

- Stake, R. (1978). The case-study method in social inquiry. *Educational Researcher*, 7, 5-8.
- Stake, R. (1980, April). *Generalizations*. Paper presented at the annual meeting of the American Educational Research Association, Boston.
- Suransky, V. (1982). *The erosion of childhood*. Chicago: University of Chicago Press.
- Thorndike, E. L. (1910). The contribution of psychology to education. *The Journal of Educational Psychology*, 1, 5-12.
- Turner, T. (1973). Piaget's structuralism. *American Anthropologist*, 75, 351-373.
- Vallence, E. (1977). The landscape of the great plains experience: An application of curriculum criticism. *Curriculum Inquiry*, 7, 87-105.
- Weiss, C. (1982). Policy research in the context of diffuse decision making. *Journal of Higher Education*, 53, 619-639.
- Wolcott, H. (1987). The teacher as enemy. In G. Spindler (Ed.), *Education and the cultural process* (pp. 136-150). Prospect Heights, IL: Waveland Press.

6 Increasing the Generalizability of Qualitative Research

JANET WARD SCHOFIELD

TRADITIONAL VIEWS OF GENERALIZABILITY

Campbell and Stanley (1963) laid the groundwork for much current thinking on the issue of generalizability just over twenty-five years ago in a groundbreaking chapter in the *Handbook of Research on Teaching*. They wrote, "External validity asks the question of generalizability: To what populations, settings, treatment variables, and measurement variables can the effect be generalized?" (p. 175; emphasis in original). They then went on to list four specific threats to external validity: the interaction of testing and the experimental treatment, the interaction of selection and treatment, reactive arrangements, and the interference of multiple treatments with one another. Although Campbell and Stanley specifically included populations, settings, treatments, and measurement variables as dimensions relevant to the concept of external validity, the aspect of external validity that has typically received the lion's share of attention in textbook and other treatments of the concept is generalizing to and across populations. This may well be due to the fact that, because of advances in sampling theory in survey research, it is possible to draw samples from even a very large and heterogeneous population and then to generalize to that population using the logic of probability statistics.

Campbell and Stanley (1963), as well as many others in the quantitative tradition, see the attempt to design research so that abstract generalizations can be drawn as a worthy effort, although issues connected with internal validity are typically given even higher priority. Thus researchers in the quantitative tradition have devoted considerable thought to the question of how the generalizability of experimental and quasi-experimental studies can be enhanced. Such efforts are

consistent with the fact that many quantitatively oriented researchers would agree with Smith (1975) that "the goal of science is to be able to generalize findings to diverse populations and times" (p. 88).

In contrast to the interest shown in external validity among quantitatively oriented researchers, the methodological literature on qualitative research has paid little attention to this issue, at least until quite recently. For example, Dobbert's (1982) text on qualitative research methods devotes an entire chapter to issues of validity and reliability but does no more than mention the issue of generalizability in passing on one or two pages. Two even more recent books, Kirk and Miller's *Reliability and Validity in Qualitative Research* (1986) and Berg's *Qualitative Research Methods for the Social Sciences* (1989), ignore the issue of external validity completely. The major factor contributing to the disregard of the issue of generalizability in the qualitative methodological literature appears to be a widely shared view that it is unimportant, unachievable, or both.

Many qualitative researchers actively reject generalizability as a goal. For example, Denzin (1983) writes:

The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. For the interpretivist every instance of social interaction, if thickly described (Geertz, 1973), represents a slice from the life world that is the proper subject matter for interpretive inquiry. . . . Every topic . . . must be seen as carrying its own logic, sense of order, structure, and meaning. (pp. 133-134)

Although not all researchers in the qualitative tradition reject generalization so strongly, many give it very low priority or see it as essentially irrelevant to their goals. One factor contributing to qualitative researchers' historical tendency to regard the issue of external validity as irrelevant and hence to disregard it is that this research tradition has been closely linked to cultural anthropology, with its emphasis on the study of exotic cultures. This work is often valued for its intrinsic interest, for showing the rich variety and possible range of human behavior, and for serving a historical function by describing traditional cultures before they change in an increasingly interconnected and homogeneous world. For researchers doing work of this sort, the goal is to describe a specific group in fine detail and to explain the patterns that exist, certainly not to discover general laws of human behavior.

Practically speaking, no matter what one's philosophical stance on the importance of generalizability, it is clear that numerous characteris-

tics that typify the qualitative approach are not consistent with achieving external validity as it has generally been conceptualized. For example, the traditional focus on single-case studies in qualitative research is obviously inconsistent with the requirements of statistical sampling procedures, which are usually seen as fundamental to generalizing from the data gathered in a study to some larger population. This fact is often cited as a major weakness of the case study approach (Bolgar, 1965; Shaughnessy & Zechmeister, 1985).

However, the incompatibility between classical conceptions of external validity and fundamental aspects of the qualitative approach goes well beyond this. To give just one example, the experimental tradition emphasizes replicability of results, as is apparent in Krathwohl's (1985) statement: "The heart of external validity is replicability. Would the results be reproducible in those target instances to which one intends to generalize—the population, situation, time, treatment form or format, measures, study designs and procedures?" (p. 123). Yet at the heart of the qualitative approach is the assumption that a piece of qualitative research is very much influenced by the researcher's individual attributes and perspectives. The goal is *not* to produce a standardized set of results that any other careful researcher in the same situation or studying the same issue would have produced. Rather it is to produce a coherent and illuminating description of and perspective on a situation that is based on and consistent with detailed study of that situation. Qualitative researchers have to question seriously the *internal* validity of their work if other researchers reading their field notes feel the evidence does not support the way in which they have depicted the situation. However, they do not expect other researchers in a similar or even the same situation to replicate their findings in the sense of independently coming up with a precisely similar conceptualization. As long as the other researchers' conclusions are not inconsistent with the original account, differences in the reports would not generally raise serious questions related to validity or generalizability.

In fact, I would argue that, except perhaps in multisite qualitative studies, which will be discussed later in this paper, it is impractical to make precise replication a criterion of generalizability in qualitative work. Qualitative research is so arduous that it is unlikely that high-quality researchers could be located to engage in the relatively unexciting task of conducting a study designed specifically to replicate a previous one. Yet studies not designed specifically for replication are unlikely to be conducted in a way that allows good assessment of the replicability issue. Of course it is possible, even likely, that specific ideas or conclusions from a piece of qualitative work can stimulate further research of

a qualitative or quantitative nature that provides information on the replicability of that one aspect of a study. However, any piece of qualitative research is likely to contain so many individual descriptive and conceptual components that replicating it on a piece-by-piece basis would be a major undertaking.

THE INCREASING INTEREST IN GENERALIZABILITY IN THE QUALITATIVE TRADITION

In the past decade, interest in the issue of generalizability has increased markedly for qualitative researchers involved in the study of education. Books by Patton (1980), Guba and Lincoln (1981), and Noblit and Hare (1988), as well as papers by Stake (1978), Kennedy (1979), and others, have all dealt with this issue in more than a cursory fashion. Two factors seem to be important in accounting for this increase in attention to the issue of generalizability. First, the uses of qualitative research have shifted quite markedly in the past decade or two. In the area of education, qualitative research is not an approach used primarily to study exotic foreign or deviant local cultures. Rather it has become an approach used widely in both evaluation research and basic research on educational issues in our own society. The issue of generalizability assumes real importance in both kinds of work.

The shift in the uses of qualitative work that occurred during the 1970s was rapid and striking. The most obvious part of this shift was the inclusion of major qualitative components in large-scale evaluation research efforts, which had previously been almost exclusively quantitative in nature (Fetterman, 1982; Firestone & Herriott, 1984). The acceptance of qualitative research as a valid and potentially rich approach to evaluation progressed to the point that Wolcott (1982) wrote, with only some exaggeration, "By the late 1970s the term 'ethnography' . . . had become synonymous with 'evaluation' in the minds of many educators" (p. 82). Evaluations are expensive and time-consuming undertakings. Although formative evaluations are usually site-specific, the worth of a summative evaluation is greatly enhanced to the extent it can inform program and policy decisions relating to other sites. In fact, as Cronbach (1982) points out, when summative evaluations are reported, no more than a fraction of the audience is interested primarily in the specific program and setting that was the object of the study. Even at the study site itself, by the time the evaluation is completed, changes may well have occurred that have important consequences for program functioning and goal achievement. Thus the ques-

tion of whether an evaluation's findings can usefully be generalized to a later point in time at the site at which the evaluation was conducted is an issue that, although often ignored, requires real consideration.

The issue of generalizability is also salient for more basic qualitative research on educational issues in this country. Funding agencies providing resources for qualitative studies of educational issues are presumably interested in shedding light on these issues generally, not just as they are experienced at one site. For example, I am currently directing a qualitative study of computer usage in an urban high school. It is clear that the impetus for the funding of this study by the Office of Naval Research derived from concerns about the Navy's own computer-based education and training efforts, not from concerns about the public schools. Quite apart from the goals of funding agencies, many qualitative researchers themselves hope to accomplish more than describing the culture of the specific school or classroom that they have chosen to study. For example, Peshkin (1982) writes of his study of school and community in a small town in Illinois, "I hoped . . . to explicate some reality which was not merely confined to other places just like Mansfield" (p. 63), a hope tellingly reflected in the title of his book, *Growing Up American* (1978), as opposed to "Growing Up in Illinois" or "Growing Up in Mansfield." This desire to have one's work be broadly useful is no doubt often stimulated by concern over the state of education in our country today. It is also clearly reinforced by the fact that, unlike most readers of ethnographic reports of exotic cultures, most readers of qualitative reports on American education have had considerable exposure during their own school years to at least one version of the culture described. Thus, unless the researcher chooses a very atypical site or presents an unusually insightful analysis of what is happening, the purely descriptive value of the study may be undercut or discounted.

