

Educational Researcher

<http://er.aera.net>

Alternative Arguments for Generalizing From Data as Applied to Qualitative Research

WILLIAM A. FIRESTONE

EDUCATIONAL RESEARCHER 1993; 22; 16

DOI: 10.3102/0013189X022004016

The online version of this article can be found at:
<http://edr.sagepub.com/cgi/content/abstract/22/4/16>

Published on behalf of



<http://www.aera.net>

By



<http://www.sagepublications.com>

Additional services and information for *Educational Researcher* can be found at:

Email Alerts: <http://er.aera.net/cgi/alerts>

Subscriptions: <http://er.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

Alternative Arguments for Generalizing From Data as Applied to Qualitative Research

WILLIAM A. FIRESTONE

One criticism about qualitative research is that it is difficult to generalize findings to settings not studied. To explore this issue, I examine three broad arguments for generalizing from data: sample-to-population extrapolation, analytic generalization, and case-to-case transfer. Qualitative research often uses the last argument, but some efforts have been made to use the first two. I suggest that analytic generalization can be very helpful for qualitative researchers but that sample-to-population extrapolation is not likely to be.

Educational Researcher, Vol. 22, No. 4, pp. 16-23.

The last 3 decades have seen considerable expansion in the use of qualitative methods in educational research. This development is apparent in the growing pluralism in discussion of research methods (e.g., Jaeger, 1988) and can be seen in casual perusal of the major research journals in the field. Qualitative methods are useful for understanding the perspectives of students, teachers, parents, and others; for clarifying the processes that take place in classrooms, during program implementation, and in other areas; and for generating hypotheses for testing through other methods (Lofland & Lofland, 1984; Patton, 1990). Yet there continue to be concerns about this approach. One of the more frequent criticisms, even among its advocates, is that it appears hard to generalize qualitative findings to settings other than those studied (Patton, 1990; Yin, 1989).

Generalizability is clearly not the strength of qualitative research (McGrath, 1982). Yet past reservations turn out to be overstated. In this article, I argue that qualitative methods are not at any great disadvantage although there are things researchers in this tradition can do to strengthen their case. Generalizing from data is always problematic at best. Since Hume (quoted in Campbell & Stanley, 1963), philosophers and researchers have understood that generalization requires extrapolation that can never be fully justified logically. When researchers generalize, they really make claims about the applicability of their findings to other settings. Readers must always assess those claims critically.

A researcher generally uses some mix of three broad arguments to make the case for the generalizability of findings. These claims are arguments in the sense that the author asserts a conclusion that cannot be fully proven. Yet they are more than mere rhetoric because they are built into the study design and analysis procedures. There are standards of good practice about what must be done to make these arguments plausible. When assessing the author's claim to generalizability of findings, the reader must assess how well these standards have been applied.

The three arguments are (a) extrapolation from sample to population, (b) analytic generalization or extrapolation using a theory, and (c) case-to-case translation. In practice,

these arguments have been linked to certain broad clusters of methods: sampling with survey research, analytic generalization with experimental and quasi-experimental methods, and case-to-case translation with qualitative methods. However, these associations are tendencies, not rules. Recent applications of qualitative research to applied fields have featured efforts to build on aspects of survey research (Miles & Huberman, 1984; Yin & Heald, 1975) and experimental and quasi-experimental methods (Kennedy, 1979; Yin, 1989), including their strategies for generalizing.

In the rest of this article, I first describe these three arguments for generalizing. Next, to place the weaknesses of case-to-case transfer in context, I point to some of the problems with more conventional arguments for generalizing. Finally, I assess the applicability of these two other strategies to qualitative research. What comes out of this analysis is an identification of some affinities between qualitative research and analytic generalization.

Arguments for Generalizing

The strongest argument for generalizing is usually thought to be extrapolation from a sample to a population. This belief is apparent in writing comparing different methodologies (e.g., McGrath, 1982), and it is broadly accepted by both qualitative researchers (Patton, 1990) and the general public. This argument relies on sampling and probability theory. To make the argument, one first identifies a population of interest and then draws a sample of that population to study. If the sample is drawn randomly so each member of the population has an equal opportunity of falling in, sampling theory can be used to make inferences about how closely characteristics of the sample reflect the larger population. One can quantify such inferences with the confidence interval—the range around the sample value within which the population value will fall with a given probability. One's claim is stronger when the sample is representative of the population since a highly skewed sample could be drawn by chance. For instance, a Gallup poll to predict a presidential election could consist entirely of Republicans.¹ There are many variants on simple random sampling, but the underlying argument is the same. Inferences from sample to population are strengthened by knowledge about the probability that certain kinds of cases will fall into the sample (Lin, 1976).

