theirs is not causal research but a description of a process, or an exploration of a phenomenon, or a report of a relationship. For instance, I did not say that my own participant observation study of sojourners in India was a causal analysis; instead, I claimed to "describe the socialization of aliens into their role" (Kidder, 1977, p. 48). In fact, I wished to say that being in India for a long time and occupying a high status caused sojourners to become more alien rather than more Indian. Without a true experiment, however, I did not make such a bold claim. This same problem besets quantitative researchers with nonexperimental designs. In the same study, I had quantitative data on attitudes, and I did not make causal assertions about those data either. Instead, I reported a relationship: "Favorability toward the host country . . . [was] examined in relation to the respondent's occupational status and elapsed time in India" (Kidder, 1977, p. 48). Reports of relationships and descriptions of processes are retreats to a safe position, a position that cannot be attacked because the writer does not claim to know what caused what.

There are many reasons for the reluctance to be so direct as to assert "this caused that," including measurement error and bias. But one problem plagues qualitative research in particular, because qualitative researchers work inductively rather than deductively. They do not begin with a hypothesis; instead, they generate hypotheses from their data. Analytic induction proceeds in the direction opposite from hypothetico-deduction. Hypothetico-deduction researchers work from the top down; they begin with theoretical premises, predict a pattern of results, and examine the data to test the theoretical deduction. Analytic inductive researchers work from the bottom up—beginning with data and developing theoretical categories, concepts, and propositions from the data (Glaser and Strauss, 1967). By forming hypotheses to fit the data, qualitative researchers violate some of the assumptions of statistical testing in quantitative research, and this too may inhibit them from making claims about causes.

If we listen carefully to the conclusions of participant observation research, however, we hear causal assertions even if they are not as simply stated as "this caused that." In his study of how

parole officers report violations, McCleary (1978, pp. 132-135) listed five career contingencies that *cause* parole officers to ignore crimes: (1) Full reporting cuts into parole officers' "free" time. (2) Full reporting places parole officers in jeopardy because they may have to defend their decisions in a hearing. (3) Parole officers believe their performance is evaluated on the basis of the good behavior of their parolees. (4) Full reporting creates "busy work" for parole officers. (5) Parole officers limit their ability to counsel or give a man a "second chance" when they report violations.

In his study of marijuana users, Becker (1963, p. 49) also made causal assertions and identified a series of events that must take place if a person is to use marijuana for pleasure: (1) learning the technique of smoking to produce effects, (2) learning to perceive the effects, (3) learning to enjoy the effects.

And in his study of embezzlers, Cressey (1953) identified four conditions that must be present if embezzlement is to occur. The person must (1) be in a position of financial trust, (2) have a nonshareable financial problem, (3) recognize embezzlement as a possible solution for the problem, (4) develop a way to rationalize embezzlement without calling it that, so that it seems like a "loan" or a justifiable use of someone else's money. Cressey found no case of embezzlement that did not have all these elements, and he concluded that these four conditions as a set are necessary and sufficient causes of embezzlement.

The five career contingencies that McCleary listed as reasons for underreporting crimes, the three steps that Becker listed as necessary for becoming a marijuana user, and the four conditions Cressey listed as necessary for embezzlement to occur are all causes. In each case, the researchers have said "This leads to that." To make such definite statements about causes and effects, they must each have ruled out rival explanations. None of these studies was a true experiment, because the researchers did not control what happened to whom under conditions of random assignment. Instead, they each recorded what happened, to whom, when, and created implicit quasi-experimental designs.

Quasi-Experimental Designs in Participant Observation

Both Cressey and Becker confined their studies to people who exhibited the phenomenon in question—embezzlement in one

case and marijuana use in the other. They did not study populations of nonembezzlers and nonusers of marijuana, but they interviewed users and embezzlers about previous times in their lives when they had not indulged in marijuana use or embezzlement. They formulated hypotheses about the conditions necessary for marijuana use or embezzlement to occur by noting the absence of those conditions during times of nonuse or nonembezzlement. Cressey and Becker were able to study the full range of the phenomenon—embezzlement and nonembezzlement of funds, use and nonuse of marijuana—by obtaining retrospective reports from their samples of users and embezzlers. They could have reported their results in a cross-tabulation or covariance table like that in Table 1.

Table 1. Phenomenon (Embezzlement or Marijuana Use)

		<i>Present</i>	<i>Absent</i>
Prior Conditions	Present	100% of the cases	0 cases
	Absent	0 cases	100% of the cases

The right-hand column is a phantom; it contains retrospective reports of people who currently exhibit the phenomenon, not interviews with people who have never embezzled or used marijuana for pleasure. It is a within-subjects variable, based on recall. Had it been possible for either Cressey or Becker to be there before some of the convicts had embezzled or before some marijuana users learned to perceive the effects, they could have gathered on-site observations of the absence of those conditions. They relied on retrospective reports as a substitute. They did not interview samples of nonembezzlers or nonusers about why they had not indulged, because such interviews would yield reasons that have little bearing on the propositions they were testing. For instance, people may say they have never used marijuana for pleasure because they have not had the opportunity or because they consider it immoral or illegal (Becker, 1963). Such interviews would give information about the causes for not becoming a marijuana user, which would be a separate study.

We can depict both Becker's study of marijuana users and

Cressey's study of embezzlers as time series designs. We borrow the following notation from Campbell and Stanley (1966):

X = treatment or cause
 O = observation or effect

The causes are the necessary conditions for the marijuana use or embezzlement to occur. The effects are the reports that marijuana use or embezzlement did or did not occur. There are two differences between a traditional time-series design and the design of these two studies: (1) the X 's are not single causes but a combination of necessary conditions, and (2) the X 's and O 's are measured through retrospective self-reports rather than the investigators' on-site observations or archives.

