11. CASE STUDY EVALUATIONS: A DECADE OF PROGRESS?

ROBERT K. YIN

The American Evaluation Association's tenth-anniversary theme, "A Decade of Progress," was the inspiration and point of departure for the present article. Leonard Bickman invited me to reflect on the use of case studies in evaluation. In considering the use of case studies, the tenth-anniversary theme quickly became a question rather than an assertion: "Has the case study produced a decade['s worth] of progress?" This article addresses the question first by defining the case study method, then by examining the use of the case study method from a historical perspective, and finally by commenting on the progress (or lack of progress) during the past decade (roughly 1987 to 1997).

DEFINITION OF THE CASE STUDY METHOD

Critical to the discussion is the definition of the case study method. As will be pointed out shortly, two different types of research have been confused within the rubric of the case study method. Selecting one or the other type will yield different interpretations regarding the possible progress over the decade's time. Therefore, for the sake of discussion, the case study method may be briefly profiled as follows (Yin, 1994a, 1997).

A Three-Featured Profile

First, the method depends on the use of—and ability to integrate in converging fashion (some would say "triangulate")—information from multiple sources of evi-

From *New Directions for Evaluation,* 1997, *76*, 69–78.

dence.The evidence may include direct observations, interviews, documents, archival files, and actual artifacts. The facts and conclusions for the case study will be built around the consistency of data from these sources, and these facts and conclusions may be expressed in both quantitative and qualitative terms.

Second, the method implicitly assumes a richness of data because a case study is intended to examine a phenomenon in its real-life context. Often, the boundary between the phenomenon and the context is not sharp, and inherent in all case studies is the potentially important influence of contextual conditions. A major investigative concomitant—usually taken for granted—is the need to collect case study data in the field, thereby collecting data about the context, although under unusual circumstances a case study can be conducted from library and secondary sources alone. A major technical concomitant is that case studies will always have more variables of interest than data points, effectively disarming most traditional statistical methods, which demand the reverse situation.

Third and last, the case study method includes research that contains single case studies as well as multiple-case studies. The process of generalizing the results of either type of case study depends on the development, testing, and replication of theoretical propositions (analytic generalization)—rather than any notions based on the selection of numeric samples and extrapolating to a population (statistical generalization). Especially helpful is the specification and testing of rival theories or explanations, which can even take place within a single case study; in a multiple-case study one possible rationale for case selection is that certain cases have been included because they represent rivals.

Methods Falling Within And Outside The Profile

Profiling the case study in this manner provides a broad umbrella for different styles of case study research, including those based on differing philosophies of science. For instance, Bob Stake's recent book characterizes my case study research as "quantitative," appearing to contrast strongly with his own "qualitative" approach (Stake, 1995). However, examination of both approaches reveals similar ingredients. Although the qualitative approach gives less attention to multiple-case situations, it clearly draws on the same multiple sources of evidence and is concerned with the richness of case and context. Stake also agrees that the matter of defining the "case" requires close attention. Further, in discussing case study data collection and analysis, he devotes an entire chapter of his book to triangulation.

At the same time, the profile excludes certain methods that have sometimes been confused with case study research and evaluation.The primary exclusion is the classic ethnographic study—commonly using the participant-observer method (Jorgensen, 1989). Such a study traditionally focused on a preliterate society, resulting in evidence based mainly on observations and discussions but with little opportunity to rely on documentary or archival records. Ethnographic methods have been used in a variety of contemporary settings (Fetterman, 1989), including the study of organizations (such as Leonard-Barton, 1987). For evaluations, two advocates of the ethnographic method note that its strength is maximized where a strong clash in

values permeates an organization or project (Lincoln & Cuba, 1986). However, because the participant-observer is limited in the ability to cover multiple events occurring at the same time, many ethnographic studies also tend to be studies of small groups within a culture (or organization), rather than systematic coverage of the whole culture (or organization). If more than participant-observation is used in doing an evaluation, the resulting study may begin to resemble and be considered a case study. To this extent, judgments about inclusion or exclusion must, as always, appreciate the actual array of techniques being used, not just broad labels.