Probability sampling requires large populations and large

WILLIAM A. FIRESTONE is professor of education and director of the Center for Educational Policy Analysis in New Jersey, Rutgers Graduate School of Education, 10 Seminary Place, New Brunswick, NJ 08903. He specializes in qualitative research, educational policy, and teacher working conditions and professionalism.

samples. However, the logarithmic relationships among population size, sample size, and confidence interval create useful efficiencies. For instance, as populations expand above 10,000, further increases make very little difference in the size of the sample needed. Hence, national polling organizations predict presidential elections within a few percentage points with samples of only 1,000 respondents. On the other hand, beyond a certain point, very large numbers of respondents are needed to get very small reductions in the size of the confidence interval (Lin, 1976).

Analytic generalization does not rely on samples and populations. According to Yin (1989, p. 44), "in analytic generalization, the investigator is striving to generalize a particular set of results to a broader theory." To generalize to a theory is to provide evidence that supports (but does not definitively prove) that theory. Generalizing to a theory is different from generalizing to a population. When Pfeffer and Lawler (1980) found no relationship between salary and job satisfaction for college professors with no recent job offers but a strong relationship for those who had such job offers, they said something about an occupational group; they generalized to that population. At the same time, they provided further evidence for cognitive dissonance theory; they generalized to a theory. Generalizing to a theory differs from generalizing to a population more when the theory is intended to apply across a wider range of specific populations and settings.

When one generalizes to a theory, one uses the theory to make predictions and then confirms those predictions. In a specific study, predictions hold under specific conditions. If the predictions hold only under those conditions, they become *scope conditions* that limit the generalizability of the theory. Cognitive dissonance theory is very broad because it can be applied in a wide variety of conditions. The scope of classroom management theories is limited in comparison. What complicates matters is that one uses theory—sometimes the same theory, sometimes different theory, sometimes formal theory, sometimes common sense—to anticipate or recognize relevant conditions that may affect the application of a theory.

The importance of these additional conditions becomes apparent in the ways one uses analytic generalization to increase confidence in the theory. According to experimentalists (e.g., Kiess & Bloomquist, 1985), a single study generally provides weak support for a theory. For them, studies must be replicated. Yet not all replications support a theory in the same way. Replications under conditions that exactly repeat the original study are most useful for establishing reliability. When conditions vary, successful replication contributes to generalizability. Similar results under different conditions illustrate the robustness of the finding.

Results that differ from the original study can also support a theory (Yin, 1989). That happens when the initial theory is set within a web of more complex and far-reaching theories. Replications that are similar to the first study in ways specified by these other theories should yield similar results. However, where ancillary conditions vary, the results should differ in predictable ways. For instance, Tobin (1986) found that lengthening the time a teacher waits for a response after asking questions increased the elaborateness of student answers. Most people would expect a replication to lead to similar results. However, if one theorized that the relationship between teacher waiting and student

answers depended on the cooperativeness of the students, one could identify conditions where longer teacher wait time would not change student answers but increase off-task behavior. The point of this example is that sometimes analytic generalization attempts to show that a theory holds broadly across a wide variety of circumstances, but sometimes it identifies the scope of a theory—that is, the conditions under which it applies.

Another way to increase confidence in a finding through analytic generalization is to anticipate threats to doing so, what Cook and Campbell (1979) call threats to external validity. Suppose for a moment that one conducts an experiment that finds that whole language instruction increases student comprehension in the lab school of one's university. There are at least three reasons why this finding might not generalize beyond the study. The students might be especially bright (interaction of selection and treatment); the teachers might have strong previous commitments that are compatible with whole language instruction and that contribute to an especially adept implementation of the approach (interaction of setting and treatment); or something else might be happening at the same time that increases comprehension (interaction of history and treatment). The whole field of quasi-experimental design has been developed to help anticipate these threats and find ways to organize one's data collection and analysis to minimize their effects.

Although threats to external validity come from knowledge about measurement and study design, the principle is the same as with replication: The theory in question is embedded in a broader web of theories extending into common sense knowledge. One uses these ancillary theories to link specific study findings to the theory of interest.

While the first two arguments for generalizing have a relatively long history in the social sciences, case-to-case transfer comes out of recent efforts to use qualitative methods for program evaluation. It is the least familiar argument for generalization to educational researchers, but it is becoming more prevalent, especially among those with a qualitative bent (Erickson, 1992; Lincoln & Guba, 1985). It is important to note, however, that while there is considerable overlap between the study of cases and qualitative research, the two are not the same. Qualitative research focuses on up-close observation of behavior in settings as well as interviewing people in those settings and collecting and analyzing documents and artifacts. Its purposes are to describe those settings and understand the definitions of those settings held by people in them (Firestone, 1987; Van Maanen, 1982). The emphasis on up-close description fits nicely with an interest in cases, but there are traditions in qualitative research in which the case becomes difficult to identify and delimit (Strauss & Corbin, 1990) as well as case studies that are largely quantitative.

Case-to-case transfer occurs whenever a person in one setting considers adopting a program or idea from another one. For instance, a district testing coordinator reading a case study on alternative forms of assessment used in another district might try to decide whether to adopt those practices within the district. The testing coordinator must know what conditions increase the utility of the reported results for the coordinator's own district.