In his analysis of what causes people to become marijuana users, Becker identified three steps or causes that work in sequence. The three steps combined we will represent as follows: X_1 - X_2 - X_3 . The first step, X_1 , is learning to smoke "properly" to produce physical symptoms. This is not sufficient to cause marijuana use alone, however, because the physical symptoms are not always obvious. The novice must learn to detect them. Having cold hands or feet, developing an intense hunger, and having fits of laughter are signs known to initiates, but novices often need to have their own symptoms pointed out before they recognize them as signs of being high. This is the second step, X_2 . Having learned to produce and perceive the effects, a novice becomes a continual user only if he or she also learns to enjoy those effects. The effects are not obviously pleasant, as the following example shows: "A new user had her first experience of the effects of marijuana and became frightened and hysterical. She 'felt like she was half in and half out of the room,' and experienced a number of alarming physical symptoms" (Becker, 1963, p. 56). Experienced users help novices reinterpret such feelings as pleasant, exciting, and sought after. This is the final step, X_3 .

Becker's evidence consisted of retrospective reports from people who had tried smoking and either continued or quit as a result of the presence or absence of conditions X_1 , X_2 , and X_3 . When people first begin smoking, they try it a second and third time, even if they do not experience all the conditions (X_1 - X_2 - X_3) during their first try, either because their friends encourage them

or because they believe there is more to the experience than what they felt. For instance, the following report describes a progressive accumulation of experience until all three conditions were met:

"I didn't get high the first time. . . . The second time I wasn't sure, and he [smoking companion] told me, like I asked him for some of the symptoms or something, how would I know, you know. . . . So he told me to sit on a stool. I sat on—I think I sat on a bar stool—and he said, 'Let your feet hang,' and then when I got down my feet were real cold, you know.

"And I started feeling it, you know. That was the first time. And then about a week after that, sometime pretty close to it, I really got on. That was the first time I got on a big laughing kick, you know. Then I really knew I was on" [Becker, 1963, pp. 49-50].

We can diagram this report as a time series, in which the effect (becoming a regular user) appeared only after all three necessary conditions were present:

<i>O</i>	X_1	<i>O</i>	X_2-X_2	<i>O</i>	$X_1-X_2-X_3$	<i>O</i>
not regular user	smoked, felt no effects	not regular user	smoked, felt effects	not regular user	smoked, felt and enjoyed effects	became regular user

As evidence that all three conditions must be present for a person to become a regular user, Becker reports cases where one or more conditions were absent and a person did not become a marijuana user. For instance:

"It was offered to me and I tried it. I'll tell you one thing. I never did enjoy it at all. I mean, it was just nothing that I could enjoy. [Well, did you get high when you turned on?] Oh, yeah, I got definite feelings from it. But I didn't enjoy them. I mean I got plenty of reactions, but they were mostly reactions of fear" [Becker, 1963, p. 54].

In another instance, a person who had been a regular user quit because he had some unpleasant and frightening experiences which made him discontinue use for three years:

“It was too much, like I only made about four tokes, and I couldn’t even get it out of my mouth, I was so high, and I got real flipped . . . I walked outside, and it was five below zero, and I thought I was dying . . . I fainted behind a bush. I don’t know how long I laid there . . . all weekend I started flipping, seeing things there and going through hell, you know, all kinds of abnormal things . . . I just quit for a long time then” [Becker, 1963, p. 57].

If we draw the time series for this person, it includes the onset, offset, and onset of marijuana use, as follows:

O O O O	$X_1-X_2-X_3$	O	$X_1-X_2-X_3$	O	X_1-X_2	O	X_1-X_2	O
nonuse		onset and use		offset and nonuse				
		$X_1-X_2-X_3$						
		onset						

Not all participant observation relies on retrospective reports alone to deduce the pattern of the career. McCleary, in his study of the socialization of parole officers, write field notes while he observed parole officers. He used both his own observations of their behavior and their retrospective reports of what happened earlier. His work provides a useful illustration of how qualitative researchers can rule out threats to internal validity.

Threats to Internal Validity in Qualitative Research

McCleary’s research on parole officers describes both the causes for underreporting and the causes for full reporting among parole officers. Parole officers do not always ignore crimes or violations by their parolees. If parole officers can benefit from reporting incidents, they do so.

McCleary (1977, p. 576) makes a straightforward causal assertion: “Routine career contingencies reward parole officers for underreporting deviant behavior. . . . The exceptions to this rule

are situations where the parole officer (PO) realizes some benefit from reporting an incident."

In addition to the five causes of underreporting listed earlier, McCleary identified three causes or occasions for full reporting:

1. *To threaten a parolee.* Parolees know their dossiers can be used against them, so a PO may threaten to report an incident to control a particular parolee.
2. *To get rid of a parolee.* By reporting incidents that show a parolee is in need of special treatment, a PO can have that parolee transferred from his caseload to a special treatment program.
3. *To protect himself.* A PO can write full reports to demonstrate he is doing his job, following the rules, and keeping track of "dangerous men."

To diagram the implicit time-series designs for McCleary's study, we adopt the following notation:

- O = observations of parole officers' reports of incidents
 X_U = routine career contingencies that reward underreporting
 X_F = special conditions that reward full reporting

Some of the observations of full reporting of incidents were made directly by McCleary. Some of the observations were told to him by parole officers. Since full reporting was the exception to the rule, McCleary observed fewer of those instances and had to rely on informants. For instance, when two newspapers began an investigation of the department of corrections, parole officers panicked and began putting their files in order to protect themselves and their supervisors. One parole officer told McCleary: "That's the only time I ever saw all nine POs in the office at the same time. We wrote a lot of paper that week. When the story hit, we were all protected" (McCleary, 1978, p. 145).