The Importance Of The Profile As A Statement About Case Study Design

In her major historical overview of case studies in American methodological thought, Platt (1992) characterized our profile of the case study method as giving greater emphasis to case study *design* rather than *data collection. The* distinctiveness of the design, especially with the number of potentially relevant variables far exceeding the number of data points (often, only a single data point or case), forces investigators to use different strategies for establishing internal, external, and construct validity, compared to experimental or quasi-experimental research. Likewise, the need to pursue analytic and not statistical generalizations means that cross-case strategies must go beyond merely counting the number of cases, as if they were a sample of anything.

At the same time, the basic profile should not be construed as ignoring issues of data collection. Case study investigators must be intensely concerned with collecting data in a reliable and rigorous manner. In doing data collection, case study investigators also must struggle with the problem of divulging identities or maintaining the confidentiality and anonymity of sources and even of the case itself.

THE USE OF THE CASE STUDY METHOD FROM A HISTORICAL PERSPECTIVE

From a historical perspective, Platt (1992) traces the practice of doing case studies back to three strands of research during the early twentieth century: the conduct of life histories, the work of the Chicago school of sociology, and casework in social work. She then shows how participant-observation emerged as a common data collection technique in doing these case studies. However, over time the data collection technique eventually became confused with the entirety of the case study method. The effect of this confusion on social science was dramatic, as traced by Platt. Prior to 1970, she found that 29 out of 31 textbooks covered the topic of case studies, yet from 1970 to 1979, 18 out of 30 textbooks published failed to mention case studies *at all.* Instead, these textbooks usually discussed participant-observation or other forms of "fieldwork" as alternative data collection techniques, reflecting the only coverage given to qualitative research.

In evaluation, this trend was serendipitously reinforced during the same period of time by the classic work of Campbell and Stanley (1963) in describing their variety of "quasi-experimental" designs. Their work—used for many years and by nearly every scholar as the defining text for evaluation research—unfortunately disparaged case studies as a "pre-experimental" form of research (the infamous "one-

shot case study"). Thus, Platt notes that, even when mentioned in textbooks, the orthodoxy of the times frequently treated case studies not as a distinct method but as an optional part of the exploratory work that might occur during the early stages of the complete research process. Coverage of the case study method by evaluation textbooks continues to be spotty to this day. As but one example of a complete and consistent oversight, none of Rossi and Freeman's first five editions of their popular evaluation textbook—1979, 1982, 1985, 1989, and 1993—contains the term *case study,* much less a discussion of it as a method.

Platt credits the first edition (1984) of my book on case study research (Yin, 1994a) as having raised fresh consciousness over the method. (Important publications leading up to the first edition were Yin, 198la and 1981b.) The fact that the significant features of the case study—as profiled in the previous section of this paper—focused on design rather than data collection distinguished the method from participant-observation. Further, according to Platt, the couching of the method within a practical format readily encouraged more people to do case studies. Whether as a result of my work or not, during the late 1980s there appears to have been much more attention given to the case study method, mainly for research but also in evaluation, especially federally-supported evaluations of education programs. The 1990s have now produced whole texts, again, about the case study (such as Feagin, Orum & Sjoberg, 1991; Ragin & Becker, 1992; Stake, 1995). In addition, professions such as public administration (Agranoff & Radin, 1991), business administration and management information systems (Benbasat, Goldstein, & Mead, 1987; Cash & Lawrence, 1989), and social work (Rubin & Babbie, 1993; Gilgun, 1994) have reincorporated case study research into their repertory of research methods, not just using case studies as a teaching tool.