According to Kennedy (1979), well-developed traditions of transferring learning from one case to another have been

developed in law, medicine, and clinical psychology. When deciding whether to apply a precedent, judges compare precedents to the current case in terms of four criteria. The first is the material facts. If a curriculum coordinator is using a qualitative case study of a program to decide if that program should be used in the coordinator's district, material facts might include the fit between the content of the program described in the case study and the subject matter area of local interest, the similarity of the students in the two situations, and so on. A second legal criterion is appropriateness. This requires making value judgments about the fairness or, in the case of educational programs, the rightness of the goals for the program described in the current situation.

The third criterion is the reason for the decision. As Kennedy (1979) points out, the study's author may conclude that the program failed because it works only with certain groups of students. However, if the curriculum coordinator wants to help just such a target group, that conclusion provides reason to use the program. The final criterion is the generality of the decision. Was the precedent decision based on a technicality or on more fundamental grounds? Does the case study suggest that the program worked because of some specific circumstances or that some more broadly applicable principle was at work? The identification of more general processes or findings encourages broader use of the study conclusions.

While transfer of findings from one case study to another is done by the reader, the researcher has an obligation to provide a rich, detailed, thick description of the case. This is because the researcher's theories about the conditions that affect the applicability of study conclusions are less important than those of the reader. One cannot know the situations in which readers are likely to consider applying study findings. Therefore one must describe a broad range of background features, aspects of the processes studied, and outcomes so readers have enough information to assess the match between the situation studied and their own, especially since their situations might be quite different.

Stake (1978, p. 7) concurs in the utility of case-to-case translation. When "there is a need for generalization about that particular case or generalization to a similar case rather than generalization to a population of cases," he believes "the demands for typicality and representativeness yield to needs for assurance that the target case is properly described. As readers recognize essential similarities to cases of interest to them, they establish the basis for naturalistic generalization." For Stake, however, the link between the case described and the application setting comes less through the formal matching criteria important to Kennedy than through the tacit knowledge of the reader. Yet this reasoning, too, leads to a need for thick description to help the reader bridge the gap between the written case and the application setting.

While Kennedy finds case-to-case transfer deficient in comparison with analytic generalization, Lincoln and Guba (1985) believe the former is the only defensible strategy. For them, the burden of proof for transferability lies less with the investigator than with the reader. The investigator's responsibility ends with providing sufficient descriptive data to make such similarity judgments possible.

In sum, the sample-to-population argument is strengthened by the appropriate application of sampling theory.

Analytic generalization is facilitated by specifying the conditions under which a study is done and their relevance to multiple theories. That knowledge is used to create controls, quasi-experimental designs, or replications that strengthen generalization. Case-to-case transfer is enhanced by thick description that allows assessment of the applicability of study conclusions to one's own situation.

Limitations to Conventional Arguments

Case-to-case transfer goes particularly well with qualitative research, which attends so carefully to the particulars of the phenomenon under study. However, this argument is especially unsatisfying to many researchers partly because it is unconventional, partly because the responsibility (and rewards) of drawing broader implications from a study are shifted to others, and partly because it is unconventional, partly because the researcher's influence over the use of study findings is so clearly limited. While educational researchers may be more familiar with other arguments, it is worth reviewing some of the limits inherent in them.

Although sample-to-population inference has long been the touchstone for generalization in social and behavioral science, some researchers have never been impressed with this argument. An example of generalization without sampling that Campbell finds telling enough to repeat (1969, 1986) is the hydrolysis of water. The experiment showing that water could be decomposed into hydrogen and oxygen atoms was done without drawing samples of water or the copper wire that ran the charge through it; yet it has withstood the test of time. Moreover, Campbell (1986) concludes that most high-quality scientific program studies have used illustrative rather than random samples, an approach he endorses while arguing that generalization is not necessarily lost as a result.

The major difficulty with the sample-to-population argument is that it is difficult to sample all the things that must be sampled to make a generalization. Consider all the studies of the real-world effects of "treatments," whether they be instructional strategy, curriculum, or social program. According to Cronbach (1982), there are at least four classes of things one would want to sample in such a study: the units affected (students, classrooms, schools, etc.) the treatments, the outcomes (effects as conceptualized and as measured), and the settings, including culture, historical period, and a variety of other dimensions. Randomly selecting treatments, that is, various ways to do whole language instruction, or observations, such as measures of higher-order thinking (or even various concepts of cognitive effects one might want to impact), is almost never done. One typically makes these choices analytically and takes those decisions into account when generalizing the results. Moreover, time can be an especially challenging dimension to sample. One can sample forward through the use of longitudinal designs, but to go backward, one is limited to retrospective histories and existing documents that may not fit the study's analytic framework.

Conventional sampling theory is very useful for choosing the units to study, but only when one has fully itemized the population. Mobile populations (immigrants or migrant students) and populations that hide their activities (such as criminals) are difficult to sample. Moreover, in order to know if a sample is representative of a population, one must have information on how theoretically relevant char-

acteristics of that population are distributed. Without such information, it is hard to know how representative a randomly selected sample is. Finally, careful sampling is not particularly helpful when the population sampled is not the population of interest. The technical apparatus of sampling loses much of its value when the setting the reader is interested in is not included in the population sampled and one must use informal knowledge to make the connection.