The research, therefore, contains a mixture of McCleary's direct observations of X 's and O 's and indirect observations given to him by his respondents. For instance, a time series for a new recruit may look like the following:

O O O X_UO X_UO X_UO

The initial *O*'s are stories about that recruit rather than McCleary's firsthand observations. He was told that this new PO first wrote overlong reports and had to learn to underreport: "His first reports give vivid descriptions of the 'personal, social, financial, family, employment, and psychological problems' of the clients he has contacted. He is quite proud of these reports, so he is shocked when his supervisor hands them back to him. All are rejected. All require 'further investigation' or 'more work'" (McCleary, 1978, p. 159). Contrary to the novice's belief that his reports were too brief, he gradually learned that they were too long. He learned that if he wanted to make statements about a client's financial problems, he had to back up his claims by contacting the client's employer, family, friends, and landlord. "Each sentence of a three-page report will require ten pages of corroboration. . . . When he seeks his trainer's help with the reports, he is advised to keep his reports brief. . . . The novice knows that DC [department of corrections] rules and regulations call for more detailed reports and he points this out to his trainer. The trainer counters this argument by pointing out the many ways in which reports can be misinterpreted by DC officials and by other criminal justice agencies" (McCleary, 1978, p. 159). Full reporting means extra work and perhaps extra trouble for parole officers.

Sometimes the extra trouble is worth it, however, and parole officers report incidents to coerce a parolee, to get rid of a troublesome parolee, or to protect themselves in case of investigations. Since full reporting is the exception rather than the rule, McCleary does not describe any long series of incidents that led to a parole officer's learning to write full reports. Instead, he presents one-shot incidents and quick lessons that are short or truncated time series. For instance, an office supervisor said:

"There's something final about the written word. A PO can talk himself blue in the face and get no response from the parolee. If he puts it down on paper, though, he gets results. Parolees know that once something gets written down, it's final. I encourage my POs to utilize their reports that way" [McCleary, 1978, p. 140].

This and the following description from a parole officer summarize what may be years of experience. They condense a series of X_{FO} X_{FO} X_{FO} into a single report of X_{FO} :

“One time I had this junkie—a white kid. His mother was causing problems for him. She didn’t have any respect for me or for the job I was trying to do. Well, I finally gave him an official warning and mailed a copy to his mother. I never had any trouble with him after that” [McCleary, 1978, p. 139].

Unlike Becker’s analysis of the causes of becoming a marijuana user, McCleary’s analysis does not specify a sequence of conditions. The five causes of underreporting need not all occur together or in a particular order. They operate singly or in combination. For this reason, I have simply used the X_U to symbolize any combination of causes for underreporting and X_F to symbolize any combination of causes of full reporting. The time-series lines shown above have different lengths, and the observations (either direct or indirect) are not evenly spaced. Some observations or stories about individual parole officers are so brief that they qualify only as one-shot observations rather than actual time series, as in the preceding example, which quotes a parole officer as saying that he never again had trouble after he sent an official warning to a parolee, with a copy mailed to his mother. No one of the lines is as regular and elegant as the archival data described by Campbell in his discussions of time series (Campbell and Stanley, 1966; Cook and Campbell, 1979). This is because McCleary did not “design” his research to take the shape of a time-series design. He did not know what the causes or X ’s were before he began. He discovered them through analytic induction, so the design is an *ex post facto* or retrospective creation. Nonetheless, we can decipher the design that emerges, and we can examine the validity of the causal assertions by seeing whether there are plausible rival explanations. The list of rival explanations that follows is taken from Campbell and Stanley (1966) and Cook and Campbell’s (1979) lists of threats to internal validity.

- *Selection.* We need not concern ourselves with this because McCleary did not try to compare one group of parole officers

with another. He traced the careers of individuals and recorded how they changed, so preexisting selection factors make no difference in his research. Selection is no threat to time-series designs.

- *History.* Is it plausible that some historical quirk was the cause of underreporting? Could something like a change in the police chief, or police department regulations, or mayor, or criminal justice regulations or legislation (for example, in a 1971 decision in the case of *Morrissey v. Brewer*, the Supreme Court declared that agencies must protect the due process rights of parolees) explain the underreporting? None of these are plausible explanations, because McCleary tracked the parole officers' behaviors over time and found that each one had to learn to write brief reports. Moreover, their shortened reports did not appear simultaneously, so no single event precipitated a sudden drop in reporting. History does not explain the underreporting.
- *Maturation.* As POs became tired, fatigued, or sophisticated, did they underreport incidents? There was no evidence that fatigue or boredom alone caused underreporting, because POs would immediately give *full* reports when the contingencies changed, no matter how fatigued.
- *Testing.* Did McCleary's observations sensitize the POs and make them change their reporting? If they were aware of what he was observing, they might have done and reported *more* paperwork rather than less, since being observed might have caused them to feel evaluated. Thus, McCleary's presence was not a likely cause of underreporting; more important were the threatened checks by the department of corrections, which led to a flurry of report writing and office attendance. Becker and Geer (1957) compare the demand characteristics of participant observation and isolated interviews and point out that, because the interview is a unique situation, interviews are more likely to produce reactive effects and effects of testing. In the isolation and anonymity of an interview, respondents are free from the constraints of social intercourse. Interviewers hope that respondents will tell the "truth" under such conditions; but respondents are also free to lie or respond to the demand characteristics of the interview (see Rosenthal and Rosnow, 1969). In an ongoing social setting such as a parole office, the daily constraints of the job, the so-

cial group, and the expectations of one's associates are greater than the constraints of being observed, so testing is less likely to influence a respondent's subsequent behavior.