So far I have argued that qualitative research's shift in both purpose and locale in the last decade or two has contributed to an increased interest in generalizability among qualitative researchers. There is yet one other factor contributing to this trend—the striking rapprochement between qualitative and quantitative methodologies that has occurred in the last decade (Cronbach et al., 1980; Filstead, 1979; Reichardt & Cook, 1979; Spindler, 1982). Exemplifying this trend is the shift in the position of Donald Campbell. Campbell and Stanley (1963) at one point contended that the "one-shot case study," which is one way of describing much qualitative research, has "such a total absence of control as to be of almost no scientific value" (p. 176). However, more recently Campbell (1979) wrote a paper to "correct some of [his] own

prior excesses in describing the case study approach" (p. 52) in which he takes the, for many, rather startling position that when qualitative and quantitative results conflict, "the quantitative results should be regarded as suspect until the reasons for the discrepancy are well understood" (p. 52).

One result of the rapprochement that has occurred is that qualitative and quantitative researchers are more in contact with each other's traditions than had typically been the case heretofore. As is often the case when a dominant tradition makes contact with a minority one, the culture and standards of the dominant group make a significant impact on the members of the minority group. This trend has most likely been reinforced by the fact that a great deal of the qualitative research on education conducted in the past fifteen years has been embedded within multimethod evaluation projects undertaken by private research firms that have traditionally specialized in quantitative research. Thus the concept of external validity and the associated issue of generalizability have been made salient for qualitative researchers, whose own tradition has not predisposed them to have given the issue a great deal of thought.

RECONCEPTUALIZING GENERALIZABILITY

Although many qualitative researchers have begun to recognize the importance of dealing with the issue of generalizability, it is clear that the classical view of external validity is of little help to qualitative researchers interested in finding ways of enhancing the likelihood that their work will speak to situations beyond the one immediately studied—that is, that it will be to some extent generalizable. The idea of sampling from a population of sites in order to generalize to the larger population is simply and obviously unworkable in all but the rarest situations for qualitative researchers, who often take several years to produce an intensive case study of one or a very small number of sites. Thus most of the work on generalizability by qualitative researchers in this decade has dealt with developing a *conception* of generalizability that is useful and appropriate for qualitative work.

A second approach to the issue of generalizability in qualitative research has been very different. A number of individuals have worked on ways of gaining generality through the synthesis of preexisting qualitative studies. For example, Noblit and Hare (1988) have recently published a slim volume on meta-ethnography. Substantially earlier, Lucas (1974) and Yin and Heald (1975) had developed what they call the

"case survey method." Ragin (1987) has presented yet another way of synthesizing qualitative studies, one that employs Boolean algebra. I will discuss these approaches to generalizing from qualitative case studies briefly at the end of this chapter. At the moment, I would like to focus on issues connected with the first approach—that is, with transforming and adapting the classical conception of external validity such that it is suitable for qualitative work.

Important and frequently cited discussions of conceptions of generalizability appropriate in qualitative work can be found in Guba and Lincoln (1981, 1982), Goetz and LeCompte (1984), and Stake (1978). Guba and Lincoln's stance on the issue of generalizability is aptly summarized in two excerpts of their own words. Guba and Lincoln write:

It is virtually impossible to imagine any human behavior that is not heavily mediated by the context in which it occurs. One can easily conclude that generalizations that are intended to be context free will have little that is useful to say about human behavior. (1981, p. 62)

They go on to say:

The aim of (naturalistic) inquiry is to develop an idiographic body of knowledge. This knowledge is best encapsulated in a series of "working hypotheses" that describe the individual case. Generalizations are impossible since phenomena are neither time- nor context-free (although some transferability of these hypotheses may be possible from situation to situation, depending on the degree of temporal and contextual similarity). (1982, p. 238)

Given these views, Guba and Lincoln call for replacing the concept of generalizability with that of "fittingness." Specifically, they argue that the concept of "fittingness," with its emphasis on analyzing the degree to which the situation studied matches other situations in which one is interested, provides a more realistic and workable way of thinking about the generalizability of research results than do more classical approaches. A logical consequence of this approach is an emphasis on supplying a substantial amount of information about the entity studied and the setting in which that entity was found. Without such information, it is impossible to make an informed judgment about whether the conclusions drawn from the study of any particular site are useful in understanding other sites.

Goetz and LeCompte (1984) place a similar emphasis on the importance of clear and detailed description as a means of allowing decisions

about the extent to which findings from one study are applicable to other situations. Specifically, they argue that qualitative studies gain their potential for applicability to other situations by providing what they call "comparability" and "translatability." The former term

refers to the degree to which components of a study—including the units of analysis, concepts generated, population characteristics, and settings—are sufficiently well described and defined that other researchers can use the results of the study as a basis for comparison. (p. 228)

Translatability is similar but refers to a clear description of one's theoretical stance and research techniques.

Stake (1978) starts out by agreeing with many critics of qualitative methods that one cannot confidently generalize from a single case to a target population of which that case is a member, since single members often poorly represent whole populations. However, he then goes on to argue that it is possible to use a process he calls "naturalistic generalization" to take the findings from one study and apply them to understanding another *similar* situation. He argues that through experience individuals come to be able to use both explicit comparisons between situations and tacit knowledge of those same situations to form useful naturalistic generalizations.

Several major themes can be found in the work of qualitative researchers who have written recently on the concept of generalizability. Whether it is Guba and Lincoln (1981, 1982) writing of fittingness, Goetz and LeCompte (1984) writing of translatability and comparability, or Stake (1978) discussing naturalistic generalizations, the emerging view shared by many qualitative researchers appears to involve several areas of consensus. First of all, there is broad agreement that generalizability in the sense of producing laws that apply universally is not a useful standard or goal for qualitative research. In fact, most qualitative researchers would join Cronbach (1982) in arguing that this is not a useful or obtainable goal for any kind of research in the social sciences. Second, most researchers writing on generalizability in the qualitative tradition agree that their rejection of generalizability as a search for broadly applicable laws is not a rejection of the idea that studies in one situation can be used to speak to or to help form a judgment about other situations. Third, as should be readily apparent from the preceding discussion, current thinking on generalizability argues that thick descriptions (Ryle, cited in Geertz, 1973) are vital. Such descriptions of both the site in which the studies are conducted and of the site to which one wishes to generalize are crucial in allowing

one to search for the similarities and differences between the situations. As Kennedy (1979) points out, analysis of these similarities and differences then makes it possible to make a reasoned judgment about the extent to which we can use the findings from one study as a "working hypothesis," to use Cronbach's (1982) term, about what might occur in the other situation. Of course, the generally unstated assumption underlying this view is that our knowledge of the phenomena under study is sufficient to direct attention to important rather than superficial similarities and differences. To the extent that our understanding is flawed, important similarities or differences may inadvertently be disregarded.

THREE TARGETS OF GENERALIZATION

Given the growing emphasis on generalizability in qualitative research and the emerging consensus about how the concept of generalizability might most usefully be viewed by qualitative researchers, two questions present themselves:

To what do we want to generalize?

How can we design qualitative studies in a way that maximizes their generalizability?

It is to these two questions that I will devote the majority of the rest of this chapter. Although I will use the term *generalize* here and elsewhere, it is important that the reader recognize that I am not talking about generalization in the classical sense. Rather, I use it to refer to the process as conceptualized by those qualitative researchers to whose work I have just referred.

I believe that it is useful for qualitative researchers interested in the study of educational processes and institutions to try to generalize to three domains: to *what is*, to *what may be*, and to *what could be*. I will deal with these possibilities one at a time, providing the rationale for striving to generalize to each of these kinds of situations and then suggesting some ideas on how studies can actually be designed to do this.

Studying What Is

From one perspective the study of any ongoing social situation, no matter how idiosyncratic or bizarre, is studying *what is*. But when I use the phrase *studying what is*, I mean to refer to studying the typical, the

common, or the ordinary. The goal of describing and understanding cultures or institutions as they typically are is an appropriate aim for much current qualitative research on educational institutions and processes. If policy makers need to decide how to change a program or whether to continue it, one very obvious and useful kind of information is information on how the program usually functions, what is usually achieved, and the like. Thus the goal of studying *what is* is one important aim for many kinds of summative evaluations. It is also appropriate outside of the area of evaluation for researchers hoping to provide a picture of the current educational scene that can be used for understanding or reflecting on it and possibly improving it. Classic works of this type that focus primarily on *what is* are Wolcott's *The Man in the Principal's Office* (1973) and Jackson's *Life in Classrooms* (1968). If one accepts the goal of designing research to maximize the fit between the research site and *what is* more broadly in society, an obvious question that arises is how this can be accomplished within the context of the qualitative tradition.

Studying the Typical. One approach sometimes used is to study the typical (Bogdan & Biklen, 1981; Goetz & LeCompte, 1984; Patton, 1980; Whyte, 1984). Specifically, I would argue that choosing sites on the basis of their fit with a typical situation is far preferable to choosing on the basis of convenience, a practice that is still quite common.

The suggestion that typicality be weighed heavily in site selection is an idea that needs to be taken both more and less seriously than it currently is. When I say that it needs to be taken more seriously than it currently is, I am suggesting that researchers contemplating selecting a site on the basis of convenience or ease of access need to think more carefully about that decision and to weigh very carefully the possibility of choosing on the basis of some other criterion, such as typicality. When I say that the strategy of selecting a typical site needs to be taken less seriously than it may sometimes be, I intend to point out that choosing a typical site is not a "quick fix" for the issue of generalizability, because what is typical on one dimension may not be typical on another. For example, Wolcott (1973) chose to focus his ethnographic study of a principal on an individual who was typical of other principals in gender, marital status, age, and so forth. This choice most likely substantially enhanced the range of applicability or generalizability of his study. Yet such a typical principal operating in an atypical school or an atypical system or even an atypical community might well behave very differently from a typical principal in a typical school in a typical system. The solution to this dilemma cannot be found in choosing

typicality on every dimension. First of all, not too many typical principals operate in environments that are typical in every way. So this strategy gains less in the realm of generalizability or fittingness than it might appear to at first glance. More important, even if one could achieve typicality in all major dimensions that seem relevant, it is nonetheless clearly true that there would be enough idiosyncrasy in any particular situation studied so that one could not transfer findings in an unthinking way from one typical situation to another.