A different problem plagues analytic generalization. The theory of interest is embedded in and depends upon a whole set of additional assertions to specify the conditions under which it holds. As an example, Cronbach (1982) describes the grim history of aerial bombing. Experience in the 1920s in what was then "Mesopotamia" indicated that bombing villages was effective for putting down revolts. This experience, he says, was the root of the British strategy for bombing Germany during World War II. The failure of that strategy in the latter case shows how characteristics of those "treated"—that is, bombed—and the setting for the treatment—that is, qualities of leadership and political organization—can affect the outcome.

The bombing example is a dramatic instance of how a variety of factors can interact with the main variables to cloud or condition the effect. Undetected and uncontrolled interactions can lead one to falsely believe that observed main effects are significant and can also mask potentially important ones. Interactions are the "Achilles' heel of the behavioral sciences" according to Cook & Campbell (1979, p. 237). Cronbach's (1975) work on aptitude-treatment interactions in the classroom and various attempts to develop contingency theories of organizational design (Galbraith, 1973) illustrate how extensive such interactions are. Quasi-experimental researchers have successfully developed a variety of designs to control for conditions that might interact with the variables specified in the theory of interest (e.g., Cook & Campbell, 1979). There are model designs to account for many potentially interacting factors although the sheer number can lead to unwieldy research plans and reduce precision of measurement. The difficulty for the researcher is specifying which additional factors must be taken into account. This is a substantive, judgmental task upon which experts in a field might disagree.

One other important point about both conventional arguments for generalizing is that they are of limited utility for making decisions in specific cases. Cronbach (1982) uses the example of the Coleman comparison of public and private schools as an example (Coleman, Hoffer, & Kilgore, 1981). These findings hold on the average, he suggests. However, they do not help the non-Catholic parent decide whether to send a child to a public or private school, for the parent must decide between two *particular* schools while the study presents average differences between categories.

More generally, applications from a study require going from a population to a specific instance or, if analytic generalization is used, from a theory supported by the research to a specific set of conditions that might be different from those studied in ways that are important, but not specified. Cronbach (1982) suggests that conclusions based on both sample-to-population and analytic arguments are most useful for decision makers like members of Congress or governors who deal with large aggregates where unmeasured conditions may average out. Decision makers like parents and teachers whose scope of application is much smaller

must engage in a kind of transfer much like that described by Kennedy (1979) and Lincoln and Guba (1985) even if the researchers make other kinds of arguments.

Conventional Arguments and Qualitative Research

Applied qualitative researchers seek to increase the capacity to generalize from their studies. One step is the spread of multisite qualitative research (Herriott & Firestone, 1983). Such studies expand one's repertoire beyond case-to-case transfer but raise a question about whether analytic or sample-to-population arguments are more useful.

One way to apply the analytic argument is to think of cases as analogous to experiments (Yin, 1989). There is an important difference. The experimenter develops a theory deductively, anticipates plausible rival explanations, and designs controls for them. As Campbell explains (1975), the qualitative researcher develops a theory inductively through a series of successive approximations, each of which is compared with the available data. Important parts of this process include imagining plausible rival explanations and seeking out evidence to assess them and more generally seeking out negative evidence for the preferred theory of the moment. Usually, the result is a fairly complex analysis where some of the "rival" explanations are built into the overall theory of the case. In both the experiment and the qualitative case, however, a theory is developed and in some sense "tested" in light of rival explanations.

There are a number of ways for the qualitative researcher to generalize analytically. Some are possible within single cases. Goetz and LeCompte (1984) have extended the Campbell-Stanley-Cook mode of thinking about rival explanations to ethnographic research. They suggest that predictable threats to generalizability can be organized under the broad headings of selection, setting, history, and construct effects. In keeping with qualitative approaches, which avoid active manipulation, they recommend searching for data within cases to explore such hypotheses rather than than using upfront design strategies. Otherwise, the logic is very similar.

One can also select single cases to maximize their use for generalizing about theories (Patton, 1990; Yin, 1989). One way is to select a critical case, one that is particularly fruitful for comparing different theories. Allison's (1971) study of the Cuban missile crisis used one dramatic event to compare three prominent theories about decision making. Another approach is the deviant case, one that clearly diverges from a well-established generalization. One such generalization was that unions tend to be run oligarchically with little sharing of power with the rank-and-file. Lipset, Trow, and Coleman's (1956) study of the typographer's union identified the internal mechanisms that made this one case much more democratic than previous theory predicted. Such cases are particularly elegant ways for extending and refining theories. They also help to generalize by specifying the conditions under which a theory holds. However, they tend to be rare, in part because they require well-established theories upon which to build.