- *Instrumentation.* Did McCleary change his way of perceiving or reporting what was going on? Was the change in the observer rather than the observed? This would be a plausible threat if McCleary had observed one cohort of parole officers, all of whom changed at the same time. Had he observed full reporting at the beginning of this research and underreporting at the end (or vice versa), instrumentation would be a rival explanation because changes in his research procedures could have coincided with the purported changes in the parole officers. Two features of McCleary's research make instrumentation implausible as a rival explanation. The first is that the POs were not a cohort whom McCleary observed from beginning to end. They were an assorted group who had been in the profession for varying lengths of time; POs entered and dropped out during the course of his observations. The second feature is that much of McCleary's data of how POs learned to underreport came not from his direct observations but from retrospective reports by other observers—supervisors and the POs themselves. Both of these features desynchronized changes in the parole officers and changes in the researcher or his instruments, making instrumentation an implausible rival explanation.
- *Regression.* Were the POs observed by McCleary an extreme group of prolific report writers whose behavior could change in only one direction? If anything, the nature of a PO's job attracts people who have another occupation, such as operating a restaurant or studying for a master's degree. They seek parole work because it does not require full-time office attendance, and it is unlikely that they would want to fill their spare time with unnecessary paperwork. They are, therefore, not likely to be compulsive report writers who can only become less so. There is also no indication that those POs whom McCleary observed or heard about were an unusual sample of officers. No one ever described himself or another as being atypical. We can be reasonably confident, therefore, that neither POs as a whole nor those whom McCleary observed represent an extreme sample whose behavior could only regress toward the mean.

- *Mortality*. Do those POs who insist on full reporting drop off the force? This is possible, but McCleary has no evidence that noncompliers left. He does have evidence that an original non-complier changed because of social pressure:

A new PO in one of the branch offices often wrote lengthy site investigation reports. The other POs in the office disapproved of this, and from time to time would make jokes at the nonconformist's expense. The jokes were meant to be a norm-enforcing mechanism, but as the new PO was not aware of the norm for site investigation reports, the jokes were interpreted as a personal attack: "Whitney called me Sally Social Worker. I don't think that's funny and I don't think he means it as a joke either. He only says it when the other fellows are around. I'm just trying to fit in here. . . . My site investigation reports do have something to do with it but I don't think that it's anybody's business. I'm not saying that anybody has to do it my way. I just happen to think that the site investigation is the most important part of this job. It's a free country, though, and if other people have a different opinion, that's their right. I don't tell them how to do their work and I don't want them telling me how to do mine" [McCleary, 1978, p. 58].

Within a few weeks this PO became "one of the fellows," as a result of both the group's ridicule and his supervisor's refusal to accept long reports.

No participant observation study has ever included an explicit discussion of these threats to internal validity, but this exercise demonstrates how it could be done. In fact, if the researchers themselves were asked to rule out the threats to validity, they could do so even more convincingly, because they have much more data to draw on than appears in the published report. What makes it possible for participant observers to rule out threats to validity in the absence of an explicit design is the richness of the data, the longitudinal observations, and the nonsimultaneity of treatments across persons. The researcher has many pieces of in-

formation about each person or incident. Much of this information is gathered without the researcher's foreknowledge of how useful it will be. One of the few rules for participant observation is "write down everything," and it is this which helps the researcher rule out rival explanations and check hypotheses that are formulated after the fact.

Cressey formulated and revised his hypothesis five times before he arrived at his conclusion about the causes of embezzlement. Each time he formulated a new hypothesis, he checked it against not only new interviews but also all of his previously recorded interviews and observations. This *ex post facto* procedure is a necessary practice in participant observation. It forms the basis for analytic induction and negative case analysis. Negative case analysis requires that the researcher look for disconfirming data in both past and future observations. A single negative case is enough to require the investigator to revise a hypothesis. When there are no more negative cases, the researcher stops revising the hypothesis and says with confidence "This caused that."

This process of revising hypotheses with hindsight which gives the qualitative researcher confidence in his or her findings is the same process that gives a quantitative researcher doubts. It threatens the statistical conclusion validity of findings. Although qualitative researchers do not use statistics and therefore would not discuss statistical conclusion validity, I will examine the procedure from the point of view of a quantitative researcher, because negative case analysis still remains suspect from a hypothetico-deductive standpoint.

Statistical Conclusion Validity and Negative Case Analysis

Cressey (1953) reports in detail how he used negative case analysis to formulate and revise his hypothesis five times: "The first hypothesis . . . was that positions of financial trust are violated when the incumbent has learned in connection with the business or profession in which he is employed that some forms of trust violation are merely technical violations and are not really 'illegal' or 'wrong,' and, on the negative side, that they are not violated if this kind of definition of behavior has not been learned" (p. 27).

He formulated this hypothesis on the basis of other literature and research on white-collar crime. After interviewing only a few inmates, however, he learned that they knew all along that embezzling was illegal. He revised his hypothesis, therefore, to read: "Positions of trust are violated when the incumbent defines a need for extra funds or extended use of property as an 'emergency' which cannot be met by legal means" (p. 27).

Some embezzlers whom he subsequently interviewed said theirs had been an emergency, but others said they had taken the money even without a financial "emergency." Still others said there had been financial emergencies earlier in their lives, when they had not embezzled. Both of these groups contradicted the second hypothesis, so Cressey developed a third: "It shifted the emphasis from emergency to psychological isolation, stating that persons become trust violators when they conceive of themselves as having incurred financial obligations which are . . . nonsocially sanctionable and which . . . must be satisfied by a private means" (p. 28). He checked this hypothesis against both subsequent and previous interviews. These checks revealed that "in a few of them there was nothing which could be considered as a financial *obligation*, that is, as a debt which had been incurred in the past for which the person at the present time felt responsible. Also, in some cases there had been nonsanctionable obligations at a prior time, and these . . . had not been alleviated by means of trust violations" (p. 28).

Cressey revised his hypothesis again. The fourth version differed by "emphasizing this time not financial obligations . . . but nonshareable *problems* . . . that is, . . . the subject could be in financial difficulty not only because of an acknowledged responsibility for past debts, but because of present discordance between his income and expenditures as well" (p. 29). The fourth revision accounted for men who had not developed debts but who had been living above their means and had been afraid to admit this to their families or friends. It also included some who had been maintaining separate households without telling their family or friends. Again, however, there were exceptions: men who said they had experienced the nonshareable problem for a long time before they embezzled. "Some stated that they did not violate the trust at the earlier period because the situation was not in sharp

enough focus to 'break down their ideas of right and wrong' " (p. 30).