In evaluation, the 1990s also saw a major case study handbook published by no less an authority than the (now defunct) Program Evaluation and Methodology Division of the U.S. General Accounting Office (1990; an earlier version was issued in 1987). The book covers the major topics in applying the case study method to evaluations—design, site selection, data collection, data analysis, and reporting—providing guidance, illustrative examples, and warnings about common pitfalls. The book notes that, at the time, "the history of the case study as an evaluation method is little older than a decade" (p. 10). Key concepts underlying the method continued to be "triangulation" and the "rich, in-depth nature of the information sought."

Again, parallel developments in evaluation methods more generally also occurred and in retrospect may have helped to produce this revived attention. In particular, Campbell now the author and coauthor of two works bearing directly on a revised view of case studies in evaluation. The first was a rather little-known article, "Degrees of Freedom and the Case Study" (Campbell, 1975). In this article, Campbell questioned whether he and others had fully appreciated the power of the case study method in the past. He noted that if a case indeed followed the same numeric mindset of other quasi-experimental designs, every case study should have a "plethora" of explanations—because of the numerousness of variables and the paucity of data points. Instead, he showed how classic case studies arrived at satis-

factory explanations only after a long and agonizing analytic process. Something else, besides the mere tallying of variables, was at work.

The second was a widely used textbook, which was a follow-up to the classic Campbell and Stanley (1963) work, and of which Thomas D. Cook was now the first author (Cook & Campbell, 1979).The book disentangled the case study method from the earlier categorization of the case study as a quasi-experimental design with the following unequivocal statement: "Certainly the case study as normally practiced should not be demeaned by identification with the one-group post-test-only design" (the infamous one-shot case study) (p. 96). In addition, the book even contained, as one of the variant quasi-experimental designs, a design that was in fact applicable to case studies (although the book did not refer to such applicability): the non-equivalent, dependent variables design (p. 118). According to this design, an experiment or quasi-experiment may have multiple dependent variables—that is, a variety, of outcomes. If, for each outcome, the initially predicted values have been found, and at the same time alternative patterns of predicted values (including those deriving from methodological artifacts or threats to validity) have not been found, strong causal inferences can be made. Because of this applicability to case studies (which normally have multiple dependent variables), this design then became the basis for using pattern matching as an analytic technique in doing case study research (Yin, 1994a, pp. 106–110). (Campbell's contribution both to case studies and to evaluation is the topic of a much more extensive article: Yin & Bickman, forthcoming).

Whether related to my book (as credited by Platt), to changes in evaluation research more generally (as just discussed), or to yet other reasons, the gains and renewed foothold made by the case study method in appearing as part of the routine range of research and evaluation methods represent a major advance. However, although the new texts and references began to appear in the early 1990s, they were the culmination of forces that began in the late 1970s and the 1980s. As such, they are only the beginning of the story of whether there has been a decade of progress in case studies as an evaluation tool from 1987 to 1997. The rest of the story is told next.

PROGRESS (OR LACK OF IT) DURING THE PAST DECADE

If the legacy of the immediately preceding period was the increased documentation of the case study method, the 1987–1997 decade itself has produced increased use and diversification of case study tools and thus the elaboration of the method. Somewhat equivocal is the effect this diversification has had on actual case study products.

Use and Diversification Of Case Study Tools

The elaboration of the case study method begins with a more refined understanding of the uses of case studies in evaluation. The GAO volume (U.S. Government Accounting Office, 1990, p. 9) explicitly lays out at least six different situations: illustrative, exploratory, critical instance, program implementation, program effects, and

cumulative (meta-analysis of multiple-case studies done at different times). Each situation demands slightly different designs. For instance, the design for an illustrative case study may be limited to the point being illustrated, whereas the design for a program implementation or program effects case study requires extensive expression of presumed causal links. At the time of its publication, the GAO volume noted that existing reports tended to use only two of the six applications (illustrative and critical instance); that trend may have since broadened.