The fundamental insight linking multisite studies to the analytic strategy is to think of each case as a replication (Yin, 1989). Then sites can be selected to maximize generalization. This requires reference to the informal network of theories in which the theory of interest is embedded. Kennedy (1979) suggests that the sample cases should maximize

variation on relevant attributes—that is, those that might be expected to affect the phenomenon of interest. When selecting cases for such studies, one always has too many characteristics to take into account. The counsel here is to maximize diversity. The cases should also be as like the population of interest as possible. If one were particularly interested in implementing a program for early elementary students, including K-8 schools might blur the distinction between those students and middle schoolers. Conversely, the sample should have few unique cases. Special circumstances like unusually strong school leadership or access to special fiscal and training resources might lead to unusually optimistic results.

The case selection strategy will vary depending on what the investigator knows about the issues; this will change over the course of a study (Strauss & Corbin, 1990). When selecting all cases early on, one tries to maximize differences among cases as much as possible but opportunistically since relevant sources of variation are not well understood; this is what Patton (1990) calls extreme case sampling. Its purpose, however, is to facilitate the discovery of relevant explanations, not generalization.

Later, the researcher becomes more strategic as the inductive generation of explanations identifies key conditions that may affect the theory being formulated. Cases can be selected to strengthen generalization by making them vary on attributes likely to interact with those studied within cases. Comparison of cases can then identify the conditions under which those explanations hold. Such comparisons are essential for generalizing through replication. However, they require substantial knowledge of each case, and systematically comparing a large number of cases becomes difficult.²

Researchers who are primarily interested in comparing cases and want a more systematic approach than that entailed in replication often use methods that look more like survey research and appear to rely on the sample-to-population argumentation to generalize. These studies pull together information on a wide variety of cases, rate the cases in terms of "variables," and then look at the associations among the variables using displays or statistics. This strategy has a long history in comparative anthropology, political science, and sociology. Yin and Heald (1975), among others, have used it to synthesize results of cases written up previously. Their approach is a sort of meta-analysis of qualitative studies. They give a great deal of attention to measuring and controlling for study quality. This "associational" approach has also been used in recent applied research by researchers who collect data from multiple sites as a part of one study (e.g., Smith & Robbins, 1982).

This associational approach helps the researcher identify broad patterns across a wide variety of cases. Because it generates and identifies relationships among variables, this approach has been characterized as more deductive and oriented toward broad generalizations than others that stick closer to whole cases (Ragin, 1987). However, the process of reviewing case data to generate variables and identify relationships can be quite inductive and exploratory. Also, because cases are reduced to variables, one can analyze larger numbers of them than is possible when each case must be examined in depth. This advantage is extended when one uses existing case records.

The weakness of this approach is that cases lose their identity. The dynamic processes through which events unfold

to create the outcome of interest in each site are lost. One understands little about why A and B occur together, only that they do. Moreover, there is a tendency to ignore exceptions in order to identify the main patterns. The associational approach adopts the probabilistic way of thinking used in quantitative research, so exceptions are less important than when one works analytically to develop patterns where all anomalies must be accounted for (Campbell, 1975; Robinson, 1969). Finally, correlational thinking treats variables in an additive fashion when, in fact, combinations of factors may be crucial (Ragin, 1987). One might be able to bypass this limitation to some extent by testing for interactions, but usually both the number of cases and the quality of data preclude such sophisticated analyses.

One way to partially avoid losing case identity comes through the matrix and display strategies developed by Miles and Huberman (1984). These matrices provide visual ways to see relationships among variables, but the authors' emphasis on building up from individual cases encourages users to develop a firm understanding of case particulars. Moreover, some display strategies encourage analysts to bring along a great deal of information about each case in the way they fill in cells rather than to simply rank or rate cases in terms of variables.

For purposes of generalization, the strength of the associational approach is that by including moderate to large numbers of cases and emphasizing broad cross-site patterns, the researcher can protect against the idiosyncrasy that may appear in single-case studies. However, none of the arguments for generality can be easily applied. Because cases lose their identity, case-to-case transfer becomes problematic. Moreover, there is no potential for replication. Without internal analysis, cases cease to be analogous to experiments, so the "study" is conducted on the set of cases. Thus, analytic generalization is weakened. Yet the number of cases is usually too small to apply the sample-to-population argument. Even larger samples are rarely drawn randomly. If the sample is not outside the control of the investigator or drawn for purposes of convenience, it is usually drawn using some theoretical criteria.

Ragin (1987) used Boolean algebra to develop an approach that allows one to compare large numbers of cases as systematically as with the associational approach while remaining in the analytic mode. In effect it allows one to aggregate replications into more parsimonious patterns than the more informal case-to-case comparisons permits.