Carried to extremes or taken too seriously, the idea of choosing on the basis of typicality becomes impossible, even absurd. However, as a guiding principle designed to increase the potential applicability of research, it is, I believe, useful. This is especially true if the search for typicality is combined with, rather than seen as a replacement for, a reliance on the kind of thick description emphasized by Guba and Lincoln (1981, 1982), Goetz and LeCompte (1984), and Stake (1978). Selection on the basis of typicality provides the potential for a good "fit" with many other situations. Thick description provides the information necessary to make informed judgments about the degree and extent of that fit in particular cases of interest.

In arguing that qualitative researchers would do well to seek to study the typical, I am not suggesting that we study the typical defined solely by national norms. Research that followed this prescription would greatly increase our knowledge of typical situations, but in a nation as diverse as the United States, it would provide too restricted, pallid, and homogeneous a view of our educational system. My emphasis on typicality implies that the researcher who has decided on the kind of institution or situation he or she wants to study—an urban ghetto school, a rural consolidated school, or a private Montessori school—should try to select an instance of this kind of situation that is, to the extent possible, typical of its kind. Such an approach suggests, for example, that a researcher interested in studying mathematics teaching choose to observe classrooms that use a popular text and generally accepted modes of instruction, rather than falling for convenience's sake into the study of classrooms that may well do neither of these. Furthermore, to the extent preliminary investigation of possible sites suggests that some or all are atypical in certain regards, careful thought about the possible implications of this atypicality for the topic under study may help to aid in site selection.

In sum, the point of my argument here is that choosing a site for research on the basis of typicality is far more likely to enhance the potential generalizability of one's study than choosing on the basis of convenience or ease of access—criteria that often weigh more heavily

than they should. However, even if one chooses on the basis of typicality, one is in no way relieved of the necessity for thick description, for it is foolhardy to think that a typical example will be typical in all important regards. Thus thick description is necessary to allow individuals to ask about the degree of fit between the case studied and the case to which they wish to generalize, even when the fit on some of the basic dimensions looks fairly close.

Performing Multisite Studies. An alternate approach to increasing the generalizability of qualitative research was evident in the sudden proliferation in the 1970s of multisite qualitative studies. Such studies were almost always part of federally funded evaluation efforts focusing on the same issue in a number of settings, using similar data collection and analysis procedures in each place. Well-known examples of this approach include the Study of Dissemination Efforts Supporting School Improvement (Crandall et al., 1983; Huberman & Miles, 1984) and the study of Parental Involvement in Federal Educational Programs (Smith & Robbins, 1984). One of the primary purposes of conducting such multisite studies is to escape what Firestone and Herriott (1984) have called the "radical particularism" of many case studies and hence to provide a firmer basis for generalization.

The multisite studies conducted in the 1970s were extremely varied, although they were all quite expensive and tended to take several years to complete. At least two kinds of variation have special implications for the extent to which this approach actually seems likely to produce results that are a good basis for generalization to many other situations. The first of these is the number of sites studied. Firestone and Herriott's (1984) survey of twenty-five multisite case study efforts found major variation on this dimension, with one study including as few as three sites and another covering sixty. All other things being equal, a finding emerging repeatedly in the study of numerous sites would appear to be more likely to be a good working hypothesis about some as yet unstudied site than a finding emerging from just one or two sites.

A second dimension on which multisite studies vary, which is also likely to effect the degree of fit between these studies and situations to which one might want to generalize, concerns the heterogeneity of the sites chosen for study. Generally speaking, a finding emerging from the study of several very heterogeneous sites would be more robust and thus more likely to be useful in understanding various other sites than one emerging from the study of several very similar sites (Kennedy, 1979). Heterogeneity can be obtained by searching out sites that will

provide maximal variation or by planned comparisons along certain potentially important dimensions. An example of the second strategy can be found in the parental-involvement study previously mentioned. The sites chosen for study were selected to allow comparison between urban and rural settings, between those with high and low reported degrees of involvement, and so forth (Smith & Robbins, 1984). This comparative strategy is potentially quite powerful, especially if there is heterogeneity among cases within each of the categories of interest. For example, if several rather different rural cases all share certain similarities that are not found in a heterogeneous group of urban cases, one has some reasonable basis for generalizing about likely differences between the two settings. Although the most obvious comparative strategy is to select cases that initially differ on some variable of interest as part of the research design, it is also possible to group cases in an *ex post facto* way on the basis of information gathered during the fieldwork. For example, if one were studying numerous very different classrooms and found that student achievement gains were quite high in some and quite low in others, one could compare these two sets of classrooms as a strategy for trying to suggest factors that contribute to high or low gains.

In sum, the possibility of studying numerous heterogeneous sites makes multisite studies one potentially useful approach to increasing the generalizability of qualitative work to *what is*. Yet I am very hesitant to see this approach as the only or even the best solution to the problem. First, such studies can be quite expensive, and the current lull in their funding highlights the extent to which such research is dependent on federal dollars that may or may not be forthcoming. Second, as Firestone and Herriott (1984) point out, budget constraints make it likely that studies including very large numbers of sites are less likely than studies of a relatively small number of sites to be able to devote intensive and prolonged care to studying the details of each site. Thus there is typically a trade-off to be made between the increased potential for generalizability flowing from studying a large number of sites and the increased depth and breadth of description and understanding made possible by a focus on a small number of sites. In suggesting that an increased number of sites leads to increased generalizability, I am assuming that enough attention is paid to each site to ensure that problems of internal validity do not arise. To the extent such problems do arise, generalizability is obviously threatened, since one cannot speak meaningfully of the generalizability of invalid data. The fact that roughly 40 percent of the multisite studies surveyed by Firestone and Herriott (1984) involved just one or two short visits to the research site

raises serious questions about whether such studies can appropriately be categorized as qualitative research in the usual sense of that term. The term *qualitative research*, and more especially the word *ethnography*, usually implies an intensive, ongoing involvement with individuals functioning in their everyday settings that is akin to, if not always identical with, the degree of immersion in a culture attained by anthropologists, who live in the society they study over a period of one or more years (Dobbert, 1982; Spindler, 1982; Wolcott, 1975). Thus it is conceivable, though not logically necessary, that attempts to gain generalizability through studying large numbers of sites undercut the depth of understanding of individual sites, which is the hallmark of the qualitative approach as it has come to be understood.

Studying What May Be

The goal of portraying typical schools—or, for that matter, typical instances of federal educational programs as they now exist—is, I believe, worthwhile. Yet accepting this as our only or even primary goal implies too narrow and limited a vision of what qualitative research can do. I would like to suggest that we want to generalize not only to *what is* but also to *what may be*. Let me explain. Here I am proposing that we think about what current social and educational trends suggest about likely educational issues for the future and design our research to illuminate such issues to the extent possible. Let me use some of my own current research to illustrate this possibility, without implying that it is the best or only example of such an approach.

One very obvious and potentially important trend in education recently has been the increasing utilization of microcomputers in instruction. In fact, microcomputers are being adopted in schools at an almost frantic pace (Becker, 1986) in spite of tight educational budgets and a generally acknowledged tendency on the part of educational institutions to resist rapid change. There is a clear division of opinion about the likely consequences of this trend. At one extreme are those who see computers as having the capability to revolutionize education in absolutely fundamental ways. Proponents of this school of thought make the rather startling claim that “the potential of computers for improving education is greater than that of any prior invention, including books and writing” (Walker, 1984, p. 3). Others take quite a different stance, emphasizing the inherent conservatism of the teaching profession with regard to pedagogical change and the failure of other highly touted educational innovations to bring about far-reaching changes. Thus it seemed important to me to design a research project

focused on understanding the impact of computer usage on students and classrooms (Schofield & Evans-Rhodes, 1989; Schofield & Verban, 1988). One could approach this issue with an emphasis on what is. For example, it would be possible to choose a school that is presently typical in terms of the uses it makes of computers in instruction. But this strategy encounters an immediate problem if one's goal is to speak to what may be. Changes in both microcomputer technology and in individuals' level of experience with computers have been so rapid in the past decade that a study of what is today could arguably be a study of primarily historical interest by the time it gets conducted, written, and published. In hopes of not just documenting the present, which is rapidly becoming the past, but of speaking to the future, I have made a number of methodological decisions that, in their abstract form, may be of use to others interested in making their work applicable to what may be.

Studying the “Leading Edge” of Change. First, since it is hard to know what kinds of computer usage will become most typical or popular in the future, I have made a point of studying a broad array of uses rather than just one particular kind. More important, I have not looked only for heterogeneity of usage but for types of usage that are now in their infancy but that many informed observers see as likely to be common in the future. Thus I consciously chose to study a school that not only uses computers as they are currently employed around the country to teach computer programming and word processing in fairly typical ways but that also was the field test site for the kind of artificially intelligent computer-based tutor that researchers in a number of centers around the country are currently developing for classroom use (Feigenbaum & McCorduck, 1983; Lawler & Yazdani, 1987.) I see this choice as a step in the direction of increasing the chances that this work will “fit” or be generalizable to the educational issues important at the time the work is published. But this is only a mere first step.

Probing Factors Likely to Differentiate the Present from the Future. One of the big problems in trying to make one's work applicable to even the fairly near future is, as Cronbach (1975) has so eloquently argued, that people and institutions change. Thus it is logically impossible to see the future even when studying futuristic uses of artificial intelligence, because one is studying that future technology in the context of a present-day institution peopled with individuals who are shaped by the era in which they live.