A second area of use and diversification has been in preparing for and documenting case study evidence. The use of *case study protocols* to organize the data collection—protocols that are far broader in scope than a simple questionnaire—is now commonly accepted as the most desired prelude to systematic data collection. The need for a case study protocol is especially great and has become frequent practice where multiple investigators are collaborating in doing multiple case studies, but are all still part of the same overall evaluation. Similarly, the understanding that case study evidence may be contained in a separate *case study database*—different from the actual final case study report—has taken greater hold. The database may take both narrative and tabular form, a key feature being that the noted information contains explicit footnotes or references to the specific source of the evidence (thereby helping to preserve the desired *chain of evidence*). Further, the database, though not edited or intended for public presentation, nevertheless needs to be available for independent inspection by other investigators.

However, during the decade, possibly the most important advance in tools has been the use of *logic models* as part of the design in doing case study evaluations (Yin, 1992; Yin, 1993, pp. 65–68). A logic model presents the presumed causal sequence of events expressed in a series of cause-and-effect steps. Developed initially to carry out evaluability assessments (Wholey, 1979), the specification of logic models is a rewarding activity in at least two respects. First, the logic model reveals the underlying theory of a program that is being evaluated, and the specification of the model provides the guidance for the relevant data that need to be sought during the case study. Second, the process of putting a logic model together—especially when shared between program managers and evaluators working together—often yields insights that need not await the completion of all evaluation but that are immediately useful for program development.

At the same time, the proper and complete specification of logic models is still an evolving craft. Potentially worrisome is that the most common logic models still only identify different effects or stages but do not give an actual explanation of how events move from one stage to another. For instance, as shown in the upper part of Figure 5.1, the typical logic model consists of a series of boxes (stages) connected by a series of arrows (causal relations among the stages). The accompanying logical statements take the following form: "By implementing this activity (input), the program will engage the needed number of participants (output) and will eventually have the desired effect on these participants (outcome)." Left unstated is exactly how the activity will indeed engage the participants or how the effect will arise