This approach is applied after one understands one's cases well enough to have identified a fairly limited set of dichotomous variables that can be used to explain an outcome of interest. The first step is to code the cases in terms of the variables of interest and arrange the results in a truth table. Table 1 is a hypothetical truth table for a study of factors affecting program implementation. From the cases, three variables were identified: the presence or absence of teacher participation, the presence or absence of a need for the program being implemented, and the presence or absence of principal leadership. Within the table, "1" signifies the presence of a variable and "0" signifies its absence. The table itself lists the eight possible combinations of independent variables, whether each possible combination leads to successful implementation or not, and the number of cases fitting each combination. The table shows four combinations leading to successful implementation. The rules of Boolean

Table 1
A Hypothetical Truth Table for a Study of Program Implementation

| Independent Variables | | | | |
|-----------------------------------|-----------------------------|--------------------------|---------------------------|-----------------|
| Teacher participation in planning | Strong principal leadership | Program fits school need | Successful implementation | Number of cases |
| 1 | 0 | 1 | 1 | 2 |
| 0 | 1 | 0 | 1 | 5 |
| 1 | 1 | 0 | 1 | 4 |
| 1 | 1 | 1 | 1 | 3 |
| 1 | 0 | 0 | 0 | 4 |
| 0 | 0 | 1 | 0 | 2 |
| 0 | 1 | 1 | 0 | 2 |
| 0 | 0 | 0 | 0 | 1 |

Note. 1 indicates variable is present; 0 indicates it is absent.

algebra are used to come up with a more concise list of combinations promoting implementation: teacher participation when there is strong need for the program and strong principal leadership when the need for the program is weak.

While the Boolean manipulations take too much space to explain, the truth table itself illustrates some important characteristics of this approach. First, it ensures that the researcher systematically examines all cases with respect to all variables simultaneously. When analyzing cases as replications without such a table, it is easy to overlook some comparisons of variables. Associational techniques tend to look at pairs of variables unless one has enough cases to control for additional variables. Moreover, a large number of cases is needed to examine several variables at once.

Second, the emphasis is on combinations of variables—that is, what happens when strong leadership occurs in the presence of a need for the program and with both the presence and absence of teacher participation. As a result, the analysis attends to interactions. Moreover, it is holistic, at least within the framework of the variables in the theory, so it is more in keeping with more conventional qualitative analysis.

Third, the logic of the truth table forces the researcher to identify necessary variables in ways that are more like Campbell's within-case analysis than the associational approach. Each combination of independent variables is supposed to be associated with only one value for the dependent variable—that is, either presence or absence. According to Ragin (1987), when a large number of cases are present for any combination, a few exceptions are acceptable, but when a substantial proportion deviates from the typical, there is a problem. While he proposes several ways to solve such problems, analytic induction suggests that such a split indicates the presence of an unidentified variable that differentiates two sets of cases that seem alike otherwise. The rules of the truth table encourage the analyst to revisit the data to find that variable.

While the truth table has many advantages, it is reductive in two ways that more conventional qualitative research is not. First, by reducing cases to a series of dichotomous variables—even if in combination—the capacity to analyze process, which is a strength of qualitative case studies, is

lost. Second, the definitions of the situation and belief systems of those studied are at best reduced to simplified categories if not entirely ignored in favor of conceptually neat variables developed by the researcher. For highly analytic efforts at theory development, these losses may be minor. Yet they are some of the major reasons why people turn to qualitative research. In education it remains to be seen whether this extremely formal approach can be included as part of a more traditional qualitative study or whether it tends to drive that part of the work out.

The Boolean approach has both advantages and disadvantages for generalizing. The big advantage is that the study clearly generalizes to a "theoretical space" represented by the truth table. In that sense, this approach is clearly in the tradition of the analytic argument and does not rely on sample-to-population reasoning. Moreover, the theoretical space is more systematic than would be the case with less formal means.

There are at least two problems, however. The more obvious one is the recurring issue for analytic arguments for generalizations; there are likely to be a number of background conditions and unstated assumptions upon which generalization depends. For all the systematicness forced by the truth table, it is virtually impossible to assure that all relevant variables are covered in any one study.

The second problem concerns missing data. The difficulty occurs when a row in the table is not represented in the data. Such instances are likely; they often reflect the organization of society. A study looking at the effects of gender and school level on teacher attitudes, for instance, would be hard pressed to find a predominantly male elementary school. Ragin (1987) suggests a number of assumptions that can be made when using Boolean algebra to manipulate truth tables with missing combinations. For instance, one might assume that such combinations can be treated as those where the dependent variable is absent. His argument is that if missing combinations are unlikely to exist, they are unlikely to show presence of the dependent variable. This example points out that how one handles missing data depends upon substantive assumptions. His argument makes sense only if the absence of a combination does not result from case selection procedures. The strength of this ap-

proach is that it does point out missing combinations that might be important. However, how one should handle missing combinations could be subject to debate.³

Conclusion

Generalization is always based on extrapolation. Different research traditions have developed their own arguments to justify such extrapolation. In general, the argument most closely associated with qualitative research, case-to-case translation, compares favorably with the others although it does so by passing the responsibility for application from the researcher to the reader. Thus, its claims, while legitimate, are weaker.