This led Cressey to the final revision: "Trusted persons become trust violators when they conceive of themselves as having a financial problem which is nonshareable, are aware that this problem can be secretly resolved by violation of the position of financial trust, and are able to apply to their own conduct in that situation verbalizations which enable them to adjust their conceptions of themselves as users of the entrusted funds or property" (p. 30). Cressey tested this hypothesis against all the data he had gathered, against two hundred cases of embezzlement collected by another researcher, and against additional interviews that he conducted in another penitentiary. He found no negative cases.

Negative case analysis produces the perfect pattern of results shown in Table 1. All four conditions necessary for embezzlement were present in 100 percent of the cases where embezzlement occurred, and one or more conditions were absent whenever embezzlement did not occur. The method ensures a perfect correlation, because the causal hypotheses are continually revised until they fit every case. This is precisely what researchers using the hypothetico-deductive approach consider illegitimate. If the hypotheses are revised to fit the data, the researcher capitalizes on chance, and the probability levels associated with the statistics are meaningless.

The conclusions of participant observation do not depend on comparing a signal-to-noise ratio, as do the conclusions of quantitative research. They are more like the conclusions of some operant conditioning research, which depend not on statistics but on graphic illustrations that demonstrate that a reinforcer works or does not work (Hersen and Barlow, 1976). When we use negative case analysis, we continue until there are no outliers or exceptions to the rule. This is very different from quantitative analysis, which incorporates outliers or exceptions to the rule in the random error term. When we use statistical analysis, we assume there will be error variance. Statistical tests are necessary when the ratio of explained variance to error variance is not obviously great. When the ratio is big enough so that the distributions do not overlap and the difference between treatment and no-treatment conditions can be seen with the naked eye, statistical tests are super-

fluous. This is the case with qualitative analysis, because there are no outliers, no random error variance. Negative case analysis eliminates all exceptions by revising hypotheses until all the data fit.

No statistical tests are necessary once negative cases are removed, because the data clearly fit the hypotheses. The obvious fit between hypotheses and data can make qualitative research seem either very good or very bad. When the conclusions are so clear that the reader says to him or herself, "Of course, now I see it," the work may appear to be true and insightful. When the conclusions are clear and the reader says, "Of course, it's obvious," the work sometimes appears trivial, because it tells the reader nothing new. The same is true of quantitative research. When the statistical tests are significant and the effects large, the findings may be either very clear and exciting, or very clear, obvious, and uninteresting.

When statistical tests are marginally significant and the effects are small, quantitative work may be interesting but is often unconvincing without additional evidence. Similarly, when the data for qualitative analysis are sparse, and there is not abundant evidence to support the hypothesis, even though there are no negative instances, the conclusions are weak and further evidence is needed.

Therefore, even though inductive analysis violates the assumptions of statistical testing by forming hypotheses to fit the data rather than finding data to test hypotheses, the consequences are not so different. When qualitative research has abundant evidence, it rings true, the conclusions are obvious, and they may be either interesting or trivial. When qualitative research has sparse evidence, it is not persuasive, even though there are no negative cases. Abundant evidence in qualitative research results when one has made many observations and recorded many instances. This is the equivalent of having a large N . The larger the N , the more convincing the conclusions in either case.

Negative case analysis replaces statistical analysis in qualitative research. Both are means to handle error variance. Qualitative analysis uses "errors" to revise the hypothesis; quantitative analysis uses error variance to test the hypothesis, demonstrating how large the treatment effects are compared to the error variance.

Construct Validity and Reliability in Qualitative Research

Construct validity is the appropriate naming of a variable, be it a cause or an effect (Meehl, 1977; Cook and Campbell, 1979). That is, the measurement of a variable must correlate with some other measurement of the same variable, and the theory underlying the variable must be correct. Campbell and Fiske (1959) introduced the multitrait-multimethod matrix as a tool for assessing construct validity. The matrix includes some correlations that should be low in addition to others that should be high as evidence of construct validity. One variable should have low or near zero correlations with another variable which is presumed to be different from the first, even though they are measured by similar methods; and it should have high correlations with other measures of itself, even though the methods of measurement are maximally different. In both cases, the validity rests on correlations between measures which differ in either the method or the trait. Reliability, on the other hand, is the consistency or similarity or replicability of observations. To assess reliability, we make repeated measurements of the same trait with similar or identical methods. Validity and reliability, therefore, represent opposite ends of a continuum of measurement, as shown in Table 2.

Table 2. The Validity-Reliability Continuum

<i>Reliability</i>	<i>Validity</i>
Maximally similar methods of measurement	Maximally different methods of measurement

Reliability coefficients indicate how much agreement there is between maximally similar methods of measuring a concept. Validity coefficients indicate how much agreement there is between maximally dissimilar methods of measuring the same construct.

Reliability and validity, broadly construed, are requirements for both quantitative and qualitative research, but the technology for assessing them has been developed primarily in quantitative research on individual differences. Calculating reliability and validity *coefficients* requires quantifying observations. Deciding whether concepts or observations are reliable or valid does not necessarily

require quantification, however. Qualitative researchers make such decisions regularly, without the explicit calculations of reliability and validity coefficients. Mills (1951), in his book on white-collar workers, illustrated how qualitative researchers assess the reliability and validity of their observations in the absence of quantified measures. He developed the argument that clerical workers experience "status panic" as a result of the ambiguous position of their work. As a result, they "conceal the nature of their own work, and borrow prestige from the firm or industry, by identifying themselves with such phrases as 'I am with Saks,' or 'I work at Time.' They saved up their salaries and spent them for an evening at expensive places of entertainment, or for a vacation at a costly resort, in order to 'buy a feeling, even if only for a short time, of higher status.' A salesgirl dealing with 'Park Avenue' customers will try to behave with greater dignity and distinction in her off-the-job contacts than the girl who works on 34th Street" (Barton and Lazarsfeld, 1969, p. 242).