There is no completely satisfactory solution to this situation, but a partial one emerged as I grappled with the issue. It is to think through

how the present and the future are likely to differ. Then the research can be structured in a way that explicitly probes the impact of things that are likely to change over time. Of course, if the analysis of the likely differences between present and future is wrong, this approach will not be particularly useful. But if the analysis is accurate, this strategy has the potential to enhance greatly the usefulness of the study.

Let me illustrate in concrete terms how I have done this. Given the rapidity with which computers are being adopted for use in widely varying arenas of life, especially in schools, it seems a reasonable expectation that one major difference between now and five to ten years in the future is what might be called the "novelty factor." Specifically, many of today's high school students are having their first real introduction to the computer, or at least to its use for educational purposes, in their high school classrooms. However, in ten years it is rather unlikely that high school students will be having their first exposure to educational computing in the tenth or eleventh grade. I have used this assumption, which is, I think, relatively uncontroversial, to influence the shape of my study in a way that will allow it to speak more adequately to the future. For example, in interviews students were specifically asked about the impact of novelty on their reactions to the computer and its importance in shaping their feelings about computer usage. Similarly, observers in the study carefully looked for reactions that appeared to be influenced by students' unfamiliarity with the computers. Moreover, I have been careful to find out which students have had prior computer experience and what kind of experience this has been in order to see as clearly as possible whether these students differ from those for whom computer use is a completely novel experience. The fact that students were observed during the full course of the school year allowed assessment of whether any initial differences in students' reactions due to prior experience were transitory or relatively long-lasting. To the extent that novelty is crucial in shaping students' reactions, I will be forced to conclude that my study may not help us understand the future as well as it might otherwise. To the extent that students' reactions appear to be more heavily influenced by things that are unlikely to change in the near future, such as adolescents' striving for independence from adult control, the likely applicability of the findings of the study to the near future is clearly increased.

Considering the Life Cycle of a Phenomenon. The preceding discussion of the possible impact of novelty on students' reactions to educational computing brings up an important point regarding qualitative

work and the issue of generalizability. The ethnographic habit of looking at a phenomenon over substantial time periods allows assessment of one aspect of generalizability that quantitative research usually does not—of where a particular phenomenon is in its life cycle and what the implications of this are for what is happening. Qualitative research, when studying a dynamic phenomenon, is like a movie. It starts with one image and then moves on to others that show how things evolve over time. Quantitative research, in contrast, is more typically like a snapshot, often taken and used without great regard for whether that photograph happened to catch one looking one's best or looking unusually disheveled. This point can be illustrated more substantively by briefly discussing a study that I carried out in a desegregated school during its first four years of existence (Schofield, 1982/1989). The study tracked changes in the school by following two different groups of students from the first day they entered the school to graduation from that school three years later. Important changes occurred in race relations over the life of the institution and over the course of students' careers in the school. Such findings suggest that in asking about what happens in desegregated schools and what the impact of such schools is on students, it is important to know where both the students and the institution are in their experience with desegregation. Yet virtually all quantitative studies of desegregation, including, I must admit, some of my own, tend to ignore these issues completely. In fact, as I discovered in reviewing the desegregation literature (Schofield & Sagar, 1983), many do not even supply bare descriptive information on the life-cycle issue. Paying attention to where a phenomenon is in its life cycle does not guarantee that one can confidently predict how it will evolve. However, at a minimum, sensitivity to this issue makes it less likely that conclusions formed on the basis of a study conducted at one point in time will be unthinkingly and perhaps mistakenly generalized to other later points in time to which they may not apply.

Studying What Could Be

As mentioned previously, I would like to argue that qualitative research on education can be used not only to study *what is* and *what may be* but also to explore possible visions of *what could be*. By studying what could be, I mean locating situations that we know or expect to be ideal or exceptional on some *a priori* basis and then studying them to see what is actually going on there.

Selecting a Site That Sheds Light on What Could Be. When studying

what could be, site selection is not based on criteria such as typicality or heterogeneity. Rather it is based on information about either the *outcomes* achieved in the particular site studied or on the *conditions* obtaining there. Perhaps the best-known example of site selection based on outcomes is choosing to study classrooms or schools in which students show unusual intellectual gains, as has been done in the voluminous literature on effective schools (Bickel, 1983; Dwyer, Lee, Rowan, & Bossert, 1982; Phi Delta Kappan, 1980; Rutter, Maughan, Mortimore, Ouston, & Smith, 1979; Weber, 1971). For an example of site selection based on the conditions obtaining at the site, a less common approach, I will again make reference to my own work on school desegregation.

When thinking about where to locate the extended study of a desegregated school mentioned previously, I decided not to study a typical desegregated school. First, given the tremendous variation in situations characterized as desegregated, it is not clear that such an entity could be found. Second, there is a body of theory and research that gives us some basis for expecting different kinds of social processes and outcomes in different kinds of interracial schools. In fact, in the same year in which the *Brown v. Board of Education* decision laid the legal basis for desegregating educational institutions, Gordon Allport (1954) published a classic analysis of racial prejudice in which he argued that interracial contact can either increase or decrease hostility and stereotyping, depending on the kind of conditions under which it occurs. Specifically, he argued that in order to ameliorate relations between groups such as blacks and whites three conditions are especially important: equal status for members of both groups within the contact situation, a cooperative rather than a competitive goal structure, and support for positive relations from those in authority. A substantial amount of empirical and theoretical work stemming from Allport's basic insight has been carried out in the past three and a half decades, most of which supports his emphasis on the crucial importance of the specific conditions under which intergroup contact occurs (Amir, 1969; Aronson & Osherow, 1980; Cook, 1978; Pettigrew, 1967, 1969; Schofield, 1979; Schofield & Sagar, 1977; Slavin, 1980; Stephan, 1985).

It is clear that desegregating school systems often take little if any heed of the available theory and research on how to structure desegregated schools in a way likely to promote positive intergroup relations, perhaps at least partly because much of this work is laboratory based and hence may seem of questionable use in everyday situations. Thus selecting a site for study on the basis of typicality might be expected to yield a site potentially rich in sources of insight about the problems of desegregated education but weak in shedding light on what can be

accomplished in a serious and sophisticated effort to structure an environment conducive to fostering positive relations between students. Since both scholars in the area of intergroup relations and the public are well aware of the potential for difficulties in desegregated schools, the task of seeing whether and how such difficulties can be overcome seems potentially more informative and useful than that of documenting the existence of such difficulties. Thus I chose to study a site that at least approximated a theoretical ideal. My goal was not to generalize to desegregated schools as a class. Rather it was to see what happens under conditions that might be expected to foster relatively positive outcomes. If serious problems were encountered at such a site, there would be reason to think that problems would be encountered in most places or, alternatively, to revise or reject the theory that led to the site selection. However, if things went well at such a site, the study would then provide an opportunity to gain some insight into how and why they go well and into what the still-intractable problems are.

Of course, the strategy of choosing a site based on some *a priori* theoretical viewpoint or, for that matter, any seriously held expectation about it raises a difficult problem. If one is unduly committed to that viewpoint, one's analysis of both what happens and why may be heavily influenced by it, and one may not ask whether other more fruitful perspectives might emerge from a more dispassionate approach to studying the situation. This is the very danger that has led to the development of such elaborate safeguards in the quantitative tradition as the double-blind experiment. Although such procedures are rarely used in the qualitative tradition, a substantial literature on the issue of internal validity in qualitative research offers assistance with this problem to the researcher who pays it close heed (Becker, 1958; Bogdan & Biklen, 1981; Glaser & Strauss, 1967; Goetz & LeCompte, 1984; Guba, 1981; Guba & Lincoln, 1981; Kirk & Miller, 1986; Miles & Huberman, 1984a, 1984b; Patton, 1980; Strauss, 1987). Furthermore, if one's purpose is not to support or reject a specific *a priori* theory but to discover, using an approach that is as open as possible, what is actually happening in a site that was chosen with the assistance of a particular theory, problems related to internal validity are somewhat mitigated. For example, the fact that I chose to study a school that theory suggested might be conducive to positive relations did not keep me from exploring in considerable depth problems that occurred there (Sagar & Schofield, 1980; Schofield, 1981, 1982/1989).

One characteristic of the school chosen for the study was especially helpful in assessing the degree to which the theory on which the site was chosen was useful. Specifically, for various reasons, conditions in

two of the three grades in this school came much closer than conditions in the remaining grade to meeting those that theory suggests are conducive to producing positive relations. Thus it was possible to assess intergroup relations as the children went from one kind of environment to another within the school (Schofield, 1979, 1982/1989; Schofield & Sagar, 1977). This suggests one very useful strategy for studying what may be—selecting an “ideal” case and a comparative case that contrasts sharply on the relevant dimensions.

Generalizing from an Unusual Site to More Typical Ones. Although I indicated above that my goal was to learn about the possibilities and problems associated with a *certain kind* of desegregated education, I would like to argue that studying a site chosen for its special characteristics does not necessarily restrict the application of the study's findings to other very similar sites. The degree to which this is the case depends on the degree to which the findings appear to be linked to the special characteristics of the situation. Some of the findings from the study I have been discussing were clearly linked to unusual aspects of the school and hence have very limited generalizability to other situations, although they may nonetheless be important in demonstrating what is possible, even if not what is generally likely. For example, I found very low levels of overt racial conflict in the school studied (Schofield & Francis, 1982). It would obviously be misguided to conclude on the basis of this study that intergroup conflict is unlikely in all desegregated schools, since the school's emphasis on cooperation, equal status, and the like did actually appear to play a marked role in reducing the likelihood of conflict.