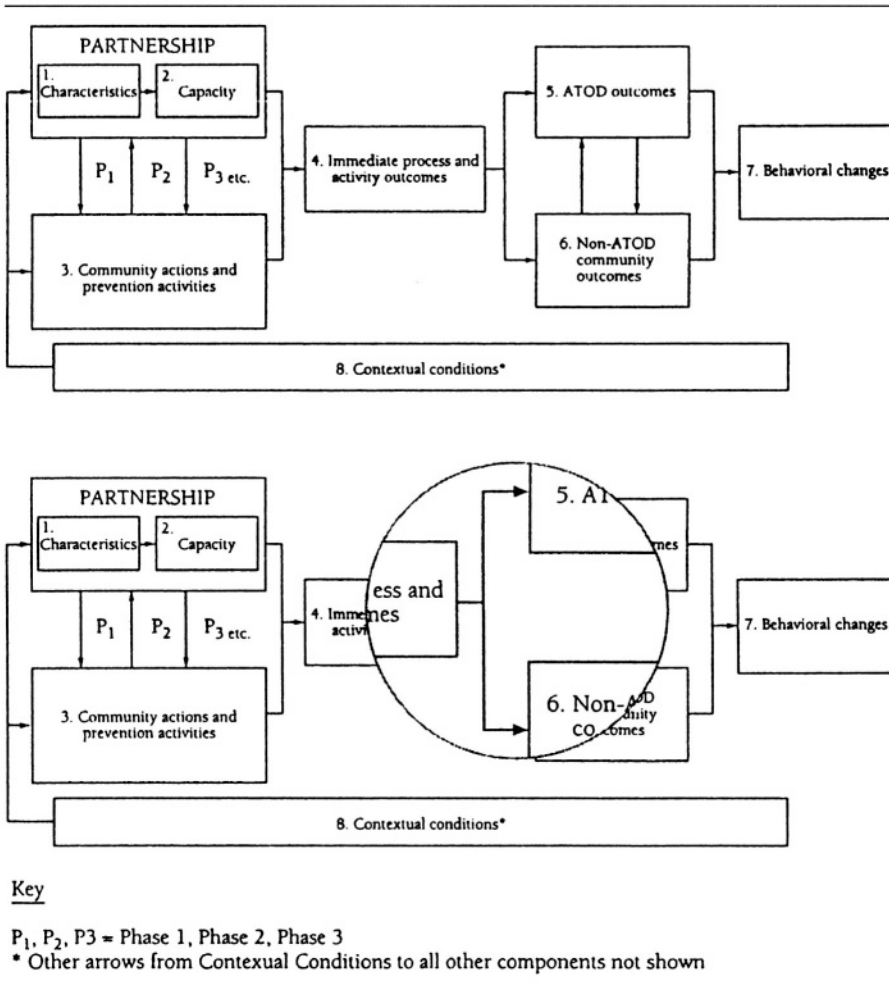

Figure 1. Relating Partnership Actions to Prevention Outcomes

from the act of participating. One possible shortcoming in these specifications has therefore been that too much attention has been given to the boxes in a logic model and not enough to the connecting arrows. The lower part of Figure 1 therefore deliberately focuses attention on one set of arrows, and the challenge to the case study investigator is to associate substantive how and why explanations with the arrows.

Case Study Practices During The Past Decade

Examinations of case study evaluations—and discussions with investigators attempting to do case studies—reveals greater use of these various case study tools over the decade. More important, the craft is now explicitly recognized as having tools and rigor, a norm going considerably beyond the earlier and crude notion that doing case studies mainly required an ability "to tell it like it is." With the increased availability of texts referring to and describing the case study method, investigators also have more ways of finding the needed guidance to practice the craft. Overall, the major progress in case study evaluations during the past decade may very well be the fact that investigators are knowingly pursuing practices that are part of a formal craft.

The process of doing case study evaluations also has become a more collaborative activity. Case study evaluators now work more closely with the officials of the program being evaluated to conduct the initial phases of evaluation, including the definition of the questions to be addressed by the evaluation, the evaluation design, and the preferred data collection methods. Similar trends, covered under such rubrics as "action research," "cluster evaluation," and "empowerment evaluation," are increasingly true of all evaluation methods, not just case studies. However, case study evaluations—focusing on concrete and readily understandable issues—lend themselves best to this new participatory type of evaluation.

Case Studies During The Past Decade

Whether all this documentation, awareness, and changes in practice have led to new and better kinds of case studies, however, is yet to be seen. In this sense, the final judgement on a decade of progress remains open. The routine case studies, frequently part of a multi-method evaluation design not limited to the case study method, appear to be better constructed and documented. For instance, 15 communities were the subject of ongoing research on as contemporary a topic as managed care (Ginsburg & Fasciano, 1996). Along the same lines, graduate students, both in this country and abroad, appear to be practicing better and more rigorous case studies, especially as part of their theses or dissertations.

But the decade has not produced any particularly distinctive case studies, such as Graham Allison's *Essence of Decision* (1971), much less any landmark multiple case study evaluation. There has also not emerged any streamlined way of sharing case studies, which still require a burdensome amount of text (and hence space) that probably precludes the creation of any journal devoted to case studies (in turn limiting the amount of professional communication about case studies). Even when a study takes over the whole issue of a 200-page journal ("Tracking Health System Change," 1996), as in the case of the managed care study previously cited (Ginsburg & Fasciano, 1996), the case studies are not presented as part of the publication. In these situations, a frequent problem is that the original case studies are usually too long, but there is great difficulty in preparing a second set of abbreviated texts.

At the same time, also possibly true is that the experience with case study evaluations bears great similarity to the experience with evaluation more generally (Yin, 1994b). For instance, evaluation as a whole may not have produced landmark studies during the past decade, in part because of the difficulty of improving on the exemplars of the past. Thus, interpreting the past decade of progress with case study evaluations shares the inevitable problem of interpreting many evaluation findings: no milestone stands out to any absolute extent; but somehow, conditions appear to be slightly better than before, based on process considerations—in this case, improved case study practices. Overall, and again as with the problem of interpreting many evaluation findings, possibly, the passage of more time is needed to provide a more revealing, if not definitive, perspective.