These repeated observations of what Mills called "status panic" lie somewhere on the continuum between reliability and validity checks. They are like reliability replications in that they are observations made by the same observer; they are like validity confirmations in that they are different manifestations of the same construct. Mills gathered many instances of status panic, not to demonstrate how frequent it is but to "demonstrate that the variable . . . does exist, that it exists in a number of manifestations, and that some of these manifestations have been observed among white-collar workers of various sorts" (McCall and Simmons, 1969, p. 244).

Repeated observations of a variable or concept in participant observation do not constitute frequency distributions. If something was recorded ten times, we cannot assume that the number 10 represents the frequency with which that event occurs in the population, because a participant observer normally does not select a random sample of people, times, or locations and obtain data from each unit in that sample to ascertain frequencies. Instead, the ten observations demonstrate that the variable exists. When participant observers list repeated instances of an event, the list serves as a reliability check. It shows that the event (and the variable it represents) occurs and that the concept is not based on

a chance observation. For instance, in my research on a hypnosis workshop, I argued that the hypnotists rewarded people who behaved like good hypnotic subjects and punished those who resisted. As evidence I listed the following statements made by two different hypnotists at different times during the workshop:

(a) "Sometimes the *very analytic* person won't go along—he thinks to himself 'The hypnotist said I'm going to go to sleep, but hell, I'm not asleep, I'm wide awake, and he's crazy!'"

(b) "The most difficult one is the compulsive. . . . He also has the constant doubt 'Am I or am I not in trance?' Or the terribly rational person who cannot allow himself to go into any fantasy would be terribly difficult."

(c) "There is a myth about hysterics—that they are better subjects. But really, *normal* persons are better subjects."

(d) "People who can throw themselves into art, nature, sports—involve themselves in a role and relinquish reality orientation for a time—these are the people who make good hypnotic subjects" [Kidder, 1972, p. 319].

These statements, and others, were made in response to questions from people in the workshop who said they did not feel they had been in a trance. They wondered whether it was "normal" to have difficulty going into a hypnotic trance. Rather than assure them that they were normal, the hypnotists countered by saying that they might be compulsive, very analytic, or terribly rational. Normal people are *better* subjects, said the hypnotists. The messages in these responses were clear to the questioners, all of whom were psychologists, "and given the public nature of the situation, no one could have ignored such rewards and punishments meted out to persons who became hypnotized and others who refused" (Kidder, 1972, p. 319).

The repeated observations of how people were rewarded for behaving like hypnotized subjects and punished for noncompliance can be regarded either as reliability or validity checks or as some combination of the two. The continuum between reliability

and validity shown in Table 2 is both a conceptual and a procedural continuum. Therefore, even in quantitative analysis, the distinction between reliability and validity becomes blurred when the methods of measurement are neither identical nor maximally dissimilar. Both technically and conceptually, reliability and validity are even less distinct in qualitative work. Nonetheless, multiple observations of measurements of the same phenomenon are important in analytic induction, because each observation provides further evidence for the argument.

One reason that qualitative researchers are less concerned with the precise form of measurement—and therefore cannot distinguish between reliability and validity checks—is that they are concerned more with the content than the form of the observation. What matters is that each additional piece of evidence is *consistent* with the other observations and not that each observation is *identical*. This is an important difference between quantitative and qualitative procedures. Reliability in fieldwork lies in an observation's not being contradicted and proved wrong rather than its being repeated in detail.

What I have said about reliability and validity in measurement in fieldwork is also true of the reliability and validity of conclusions. They rest on not being contradicted rather than on being repeated. As a graduate student in social psychology at Northwestern, I occasionally heard Donald Campbell and Howard Becker disagree over the virtues of exact replication. The problem, as Campbell defined it, was that field workers are not interested in replicating results; each would rather make his or her own unique discovery. And even if one field worker tried to replicate another's study, the attempt would probably fail, because the second researcher would probably discover something new and pursue that rather than try faithfully to reproduce the original observations. Becker considered Campbell's description accurate, but said it presents no problem because the reliability of fieldwork lies in its not being contradicted rather than its being repeated in detail. Instead of looking for the same phenomenon, one will do better to look for new information that may be *consistent* or *inconsistent* with the first. This is true for reliability across studies as well as within a single study. Therefore, rather than test-retest reliability or research-replication reliability, qualitative research calls for

something akin to an internal consistency measure of reliability. What matters in each instance is that there be no negative or inconsistent evidence. Negative case analysis exemplifies the procedure: The researcher "searches for negative or qualifying, as well as for supporting, instances. It has been said that science turns upon negative evidence," writes Strauss (see Strauss, Schatzman, Bucher, Ehrlich, and Sabshin, 1964, p. 21); and that is the rationale for the approach to reliability described above.

External Validity

Conclusions that are externally valid can be replicated across other persons, times, places, and operationalizations of the treatment and effects. The external validity of experiments and quantitative research depends on the researcher's demonstration that the same results can be found in other laboratories, or with another sample of subjects, or with other measures of the same variables. The external validity of qualitative research depends on the researcher's demonstration that similar results occur in other settings. In neither case does the researcher intend to make statements only about the subjects or observations made in that particular study.

Psychological experiments with college sophomores are never intended to be studies of college sophomores; they are intended to be studies about people in general. The same is true of qualitative research. McCleary did not intend his study of parole officers in Chicago to be read as an analysis that is true only of parole officers in Chicago, or even only of parole officers. He wished to make statements about social service agencies in general. He writes in his concluding chapter, "Rule breaking characterizes the official behavior of the social service bureaucrat" (McCleary, 1978, p. 171). Becker's study of marijuana users was not intended to be a study of those particular marijuana users, or even of marijuana users alone. The analysis says something about "the social interpretation of a physical experience which is . . . ambiguous" (Becker, 1963, p. 42). Seldom does a study purport to be about only the people or process examined in that study.