However, other findings that emerged from the study and were also related to atypical aspects of the situation may have a greater degree of applicability or generalizability than the finding discussed above. For example, I found the development of a color-blind perspective and of an almost complete taboo against the mention of race in the school studied (Schofield, 1986, 1982/1989). Since the emergence of the color-blind perspective and the accompanying taboo appeared to be linked to special characteristics of the school, I would not posit them as phenomena likely to occur in most desegregated schools. But I feel free to argue that *when* they do develop, certain consequences may well follow because these consequences are the logical outcomes of the phenomena. For example, with regard to the taboo against racial reference, if one cannot mention race, one cannot deal with resegregation in a straightforward way as a policy issue. Similarly, if one cannot men-

tion race, there is likely to be little or no effort to create or utilize multicultural curricular materials. Thus, although the taboo against racial reference may not occur in a high proportion of desegregated schools, when it does occur the study I carried out gives a potentially useful indication of problems that are likely to develop.

I would now like to turn to a third finding of the study, one so unrelated to the atypical aspects of the situation studied that it is a reasonable working hypothesis that this phenomenon is widespread. After I observed extensively in varied areas of the school and interviewed a large number of students, it became apparent that the white children perceived blacks as something of a threat to their physical selves. Specifically, they complained about what they perceived as black roughness or aggressiveness (Schofield, 1981, 1982/1989). In contrast, the black students perceived whites as a threat to their social selves. They complained about being ignored, avoided, and being treated as inferior by whites, whom they perceived to be stuck-up and prejudiced (Schofield, 1982/1989). Such findings appear to me to be linked to the black and white students' situation in the larger society and to powerful historical and economic forces, not to special aspects of the school. The consequences of these rather asymmetrical concerns may well play themselves out differently in different kinds of schools, but the existence of these rather different but deeply held concerns may well be widespread.

I have gone into some detail with these examples because I think they raise a crucial point for judging the applicability or generalizability of qualitative work. One cannot just look at a study and say that it is similar or dissimilar to another situation of concern. A much finer-grained analysis is necessary. One must ask what aspects of the situation are similar or different and to what aspects of the findings these are connected.

GENERALIZING THROUGH AGGREGATION OR COMPARISON OF INDEPENDENT STUDIES

This paper has argued that it is possible to achieve greater generalizability of qualitative research to situations of interest than is often now the case by following some of the design suggestions discussed above. However, there is another approach to increasing the generalizability of qualitative case studies that should not be ignored. This other strategy aims not at increasing the generalizability of one study or a set of

studies planned in conjunction with each other but at finding ways to aggregate, compare, or contrast already existing studies. One of these strategies was first laid out some time ago by Yin and Heald (1975). Another promising approach is suggested by Ragin's (1987) recent work on a strategy that he calls the "qualitative comparative method." A third very different approach has been outlined recently by Noblit and Hare (1988).

The Case Survey Method

Yin and Heald (1975) point out that case studies, whether qualitative or quantitative, are very prevalent in many fields. The nub of the problem from their perspective is that while "each case study may provide rich insights into a specific situation, it is difficult to generalize about the studies as a whole" (p. 371). Their solution to this problem is to propose a method for aggregating the information from separate studies. They call the method they developed the "case survey method." Basically this method consists of several steps. First, the literature relevant to one's interest is located. Then these studies are subjected to close scrutiny, so that those failing to meet certain crucial methodological requirements can be removed from the set to be analyzed. Then coders go through each of the remaining case studies with the goal of using the information contained therein to complete a set of closed-ended questions. These questions pertain to the topic of one's study and constitute the dataset ultimately used in the case survey approach. For example, Yin and Heald (1975) discuss a study of the effectiveness of urban decentralization efforts in which the closed-ended questions covered (1) the nature of the case study itself, (2) the context in which the decentralization effort occurred, (3) the characteristics of the specific effort at decentralization, and (4) five possible outcomes of decentralization. One then uses the material in the questionnaires to search for patterns on which generalizations can be based. The strategy for producing these generalizations is the use of statistical tests of association between different variables. For example, Yin and Heald report a statistically significant positive association between their judgments of the quality of specific case studies and the degree to which the study concluded that decentralization succeeded. The case survey procedure is parallel in some respects to the procedures suggested more recently by Miles and Huberman (1984a, 1984b) for aggregating data from multisite studies. However, Miles and Huberman tend not to emphasize statistical significance, perhaps because the number of studies in

many multisite qualitative endeavors is so small as to preclude attaining statistical significance unless the effects are of extraordinary strength.

As Yin and Heald (1975) acknowledge, there are clear limitations to the approach they suggest. First, of course, there must be a substantial body of literature available relevant to a particular topic for this procedure to work well. For example, Yin and Yates (1975) aggregated data from more than 250 studies of urban decentralization. When the number of available cases is small, statistical techniques lack power, since each case study must be treated as a single observation. Also, in such cases the number of variables worthy of coding may well be large compared to the number of sampling points (i.e., case studies), which also poses statistical problems. Second, the case survey method, with its emphasis on reducing the rich descriptive material provided in many case studies to uniform quantifiable data, risks ignoring unique factors that may be crucial to understanding specific cases or kinds of cases. Third, as Yin and Heald (1975) note, the case survey method may be more suited to inquiries focusing on outcomes rather than on process. Because of the numerous limitations of the case survey method, Yin (1981) has concluded that the "case-survey method should be used in highly selective situations" (p. 63) and that other methods for comparing across cases may ultimately prove more fruitful.

Unfortunately, the development of other methods for comparing and aggregating across cases, especially cases that have not been planned as part of a unified multisite effort, are not well developed. Although the work of Miles and Huberman (1984a, 1984b) and Yin (1984) provides many useful design and analysis suggestions for investigators planning multisite studies, relatively little methodological guidance is available to researchers who wish to compare studies that were designed and executed independently. A crucial difference between these two cases, of course, is that in the former one can obtain some degree of uniformity in the information gathered. This is crucial for the kinds of pattern-producing techniques suggested by Miles and Huberman and by Yin. There are other important differences as well. For example, in a multisite study with central direction it is at least theoretically possible, if not eminently practical, for one individual to have access to the raw data from all the different sites. However, such is generally not the case when one is trying to conduct comparisons of previously published case studies or ethnographies. Thus, at this point in time, our ability to achieve generalizations through the comparison of independently conducted pieces of qualitative work on a particular topic is quite limited.

The Qualitative Comparative Method

One promising new strategy for aggregating case studies has recently been developed by Ragin (1987). Ragin starts with the premise that two of the distinctive traits of case studies, and of case-oriented comparative research more generally, are their attention to cases as wholes and to the possibility that several different sets of circumstances can lead to the same outcome. He argues that most attempts to aggregate numerous case studies using quantitative approaches tend not to make use of these strengths and thus do not make full use of the data bases on which they are built. To remedy this situation, Ragin proposes an approach that he calls the "qualitative comparative method." This approach is based on Boolean algebra, the algebra of sets and logic. Although a full discussion of this technique is beyond the scope of this paper, since it would require introducing readers to the basics of Boolean algebra, it is possible briefly to discuss Ragin's general approach without becoming unduly technical.

First, Ragin's techniques can be used with widely varying numbers of case studies as one's raw data. In this regard it is more flexible than the case survey method, which is suitable only when relatively large numbers of case studies are available because of its dependence on the concept of statistical significance. Second, the techniques can be used with either preexisting case studies or with multisite studies planned with the qualitative comparative strategy in mind. All that is necessary are data that allow one to build truth tables—that is, categorical information on the variables of major interest to the analysis. Ragin argues that his approach allows one to examine complex and multiple patterns of causation, to produce parsimonious explanations, to study cases both as wholes and as parts, and to evaluate competing explanations. Ragin presents several extended and sharply contrasting examples of the varied ways in which the approach he uses can be applied. Although his approach seems better suited in many ways to aggregating qualitative case studies than the case survey method, since a Boolean approach allows one to take better advantage of the characteristic strengths of case studies, it is too early to understand completely either its full potential or the various problems that individuals using this approach will face.

Meta-Ethnography

Consideration of the techniques discussed above suggests that both the case survey method and attempts at case comparison are often

based on a logic that seeks to generalize by aggregating studies. Noblit and Hare (1988) suggest that such an effort is misdirected, arguing that efforts at aggregation tend to ignore the interpretive nature of qualitative research and to miss much of what is most important in each study. They believe it is possible to systematically compare very diverse cases in order to draw cross-case conclusions. However, they see such an effort as best conceptualized as the *translation* of studies into one another rather than as their aggregation. They call this translation "meta-ethnography."

Noblit and Hare argue that studies of similar topics can be seen as directly comparable, as essentially refutational, or as together suggesting a new line of argument. Once a preliminary look at the material to be synthesized suggests which of the above is the case, a translation and synthesis is attempted. This process may refute the initial assumption about the relation between the cases, but it would generally be expected not to do so.

In order to perform the translation and synthesis, Noblit and Hare suggest a focus on and a listing of the concepts, themes, and metaphors that the author of each study utilizes. The meta-ethnographer lists and organizes these themes and then attempts to relate them to one another. This somewhat abstract process is perhaps best clarified by a brief example. Noblit and Hare exemplify the idea of a reciprocal translation of studies by comparing Collins and Noblit's (1978) research in a desegregated school to Wolcott's study, *The Man in the Principal's Office* (1973). The comparison makes sense and, in fact, is only possible because Collins and Noblit's study laid great emphasis on the role of the principal in the desegregated school they studied. Noblit and Hare list the terms used in both studies to describe the context in which the principal functioned, the principal's behavior, and the like. The meta-ethnography then consists of a discussion of the ways in which the two situations and studies appear to be similar and different and, more important, of the extent to which the themes developed in each are adequate to handle the other ethnography as well. These judgments are based on attributes of the themes, such as their economy, cogency, and scope. For example, Wolcott describes the conduct of the principal he studied as characterized by patience and prudence. Collins and Noblit compare two different principals in a particular school. The first was said to have created negotiated order in the school. His successor, with a far different style, created what Collins and Noblit characterized as a bureaucratic order. After discussing the particulars of the two studies, Noblit and Hare (1988) conclude that a translation between them is possible but that Wolcott's metaphors are more adequate to this task

than those of Collins and Noblit. This means that Wolcott's concepts were able to capture what occurred in the Collins and Noblit study in a fuller and more adequate way than the Collins and Noblit themes fit the Wolcott study. Of course, it is possible in a meta-ethnography that none of the studies compared will have characterized its themes in a way that adequately fits all others, even though there are many parallels. In such a case, the hope is that the individual doing the meta-ethnography may be able to produce new, more inclusive concepts that work better than those from any particular study.