Efforts to strengthen the generalizability of qualitative research have focused on incorporating either sample-to-population extrapolation or analytic generalization. My assessment of these options is that the long tradition of looking for relationships among variables in a sample of cases has some utility, but it does not adequately apply sample-to-population extrapolation to qualitative work. The samples are too small, and questions remain about how they were selected and what the populations are.

Analytic generalization has more promise, partly because there are more ways to make links between cases and theories. One can look for threats to generalizability within cases. Critical and deviant cases can be used to explore or extend existing theories. Multicase studies can use the logic of replication and comparison to strengthen conclusions drawn in single sites and provide evidence for both their broader utility and the conditions under which they hold. Ragin's Boolean approach helps to examine multiple replications more systematically.

One issue that deserves further exploration is the relationship between the scope of application intended by the reader and the argument employed. When the application is limited to one or a few specific settings, differences between the three arguments seem to be reduced or even to favor case-to-case transfer. The reader must make inferences about many factors that may be left unexplored in studies relying on sample-to-population or analytic generalization. These conditions may be imperfectly described through a design stressing case-to-case transfer, but they are more likely to receive attention through this approach, with its stress on thick description. Factors unimportant to the researcher may be critical to the reader. They are more likely to show up in designs stressing the careful depiction of the setting studied. When the application is broader and findings must only hold "on average" and exceptions are allowed, more conventional arguments appear to be better than case-to-case transfer.

The argument for qualitative research has never been that its claims for generalizability are exceptionally strong. Qualitative research is best for understanding the processes that go on in a situation and the beliefs and perceptions of those in it. Still, qualitative researchers can do things to increase the broad applicability of their findings. Some of these—like providing rich, "thick" description—contribute to case-to-case reasoning. Others—like intentionally sampling for theoretically relevant diversity and replicating cases through multisite designs—are particularly useful in a more analytic approach. In any event, qualitative methods should not be avoided because of the fear that their claims for broad relevance are especially weak. That is not the case.

Notes

I want to thank Steve Barnett, Jim Giarelli, Nobuo Shimahara, Jeff Smith, and three anonymous reviewers for their comments on earlier drafts of this article.

¹A reviewer of an earlier draft pointed out that sometimes one cannot know how skewed a sample is because the population parameters are unknown. The implications of this problem are discussed below.

²This discussion seems to assume that one must specify well-delineated "cases" to sample—for example, classrooms, schools, and districts. That is not always the case. Advocates of grounded theory have developed a rather sophisticated approach to theoretical sampling for qualitative research, but they tend to look at settings that are at best ambiguously bounded, like instances of pain management where neither the spatial nor the temporal limits are clear (e.g., Strauss & Corbin, 1990).

³One implication of this need to avoid empty cells is that the Boolean approach can require a large number of cases. As one adds variables, the number of rows grows exponentially. Even a moderately complex theory with five variables has 32 rows and should have at least that many cases.

References

- Allison, G. T. (1971). *Essence of decision: Explaining the Cuban missile crisis*. Boston: Little, Brown.
- Campbell, D. T. (1969). Prospective: Artifact and control. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research*. New York: Academic Press.
- Campbell, D. T. (1975). Degrees of freedom and the case study. *Comparative Political Studies*, 2, 178-193.
- Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis*, no. 31 (pp. 67-78). San Francisco: Jossey-Bass.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Coleman, J. S., Hoffer, T., & Kilgore, S. (1981). *Public and private schools*. Chicago: National Opinion Research Center.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.
- Cronbach, L. J. (1982). *Designing and evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Erickson, F. (1992). Why the clinical trial doesn't work as a metaphor for educational research: A response to Schrag. *Educational Researcher*, 21(5), 9-11.
- Firestone, W. A. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher*, 16(7), 16-21.
- Galbraith, J. (1973). *Designing complex organizations*. Reading, MA: Addison-Wesley.
- Goetz, J., & Lecompte, M. (1984). *Ethnography and qualitative design in educational research*. New York: Academic Press.
- Herriott, R. E., & Firestone, W. A. (1983). Multisite qualitative policy research: Optimizing description and generalizability. *Educational Researcher*, 12(2), 14-19.
- Jaeger, R. M. (Ed.). (1988). *Complementary methods for research in education*. Washington, DC: American Educational Research Association.
- Kennedy, M. M. (1979). Generalizing from single case studies. *Evaluation Quarterly*, 3, 661-678.
- Kiess, H. O., & Bloomquist, D. W. (1985). *Psychological research methods: A conceptual approach*. Boston: Allyn & Bacon.
- Lin, N. (1976). *Foundations of social research*. New York: McGraw-Hill.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Lipset, S. M., Trow, M., & Coleman, J. S. (1956). *Union democracy: The inside politics of the International Typographical Union*. New York: Free Press.
- Lofland, J., & Lofland, L. H. (1984). *Analyzing social settings: A guide to qualitative observational research*. Belmont, CA: Wadsworth.
- McGrath, J. E. (1982). Dilemmas: The study of research choices and dilemmas. In J. E. McGrath, J. Martin, & R. A. Kulka (Eds.), *Judgment calls in research* (pp. 69-102). Beverly Hills, CA: Sage.
- Miles, M., & Huberman, A. M. (1984). *Qualitative data analysis*. Newbury Park, CA: Sage.