How do qualitative researchers achieve or assess external validity, particularly in the absence of replications? They do so by showing how the process they studied is *similar to* processes that

occur in other places and with other people. For instance, my research on how people become hypnotized was a study not only of hypnosis but of the phenomenon Becker observed among marijuana users. I also found that novices had to go through a series of stages to become good hypnotic subjects. The first was learning to behave as though they were hypnotized. The second was redefining that behavior to attribute it to a trance state rather than to a conscious act. The third was redefining "hypnosis" to include the behaviors and feelings the person experienced while behaving hypnotized. The steps in becoming hypnotized are similar to the steps in becoming a marijuana user. The novice first acquires some new behaviors—trancelike behaviors for hypnosis and effective smoking behaviors for marijuana use. Next, the novice must attribute those behaviors to the force of hypnosis rather than to conscious volition. Finally, the novice must change his or her criteria for what can pass as hypnosis, and call even mild states of relaxation hypnosis.

Becoming hypnotized, I argued, is like changing one's attitudes about one's own experiences and behaviors. The change is accomplished through negotiations between the hypnotist and the subject. The hypnotist's first step is to convince the subject that he or she acted like a hypnotized person. The second step is to attribute the behaviors not to compliance or coercion but to the trance state. For instance, when a subject asked: "How do you know if you were in a trance or not? I mean, I know I did some things, but I think they were all under conscious voluntary control," the hypnotist replied: "I think you can tell if someone is in a trance by looking at them . . . the facial expressions. I could walk around the room and tell who wasn't and who was, by how they responded. I thought you were, but maybe you didn't *think* you were." A second hypnotist reinforced this by saying to the same questioner: "You were actually the one that I thought went into trance the quickest" (Kidder, 1972, p. 317).

In addition to pointing out to subjects that they looked as though they had been in a trance, the hypnotists pointed out that the subjects looked and acted that way not because they had been forced but because they were under the influence of the trance. When participants said that although they behaved like hypnotic subjects, they felt they were just "playing the game," the hypno-

tists pointed out that there were no external constraints that forced them to do so. The responsibility lay with the subjects, who were asked: "Why did you feel that you wanted to play the game?"

The exchanges between hypnotists and their subjects were negotiations about who was responsible for the subject's becoming hypnotized. The following are samples of negotiated responsibility:

Hypnotist: Dr. Z tried to help those of you who weren't able to do the arm lift. He said, "For those for whom this was difficult, the arm can get very heavy" [so persons could let their arms go down instead of up]. . . . He pointed out the successes instead of failures and gave other possibilities for achieving success.

Subject: You make it sound as if it's the patient's fault instead of yours if he doesn't go into trance. . . .

Hypnotist: Well, let me say this. Earlier hypnosis was done in an authoritarian fashion—now it is much more permissive and we conceive of hypnosis as the achievement of the subject, in which the hypnotist helps. . . . Does this answer your question? [Kidder, 1972, p. 319.]

The hypnotist thus placed the responsibility with the subject. If he behaved like a hypnotized person, it was attributable to him or to his trance, not to external coercion.

The hypnotists negotiated several aspects of reality with their subjects. They negotiated the facts—whether or not someone appeared to have been in a trance; and they negotiated the locus of causality or responsibility for behaving hypnotized. I regard this as a study of negotiated realities and consider it similar to other studies of the social construction of reality (Berger and Luckman, 1967). The process I studied is like the process that Becker (1963) studied in his report on individuals who become marijuana users, and like the process Scheff (1968) studied in his report of therapist's and patient's negotiations and assessments of responsibility. I am less concerned with generalizing to other hypnosis workshops than I am with generalizing to other situations where two or more

parties negotiate definitions of what happened and who was responsible.

The external validity of research depends on what the researcher claims to have discovered. In the study of this particular hypnosis workshop, I claim to have discovered something about attitude change and negotiated responsibility more than to have discovered anything about hypnosis in general. It is likely, as the hypnotists themselves pointed out, that other, more "authoritarian" hypnotists would not place responsibility with their subjects, and the process of becoming hypnotized may be quite different. Rather than generalize to other hypnosis workshops which I have not observed, therefore, I wished to generalize to other phenomena which may appear very different on the surface but which share the stages and the negotiations that I described.

By saying this was not a study of hypnosis alone, I am not saying that no one could replicate the research or test it by examining another hypnosis workshop or by examining the same tape recordings that I analyzed. Instead, I contend that the external validity of my findings depends on their similarity to findings from very different settings, such as learning to become a marijuana user or learning to accept someone else's definition of one's ailment. This is not so different from the meaning of external validity for laboratory experiments. Their external validity depends not only on demonstrations that the same results can be obtained in another laboratory with another group of subjects but also on the apparent similarity between the laboratory observations and observations in a nonlaboratory "real-world" setting. With the exception of experiments designed to study the social psychology of the psychology experiment (for example, Orne, 1962, 1969), laboratory research is not meant to be generalizable only to other laboratories. It is designed to be an analogue of social processes outside the laboratory, which may bear no surface similarity to the laboratory procedures.

To demonstrate the external validity of any research, we must actually replicate it, varying either the subjects, the setting, the operational definitions of treatments and effects, or all of these. To assess the external validity of research short of replicating, we work inductively, and we compare the procedures and results of the study in question with the process to which the

researcher wishes to generalize. What matters is not their surface similarity but the apparent similarity of their processes, structure, or meaning. Since this comparison depends on inferred rather than obvious similarities, "the question of external validity, like the question of inductive inference, is never completely answerable" (Campbell and Stanley, 1966, p. 5).