SUMMARY AND CONCLUSIONS

Although qualitative researchers have traditionally paid scant attention to the issue of attaining generalizability in research, sometimes even disdaining such a goal, this situation has changed noticeably in the past ten to fifteen years. Several trends, including the growing use of qualitative studies in evaluation and policy-oriented research, have led to an increased awareness of the importance of structuring qualitative studies in a way that enhances their implications for the understanding of other situations.

Much of the attention given to the issue of generalizability in recent years on the part of qualitative researchers has focused on redefining the concept in a way that is useful and meaningful for those engaged in qualitative work. A consensus appears to be emerging that for qualitative researchers generalizability is best thought of as a matter of the "fit" between the situation studied and others to which one might be interested in applying the concepts and conclusions of that study. This conceptualization makes thick descriptions crucial, since without them one does not have the information necessary for an informed judgment about the issue of fit.

This paper argues that three useful targets for generalization are *what is*, *what may be*, and *what could be* and provides some examples of how qualitative research can be designed in a way that increases its ability to fit with each of these situations. Studying *what is* refers to studying the typical, the common, and the ordinary. Techniques suggested for studying *what is* include choosing study sites on the basis of typicality and conducting multisite studies. Studying *what may be* refers to designing studies so that their fit with future trends and issues is maximized. Techniques suggested for studying *what may be* include seeking out sites in which one can study situations likely to become more common with the passage of time and paying close attention to how such present

instances of future practices are likely to differ from their future realizations. Studying *what could be* refers to locating situations that we know or expect to be ideal or exceptional on some *a priori* basis and studying them to see what is actually going on there. Crucial here is an openness to having one's expectations about the phenomena disconfirmed.

A very different approach to increasing the generalizability of qualitative research is evident in the work of some scholars who have focused on how to achieve generalizability through the aggregation or comparison of extant independently designed case studies or ethnographies. The case survey approach suggested by Yin and Heald (1975) is promising in a limited number of cases in which comparable information is available from a relatively large number of studies. Case comparison strategies, such as the qualitative comparative method suggested by Ragin (1987), may be more realistic and fruitful in many areas of research; but these comparative techniques are still in the early stages of development. Noblit and Hare (1988) suggest a kind of comparison they call "meta-ethnography," which focuses on the reciprocal translation rather than the aggregation of studies. Although such an approach may have promise, it is so new that its ultimate fruitfulness is still quite untested.

Acknowledgments

Much of the research on which this paper is based was funded by the Office of Naval Research, Contract Number N00 14-85-K-0664. Other research utilized in this paper was funded by Grant Number NIE-G-78-0126 from the National Institute of Education. However, all opinions expressed herein are solely those of the author, and no endorsement by ONR or NIE is implied or intended. My sincere thanks go to Bill Firestone and Matthew Miles for their constructive comments on an earlier draft of this paper.

REFERENCES

- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, UK: Cambridge University Press.
- Amir, Y. (1969). Contact hypothesis in ethnic relations. *Psychological Bulletin*, 71, 319-342.
- Aronson, E., & Osherow, N. (1980). Cooperation, prosocial behavior, and academic performance: Experiments in the desegregated classroom. In

- L. Bickman (Ed.), *Applied social psychology annual* (Vol. 1, pp. 163-196). Beverly Hills, CA: Sage.
- Becker, H. J. (1986). Instructional uses of school computers. *Reports from the 1985 National Survey* (Issue No. 1) (pp. 1-9). Baltimore, MD: Center for Social Organization of Schools, Johns Hopkins University.
- Becker, H. S. (1958). Problems of inference and proof in participant observation. *American Sociological Review*, 23, 652-659.
- Berg, B. L. (1989). *Qualitative research methods for the social sciences*. Boston: Allyn & Bacon.
- Bickel, W. E. (1983). Effective schools: Knowledge, dissemination, inquiry. *Educational Researcher*, 12(4), 3-5.
- Bogdan, R. C., & Biklen, S. K. (1981). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn & Bacon.
- Bolgar, H. (1965). The case study method. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 28-39). New York: McGraw-Hill.
- Brown v. Board of Education* (1954). 347 U.S. 483.
- Campbell, D. T. (1979). Degrees of freedom and the case study. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 49-67). Beverly Hills, CA: Sage.
- Campbell, D., & Stanley, J. (1963). Experimental and quasi-experimental designs for research on teaching. In N. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.
- Collins, T., & Noblit, G. (1978). *Stratification and resegregation: The case of Crossover High School*. Final report of NIE contract #400-76-009.
- Cook, S. W. (1978). Interpersonal and attitudinal outcomes in cooperating interracial groups. *Journal of Research and Development in Education*, 12, 97-113.
- Crandall, D. P., et al. (1983). *People, policies and practices: Examining the chain of school improvement* (Vols. 1-10). Andover, MA: The Network.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Denzin, N. K. (1983). Interpretive interactionism. In G. Morgan (Ed.), *Beyond method: Strategies for social research* (pp. 129-146). Beverly Hills, CA: Sage.
- Dobbert, M. L. (1982). *Ethnographic research: Theory and application for modern schools and societies*. New York: Praeger.
- Dwyer, D. C., Lee, G. V., Rowan, B., & Bossert, S. T. (1982). *The principal's role in instructional management: Five participant observation studies of principals in action*. Unpublished manuscript, Far West Laboratory for Educational Research and Development, San Francisco.
- Feigenbaum, E. A., & McCorduck, P. (1983). *The fifth generation: Artificial intelligence and Japan's computer challenge to the world*. Reading, MA: Addison-Wesley.

- Fetterman, D. M. (1982). Ethnography in educational research: The dynamics of diffusion. In D. M. Fetterman (Ed.), *Ethnography in educational evaluation* (pp. 21-35). Beverly Hills, CA: Sage.
- Frustard, W. J. (1979). Qualitative methods: A needed perspective in evaluation research. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 33-48). Beverly Hills, CA: Sage.
- Freestone, W. A., & Herrnett, R. F. (1984). Multisite qualitative policy research: Some design and implementation issues. In D. M. Fetterman (Ed.), *Ethnography in educational evaluation* (pp. 63-88). Beverly Hills, CA: Sage.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz (Ed.), *The interpretation of cultures* (pp. 3-30). New York: Basic Books.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine Publishing.
- Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in education research*. Orlando, FL: Academic Press.
- Guba, E. (1981). Criteria for assessing the trustworthiness of naturalistic inquiry. *Educational Communication and Technology Journal*, 29, 79-92.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass.
- Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication and Technology Journal*, 30, 233-252.
- Huberman, A. M., & Miles, M. B. (1984). *Innovation up close: How school improvement works*. New York: Plenum Press.
- Jackson, P. W. (1968). *Life in classrooms*. New York: Holt, Rinehart and Winston.
- Kennedy, M. M. (1979). Generalizing from single case studies. *Evaluation Quarterly*, 3(4), 661-678.
- Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research*. Beverly Hills, CA: Sage.
- Krathwohl, D. R. (1985). *Social and behavioral science research: A new framework for conceptualizing, implementing, and evaluating research studies*. San Francisco: Jossey-Bass.
- Lawler, R. W., & Yazdani, M. (Eds.). (1987). *Artificial intelligence and education: Learning environments and tutoring systems* (Vol. 1). Norwood, NJ: Ablex Publishing.
- Lucas, W. (1974). *The case survey method: Aggregating case experience*. Santa Monica, CA: Rand.
- Miles, M., & Huberman, A. (1984a). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher*, 13, 20-30.
- Miles, M., & Huberman, A. (1984b). *Qualitative data analysis: A sourcebook of new methods*. Newbury Park, CA: Sage.
- Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Beverly Hills, CA: Sage.

- Patton, M. Q. (1980). *Qualitative evaluation methods*. Beverly Hills, CA: Sage.
- Peshkin, A. (1978). *Growing up American: Schooling and the survival of community*. Chicago: University of Chicago Press.
- Peshkin, A. (1982). The researcher and subjectivity: Reflections on an ethnography of school and community. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action* (pp. 48-67). New York: Holt, Rinehart and Winston.
- Pettigrew, T. (1967). Social evaluation theory: Convergences and applications. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 5). Lincoln: University of Nebraska Press.
- Pettigrew, T. (1969). Racially separate or together. *Journal of Social Issues*, 25, 43-69.
- Phi Delta Kappan. (1980). *Why do some urban schools succeed? The Phi Delta Kappa study of exceptional urban elementary schools*. Bloomington: Phi Delta Kappa and Indiana University.
- Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.
- Reichardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 1-33). Beverly Hills, CA: Sage.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
- Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, 39, 590-598.
- Schofield, J. W. (1979). The impact of positively structured contact on intergroup behavior: Does it last under adverse conditions? *Social Psychology Quarterly*, 42, 280-284.
- Schofield, J. W. (1981). Competitive and complementary identities: Images and interaction in an interracial school. In S. Asher & J. Gottman (Eds.), *The development of children's friendships*. New York: Cambridge University Press.
- Schofield, J. W. (1986). Causes and consequences of the colorblind perspective. In S. Gaertner & J. Dovidio (Eds.), *Prejudice, discrimination and racism: Theory and practice* (pp. 231-253). New York: Academic Press.
- Schofield, J. W. (1989). *Black and white in schools: Trust, tension, or tolerance?* New York: Teachers College Press. (Original work published 1982)
- Schofield, J. W., & Evans-Rhodes, D. (1989, May). *Artificial intelligence in the classroom: The impact of a computer based tutor on teachers and students*. Paper presented at the 4th International Conference on Artificial Intelligence in Education, Amsterdam, The Netherlands.
- Schofield, J. W., & Francis, W. D. (1982). An observational study of peer interaction in racially-mixed "accelerated" classrooms. *The Journal of Educational Psychology*, 74, 722-732.