- Patton, M. G. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Pfeffer, J., & Lawler, J. (1980). Effects of job alternatives, extrinsic rewards, and behavioral commitment on attitude toward the organization: A field test of the insufficient justification paradigm. *Administrative Science Quarterly*, 25, 38-56.
- Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.
- Robinson, W. S. (1969). The logical structure of analytic induction. In G. J. McCall & J. L. Simmons (Eds.), *Issues in participant observation: A text and reader* (pp. 196-205). Reading, MA: Addison-Wesley.
- Smith, A. G., & Robbins, A. E. (1982). Structured ethnography. *American Behavioral Scientist*, 26, 45-61.
- Stake, R. E. (1978). The case study method of social inquiry. *Educational Researcher*, 7(2), 5-7.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Tobin, K. (1986). Effects of teacher wait time on discourse characteristics in mathematics and language arts classes. *American Educational Research Journal*, 23, 191-120.
- Van Maanen, J. (1982). Introduction. In J. Van Maanen, J. M. Dabbs, & R. R. Faulkner (Eds.), *Varieties of qualitative research* (pp. 11-30). Beverly Hills, CA: Sage.
- Yin, R. K. (1989). *Case study research: Design and methods* (2nd ed.). Newbury Park, CA: Sage.
- Yin, R. K., & Heald, K. A. (1975). Using the case survey method to analyze case studies. *Administrative Science Quarterly*, 20, 371-381.

Received June 23, 1992

Final revision received, February 22, 1993

Accepted March 2, 1993

—UPDATE—

Call for Papers

Texas Researcher seeks manuscripts of 30 pages or less by August 31, 1993. Contact J. E. Gonzalez, Texas Center for Educational Research, P.O. Box 2947, Austin, TX 78768-2947. Phone: 512-467-0222.

The Association for Educational Research of the American Association for the Advancement of Slavic Studies solicits papers for a June 1994 conference tentatively titled **Contemporary Education in Eastern Europe, Russia, and the Soviet Successor States: Common Legacies and the Struggle for Reform**. Send proposals to Mark S. Johnson, AER-AAASS, 246 Hawthorne, Memphis, TN 38112.

Scholarships

To apply for a **Rotary Foundation Scholarship** for an academic year's study in another country, contact a local Rotary club. Although many clubs have earlier application deadlines, the final deadline in July 15, 1993.

New Program

A 3-year **evening training program in art therapy** is now available in Toronto. Write Toronto Art Therapy Institute, 216 St. Clair Ave. West, Toronto, Ont., Canada M4V 1R2, or call 416-924-6221.

Awards

The **William T. Grant Foundation** offers five awards for research contributing to the understanding of children, adolescents, and youth. Each award totals \$175,000 and is for 5 years. Write Faculty Scholars Program, William T. Grant Foundation, 515 Madison Ave., New York, NY 10022-5403. Application deadline: July 1, 1993.

Member Activities

President Clinton has nominated **Marshall Smith** to be undersecretary of education. Dr. Smith is currently dean of education at Stanford University. The president has also nominated **Sharon Robinson** for assistant for research and improvement for the Department of Education. Dr. Robinson is now director of the National Education Association's National Center for Innovation.

Conference

Automated Design for Interactive Multimedia, the Fifth Annual Utah State University Instructional Technology Summer Institute, will be held August 18-21, 1993. Contact Teresa W. McKnight, Summer Institute, Department of Instructional Technology, Utah State University, Logan, UT 84322-2830. Phone: 801-750-2779; fax: 801-750-2693.

Member Publications

Anne DiPardo, *A Kind of Passport: A Basic Writing Adjunct Program and the Challenge of Student Diversity*, NCTE, 1993.

Thomas B. Greenfield and Peter Ribbins (Eds.), *Greenfield on Educational Administration*, Routledge, 1993.

J. M. Johnston and **H. S. Pennypacker**, *Strategies and Tactics of Behavioral Research*, 2nd edition, and *Readings for Strategies and Tactics of Behavioral Research*, LEA, 1993.

James S. Kaminsky, *A New History of Educational Philosophy*, Greenwood, 1993.

Dennis E. Mithaug, *Self-Regulation Theory: How Optimal Adjustment Maximizes Gain*, Praeger, 1993.

Lee Owens and Jennifer Barnes, *Learning Preference Scales: Handbook and Test Master Set*, Australian Council for Educational Research.

Theresa Perry and **James W. Fraser** (Eds.), *Freedom's Plow: Teaching in the Multicultural Classroom*, Routledge, 1993.

George Willis, **William H. Schubert**, **Robert V. Bullough, Jr.**, **Craig Kridel**, and **John T. Holton** (Eds.), *The American Curriculum: A Documentary History*, Greenwood, 1993.