Conclusions

I have argued that qualitative research can be not only "deep" and "rich" but also "hard." The same criteria that we apply to quantitative research—criteria for internal, external, and construct validity—can be applied to qualitative research. These criteria are not relevant for all qualitative research, only for that research which contains causal assertions. When qualitative research contains causal assertions, we can discern an implicit quasi-experimental design in the research. Since participant observation is seldom explicitly designed, the timing and frequency of observations rarely match exactly any of the quasi-experimental designs illustrated in the literature. They most nearly resemble multiple time-series designs, however, because they are frequently longitudinal observations of a socialization career. The multiple observations and nonsimultaneous treatments permit the investigator to rule out threats to external validity. Although this has nowhere been done explicitly, I demonstrated how qualitative researchers can and perhaps do rule out rival explanations in arriving at their conclusions about the necessary steps, stages, or conditions for socialization to proceed. These steps or conditions are the alleged causes. They are seldom single causes, and the conclusions do not say "This caused that." They are multistage, multideterministic analyses.

The construct validity and reliability of participant observation depend on the researcher's demonstration that there are multiple instances of a given construct. These multiple observations act like the multiple items or repeated measurements of a quantitative scale—they demonstrate that the construct exists. They demonstrate the reliability of the observations. The methods of making such observations are more varied than identical, so the multiple observations also provide convergent validation. Since

procedures for demonstrating reliability and validity differ in degree, and reliability and validity lie on a continuum, I am not violating our understanding of these terms when I say that the multiple observations of a construct may provide evidence of reliability or validity or something in between.

The external validity of qualitative research is ascertainable in the same way as the external validity of quantitative research. I do not claim that qualitative studies outrank quantitative studies in external validity, because naturally occurring events or field settings do not automatically confer high external validity. Field researchers may wish to generalize not to similar field settings but to similar processes in very different settings, as was the case with my study of a hypnosis workshop. The external validity of qualitative research, like that of quantitative research, depends on the underlying rather than the surface similarity between the process studied and the processes the researcher names as analogues.

Qualitative research can be assessed by the same criteria as quantitative research. The logic of internal, external, and construct validity is the same, regardless of whether the researcher uses words or numbers. Good qualitative research, like good quantitative research, is both rich and rigorous.

References

- Barton, A. H., and Lazarsfeld, P. F. "Qualitative Support of Theory." In G. J. McCall and J. L. Simmons (Eds.), *Issues in Participant Observation*. Reading, Mass.: Addison-Wesley, 1969.
- Becker, H. S. "Becoming a Marihuana User." In *Outsiders*. New York: Free Press, 1963.
- Becker, H. S., and Geer, B. "Participant Observation and Interviewing: A Comparison." *Human Organization*, 1957, 16 (3), 28-32.
- Berger, P. L., and Luckman, T. *The Social Construction of Reality*. Middlesex, England: Penguin Books, 1967.
- Campbell, D. T. "The Mutual Methodological Relevance of Anthropology and Psychology." In F. L. K. Hsu (Ed.), *Psychological Anthropology: Approaches to Culture and Personality*. Homewood, Ill.: Dorsey Press, 1961.
- Campbell, D. T. "Qualitative Knowing in Action Research." Kurt

- Lewin Award Address presented at 82nd annual meeting of the American Psychological Association, New Orleans, Sept. 1, 1974.
- Campbell, D. T. " 'Degrees of Freedom' and the Case Study." *Comparative Political Studies*, 1975, 8, 178-193.
- Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin*, 1959, 56, 81-105.
- Campbell, D. T., and Stanley, J. C. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1966.
- Cook, T. D., and Campbell, D. T. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally, 1979.
- Cressey, D. R. *Other People's Money: A Study in the Social Psychology of Embezzlement*. New York: Free Press, 1953.
- Glaser, B. G., and Strauss, A. L. *The Discovery of Grounded Theory*. Chicago: Aldine, 1967.
- Hersen, M., and Barlow, D. H. *Single-Case Experimental Designs: Strategies for Studying Behavior Change*. Elmsford, N.Y.: Pergamon Press, 1976.
- Kidder, L. H. "On Becoming Hypnotized: How Skeptics Become Convinced: A Case of Attitude Change?" *Journal of Abnormal Psychology*, 1972, 80 (3), 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 (1), 48-60.
- Lindesmith, A. R. *Addiction and Opiates*. Chicago: Aldine, 1968.
- McCall, G. J., and Simmons, J. L. *Issues in Participant Observation*. Reading, Mass.: Addison-Wesley, 1969.
- McCleary, R. "How Parole Officers Use Records." *Social Problems*, 1977, 24 (5), 576-589.
- McCleary, R. *Dangerous Men: The Sociology of Parole*. Beverly Hills, Calif.: Sage, 1978.
- Meehl, P. E. "Construct Validity in Psychological Tests." In *Psychodiagnosis: Selected Papers*. New York: Norton, 1977.
- Mills, C. W. *White Collar*. New York: Oxford University Press, 1951.
- Orne, M. T. "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics

- and Their Implications." *American Psychologist*, 1962, 17, 776-783.
- Orne, M. T. "Demand Characteristics and the Concept of Quasi-Controls." In R. Rosenthal and R. L. Rosnow (Eds.), *Artifact in Behavioral Research*. New York: Academic Press, 1969.
- Rosenthal, R., and Rosnow, R. L. (Eds.). *Artifact in Behavioral Research*. New York: Academic Press, 1969.
- Scheff, T. J. "Negotiating Reality: Notes on Power in the Assessment of Responsibility." *Social Problems*, 1968, 16, 3-17.
- Sherif, C. W. "Bias in Psychology." In J. A. Sherman and E. T. Beck (Eds.), *The Prism of Sex: Essays in the Sociology of Knowledge*. Madison: University of Wisconsin Press, 1979.
- Strauss, A., Schatzman, L., Bucher, R., Ehrlich, D., and Sabshin, M. *Psychiatric Ideologies and Institutions*. New York: Free Press, 1964.
- Zelditch, M., Jr. "Some Methodological Problems of Field Studies." *American Journal of Sociology*, 1962, 67, 566-576.
- Znaniecki, F. *The Method of Sociology*. New York: Holt, Rinehart and Winston, 1934.