- Schofield, J. W., & Sagar, H. A. (1977). Peer interaction patterns in an integrated middle school. *Sociometry*, 40, 130-138.
- Schofield, J. W., & Sagar, H. A. (1983). Desegregation, school practices and student race relations. In C. Rossell & W. Hawley (Eds.), *The consequences of school desegregation* (pp. 58-102). Philadelphia, PA: Temple University Press.
- Schofield, J. W., & Verban, D. (1988). Computer usage in the teaching of mathematics: Issues which need answers. In D. Grouws & T. Cooney (Eds.), *Effective mathematics teaching* (pp. 169-193). Hillsdale, NJ: Erlbaum.
- Shaughnessy, J. J., & Zechmeister, E. B. (1985). *Research methods in psychology*. New York: Knopf.
- Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research*, 50, 315-342.
- Smith, A. G., & Robbins, A. E. (1984). Multimethod policy research: A case study of structure and flexibility. In D. M. Fetterman (Ed.), *Ethnography in educational evaluation* (pp. 115-132). Beverly Hills, CA: Sage.
- Smith, H. W. (1975). *Strategies of social research: The methodological imagination*. Englewood Cliffs, NJ: Prentice-Hall.
- Spindler, G. (1982). General introduction. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action* (pp. 1-13). New York: Holt, Rinehart and Winston.
- Stake, R. E. (1978). The case-study method in social inquiry. *Educational Researcher*, 7, 5-8.
- Stephan, W. J. (1985). Intergroup relations. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 2, pp. 599-658). New York: Random House.
- Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge, UK: Cambridge University Press.
- Walker, D. F. (1981). Promise, potential and pragmatism: Computers in high school. *Institut for Research in Educational Finance and Governance Policy Notes*, 5, 3-4.
- Weber, G. (1971). *How city children can be taught to read: Four successful schools*. Washington, DC: Council for Basic Education.
- Whyte, W. F. (1984). *Learning from the field: A guide from experience*. Beverly Hills, CA: Sage.
- Wolcott, H. F. (1973). *The man in the principal's office: An ethnography*. New York: Holt, Rinehart and Winston.
- Wolcott, H. F. (1975). Criteria for an ethnographic approach to research in schools. *Human Organization*, 34, 111-127.
- Wolcott, H. F. (1982). Mirrors, models, and monitors: Educator adaptations of the ethnographic innovation. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action* (pp. 68-95). New York: Holt, Rinehart and Winston.
- Yin, R. K. (1981). The case study crisis: Some answers. *Administrative Science Quarterly*, 26, 58-64.

- Yin, R. K. (1984). *Case study research: Design and methods*. Beverly Hills, CA: Sage.
- Yin, R. K., & Heald, K. A. (1975). Using the case survey method to analyze policy studies. *Administrative Science Quarterly*, 20, 371-381.
- Yin, R. K., & Yates, D. (1975). *Street-level governments: Assessing decentralization and urban services*. Lexington, MA: D. C. Heath.

*Commentary on the Papers
by Donmoyer and by Schofield*

Generalizing from Case Studies

HOWARD S. BECKER

WHY ARE WE HAVING THIS CONFERENCE?

When I was asked to participate in (yet another) conference on qualitative research in education, I wondered, "Why in the world are they having this conference?" I knew the answer, of course: Qualitative research is still somehow not respected as real science in education, so those of us who do it have to get together to answer those questions still one more time.

But why is qualitative work still not respected as real science in education? It's nothing new. For me, qualitative research on education starts with Willard Waller's *Sociology of Teaching*, published in 1932; others may want to start earlier or later, but the tradition is certainly at least fifty years old.

And the tradition thrives. A reasonably selective bibliography of solid qualitative studies of educational organizations and activities would include dozens, maybe hundreds, of items, many of them classics of social research. Such studies have added immeasurably to our understanding, not just of educational matters, but of social organization generally.

So what's wrong? There are several overlapping answers to that. One is that education, like other applied fields, is dominated not by sophisticated researchers but by laypeople who are the consumers of educational research. These consumers—teachers, administrators, and others—have had some training in research, but not very much. Often they have had just one course in which, since education is still dominated by models taken from psychology, they learn what "real research" looks like. It looks like experiments, with comparable data gathered in standardized ways from isolated subjects, summarized and tested with old-fashioned statistics. Since qualitative research—can I

substitute "field research"?—doesn't look anything like that, these consumers don't recognize it as having any value.

More sophisticated research consumers have another complaint about qualitative research. Questionnaires and tests can be examined ahead of time and purged of dangerous material. Qualitative researchers, on the other hand, won't say what they are going to do, claiming they don't know for sure. They just hang around, poking into everything. No telling what they'll ask, what they'll stumble onto, what untoward events will happen when they happen to be there to see and hear the whole thing, what students and subordinates will tell them when you aren't looking. Canny administrators try to contain snoopy sociologists, but it isn't easy. No wonder some presidents of small colleges would not let David Reisman interview their faculty about civil liberties unless they were in the room!

Worse yet, good fieldworkers do not restrict themselves to studying the help but insist on treating administrators and other important people as objects of study, whose actions are to be investigated as skeptically and objectively as those of their underlings.

Finally, field research is fundamentally useless to people who run educational institutions, not because it cannot uncover solutions to problems but because the solutions it uncovers are too radical to be "practical." Field studies usually find the causes of educational problems in the basic character of the organizations in which they arise. That inevitably suggests that the problems cannot be solved without making changes administrators cannot or do not want to make. Quantitative studies that don't look so deep make less trouble.

Here's a simple example. When my colleagues and I discussed a draft of our study of medical students (Becker, Geer, Hughes, & Strauss, 1961) with the faculty of the school we had studied, they wanted to know what our recommendations were. We said we didn't have any but would suggest solutions to particular problems they described. They said they didn't like the way students crammed for exams by memorizing stuff from books and lecture notes.

"Well, what skill do you want to test students for?"

"We'd like students to be able to examine a patient, take a history, order appropriate lab work, make a reasonable diagnosis, and suggest a plan of treatment."

"Fine. Give each student two patients to work up, then check their findings against yours."

They looked glum when we said that. "That's all very well in theory, but it's not practical."

"Why not?"

They had their own practices to attend to, their research to do, their papers to write; there was no time for such an elaborate testing procedure, not without giving up things clearly more important to them. They wanted a solution to the problem, but not at that price. We could solve the problem as originally stated, but not with those additional constraints.

But those are all reasons why the typical nonresearcher consumers of research on educational organizations don't like qualitative research and attack it by calling it unscientific. We, who are researchers, have to deal with that, but we don't have to believe that it poses deep epistemological problems. Our epistemological problems are no worse than those of social scientists working on any other kind of social organization. If we think that, say, William Foote Whyte (1943) had it more or less right when he did *Street Corner Society* (and who doesn't?), then we don't have anything special to worry about.

Nevertheless, people are worried about this problem of generalizing. I'd like to propose a diagnosis and a remedy, not because I'm sure I'm right, but provisionally, as a way of opening another perspective on the topic.

WHOM ARE WE GENERALIZING FOR?

We can agree with Donmoyer that human activity ought to be seen, à la Blumer (1969), as collective action. We can push that a little further, and perhaps avoid some of the confusions surrounding the idea of generalization, by seeing making generalizations as something social scientists do together, as a routine part of their work activity.

That view of generalizing lets us take advantage of recent work by sociologists of science (e.g., Latour, 1987; Fujimura, Star, & Gerson, 1987, and the literature cited there), who have begun to look at science not as a form of knowledge but as a kind of work. Such a view does not canonize any one form of generalization as The Right Way to do it. Rather, we can look and see what kinds of generalizations the people doing this kind of science would like to make, what kinds their situation makes possible, and what they have to do to get those generalizations.

Consumers' responses, a major aspect of the situation researchers work in, help shape the work product. Educational researchers usually make their generalizations for multiple audiences of consumers: for social scientists and theorists of education, to be sure, but also for educators, administrators, and the lay public. Each audience wants a

somewhat different kind of generalization, some of which are more possible than others.

Since, of course, we are not just studying this form of science-making, but are also engaging in it, we cannot be satisfied with that kind of neutral, detached observation. We need to go beyond analysis of the problem and find some sort of solution.

WHAT ARE WE GENERALIZING ABOUT?

Applied academic fields have their subject matter defined for them by the institutions for which they work. So the subject matter of the field of education is schools, and what educational researchers want to generalize about is schools. What I find out by studying *this* school, on this view, ought to apply to *all* schools, or all members of some subclass of schools.

We might try to redefine the subject matter as education, wherever it happens and by whomever it is carried on. We might study the way thieves teach one another the latest techniques of their trade, or the way young people teach one another to use drugs or engage in sexual activity. But that's just cheap irony, because everyone knows that those activities are not education, at least not what any reasonable layperson means by education. When we study education, we want to study schools in order to find out about all schools.

Generalization becomes a problem because, if we think of education and learning as generic social processes, there is no reason to think that those processes take place only in schools. There is no reason, for that matter, to think they take place in schools at all, even though that is the story schools tell about themselves and the story well-socialized members of our society believe or, at least, pretend to believe so that they won't appear to be nuts.

Generalization is a problem, too, because schools are not all alike in the way they are organized, so that there is no good reason to think that what we find in one will be found in any others. Both of these problems make it difficult to see what you can generalize about if the subject of your generalizations has to be schools.

Where Does Education Take Place?

If we want to generalize about education, we might just ignore schools and study the generic social processes of teaching and learning, wherever they occur and whoever is involved. This enterprise quickly

runs afoul of conventional, strongly entrenched ideas of what these words refer to. We can, of course, make words mean just what we want them to mean, and some of us could get away with this, but I doubt we could do it in any organization that claimed to be dealing with education as ordinarily defined and tried to raise money on that basis.

The word *education* assumes that people learn by being taught, that it consists, as a social enterprise, of knowledgeable people teaching people who are less knowledgeable, and typically, not surprisingly, less powerful and less well placed (children or immigrants, for instance), and that this activity goes on in schools. That's what education *is*. You can study, as a number of people have, how youngsters learn to use marijuana. You may find, as Schaps and Sanders (1970) did, that young women typically learn from their boyfriends, while the boyfriends learn from one another. You cannot, however, call that education and expect anyone to take you seriously.

But suppose that the process by which boyfriends teach girlfriends to smoke dope has a lot in common with other activities in which knowledge, skill, and ideas are passed on. It might, for instance, resemble the system described by Gagnon and Simon (1973), in which young women teach their boyfriends to engage in romance, which they have been practicing by themselves for quite a while, while the boyfriends are teaching them to engage in sex, which *they* have been practicing by *themselves* for a similar period of time. If the process works, and each learns what the other knows, they can manage to fall in love in the more or less standard way.

We might find, further, that these processes of peer teaching and mutual learning have their counterparts inside schools and other so-called educational institutions. It has certainly been true for some time that personal computer users have taught one another how to use their machines, despite or because of the more conventional standardized instruction available here and there. And it has been demonstrated in a variety of studies that students in conventional institutions teach one another how to deal with the constraints, requirements, and opportunities those places embody.

To take another variation on the standard model of education, some kinds of teaching and learning are, unlike the elementary and secondary education that form the core of interest for educational researchers, totally voluntary: Lessons in piano playing, tennis, and French are all like that. They take place in profit-making establishments, are often if not always individual, and have no fixed term, no credits, and no degrees. Students just take lessons until they feel they aren't getting anything out of it any more. The distribution of power

between student and teacher is so different from the stereotypical school that this is bound to be a somewhat different generic type. (See the discussion in Becker, 1986.)

We can very likely make generalizations about phenomena so defined, that is, phenomena defined as being alike in ways we think central to the way collective activity is carried on in them. A good example would be Erving Goffman's analysis (1961) of what places having the generic features of "total institutions" had in common with respect to the way their inmates (be they nuns, sailors at sea, or mental patients) had to live and the kinds of adjustments they made to living that way. Another would be his analysis of the characteristic social forms that grew up around people who had stigmas of various kinds (Goffman, 1963). The brilliance of his analyses was that you could see that, in the generic sense he had in mind, everyone had some sort of stigma, not just people who were blind or missing a limb, and every institution was, in some respects, a total institution. Understanding a generic phenomenon let you generalize about some large area of social life.

Generalizing this way is something like the Lincoln-Guba model Donmoyer finds inadequate (which I've always interpreted to be a lot more like what he recommends than he thinks it is). His changes seem to be mostly in the direction of finding ways for practitioners to apply knowledge and to take advantage of the kind of learning that comes from experience, the case study then playing the role of provider of vicarious experience.

I'd put it this way: Researchers cannot avoid devising working hypotheses about what they study and, having done that, will of course revise and amplify them to take account of further findings (as Donmoyer's schema theory suggests). But the hypotheses and ideas they develop in this way will probably be about things that are not related in any clear way to schools. If, however, you don't generalize about schools, in some crucial respect you haven't solved the problem most educational researchers have in mind when they talk about generalizing.

Schools Are Not Alike

Researchers who work in the field of education run afoul of one of the great scams of our society: the notion that things called by the same name are the same in other respects. Maybe. Maybe not. The most direct answer to the question of what all organizations called schools have in common is that they are called "schools." Whether that regularity entails any other regularities is problematic or, as we like to say, an empirical question.

It is possible that the people who assign such names do it by looking for just the regularities we would require before we would say that members of that class are alike. That isn't silly, but it certainly isn't guaranteed to be the way things are done. In fact, that usually isn't how people assign such names, although the mechanisms for doing that—for example, certification by the state—probably keep the assignments from being totally unrelated to some set of criteria.

It is also possible that, once an organization is called a school, it comes to resemble the ideal type. If you have the name, you probably have acquired an obligation to live up to the legal or commonsense requirements that go with it. If you don't you may not have any students, may be sued by the state's attorney or the U.S. Postal Service for fraud, or just be laughed at by the people you wanted to impress by taking the name. So, once you have the name, your organization may end up looking quite a bit like the ideal type.

Nevertheless, the name and the reality are at best loosely connected, and some things called schools may actually resemble places that go by other names more than they do other places called schools. Edgar Friedenberg (1959), for instance, had great fun talking about schools as prisons. Of course, he was being ironic, and many thought he was just plain snotty. But few doubted that he was talking about a real enough aspect of contemporary American schools, though many thought he was "going too far," leaving out more favorable aspects of the schools that would balance off the bad stuff he insisted on talking about. Some critics conceded that some schools were prisonlike but thought that it wasn't that many.

The example makes clear why we have trouble generalizing. If schools are supposed to be a certain way, and in reality they ain't the way they spozed to be, the people who run them are falling down on the job. But we ordinarily give such respectable people the benefit of the doubt, not requiring them to prove that their organization is what they say it is. It is one of the several privileges of respectability. That is a bad habit we have, taking the classifications and ideas of the people we study as though they were the way things really are, instead of just more data about what we are studying.

HOW CAN WE GENERALIZE?

Another complication of generalizing arises from the notion that we do not have a valid generalization unless we get the same result in every organization that fits our definition. But that's not true. You can

develop generalizations by seeing how each case, potentially, represents different values of some generic variables or processes.

The study of prison cultures furnishes a nice example. Students of prisons (e.g., Sykes, 1958) had demonstrated that, in the men's prisons they studied, inmates developed an elaborate culture. They created a convict government that took over many of the functions of keeping order in the joint; they developed quasi-markets in cigarettes, drugs, tailor-made clothing, and a variety of personal services; they organized sexual activity; they enforced a strict code of convict behavior emphasizing the necessity of never giving information about other prisoners to prison guards and officials.

Analysts of prison culture attributed these inventions to the deprivations of prison life: Deprived of autonomy, prisoners carved out a governmental structure that got some autonomy back for them and a convict code that preserved that autonomy; deprived of drugs, sharp clothes, and other goods they were used to in civilian life, they organized markets to provide those things; deprived of sex, they improvised prison-specific homosexual relationships that did not threaten their self-conceptions as macho men. The generalization was, prisoners develop a culture that solves the problems created by the deprivations of prison life.

So far, so good. Ward and Kassebaum (1965), with this theory in mind, studied a women's prison. They didn't find any of that. Quite the opposite. Even the officials of the prison complained about the lack of a convict code: The women were forever snitching on one another in a way that made a lot of trouble. There was no underground market in much of anything. Sex life was not organized in the predatory style of the men's prison; instead, the women developed pseudo-families, with butches acting as the husbands and fathers of a collection of wives and daughters. (See also Giallombardo, 1966.)

Do these differences invalidate the generalization that the deprivations of prison life lead to the creation of a prison culture? Do they mean that no generalizations about prisons are possible? Not at all. They mean that the generalizations are not about how all prisons are just the same, but about a process, the same no matter where it occurs, in which variations in conditions create variations in results. That's actually a classier form of generalization anyway.

In this case, the theory wasn't wrong, but you had to put in the right values of the variables to see how it was right. You could still say that the deprivations of prison life led to the creation of prison culture, but that this was true only if you understood that prison deprived women of different things than men. Women were not deprived of

autonomy because, on their own testimony, they had never had it; they had always lived under the protection of a man—a father, husband, or lover. They were, however, deprived of exactly that kind of protection. So they didn't develop a convict government, but they did develop a system of homosexual relationships in which one woman stood in as the masculine protector.

New women prisoners were especially afraid because, due to variations in the gender distributions of crime, men's prisons have a lot of professional criminals serving time for robbery, burglary, and other less violent crimes, while most women prisoners are in for drugs, prostitution, and typically amateur crimes of passion. There are, thus, more murderers in women's prisons, which as a result sound like more dangerous places, even to the murderers who know themselves as nondangerous (they just wanted to kill that one person who done them wrong). So even the murderers are looking for someone to take care of them.

Similarly, women's prisons typically allow inmates to buy things they want, like cosmetics and clothes, so that there is no need for the development of an underground market.

In short, women are deprived of different things, both because their lives on the outside and, therefore, their needs on the inside differ, and because the prison is run differently for them. Their culture responds to that difference. The generalization is still true, even though the results are quite different.

That's how we might go about generalizing about schools. In fact, I think we could reasonably say that it *is* how we go about generalizing about schools, whether we admit it or not. All that remains is to accept that as what we're doing and recognize it as a perfectly OK way to do it.

REFERENCES

- Becker, H. S. (1986). A school is a lousy place to learn anything in. In H. S. Becker (Ed.), *Doing things together* (pp. 173-190). Evanston, IL: Northwestern University Press.
- Becker, H. S., Geer, B., Hughes, E. C., & Strauss, A. L. (1961). *Boys in white: Student culture in medical school*. Chicago: University of Chicago Press.
- Blumer, H. (1969). *Symbolic interactionism*. Englewood Cliffs, NJ: Prentice-Hall.
- Friedenberg, E. (1959). *The vanishing adolescent*. New York: Dell.
- Fujimura, J. H., Star, S. L., & Gerson, E. M. (1987). Méthodes de recherche en sociologie des sciences: Travail, pragmatisme et interactionnisme symbolique. *Cahiers de Recherche Sociologique*, 5, 65-85.

