

# Some Benefits of Integrating Qualitative and Quantitative Methods in Program Evaluation, with Illustrations

Doren L. Madey

Suqua School of Business, Duke University

The integration of quantitative and qualitative methods within a single evaluation has synergistic effects in the three major phases of design, data collection, and analysis. A recently completed study, NTS Research Corporation's evaluation of the State Capacity Building Program, illustrates the benefits obtained. After reviewing the historical debate between advocates of the two methods, the program and its evaluation are described briefly. The article deals first with the contributions of qualitative methods to quantitative work; then, conversely, it examines the contributions of quantitative to qualitative methods. Some of the costs involved in such a methodological marriage are explored. In a concluding section, the applicability of integrated methodologies to other evaluations is considered.

## The Historical Debate: Qualitative vs. Quantitative

Prior to the 1970s, tightly controlled experimental designs dominated the evaluation field. Classical Fisherian experiments,<sup>1</sup> preferably using factorial designs, were extolled as the most powerful and acceptable model for evaluation (Boring, 1954; Fisher, 1935 et. seq.). Where the stringent requirements of the tightly controlled experimental design could not be met, quasi-experimental designs were considered viable alternatives (Campbell & Stanley, 1966). Program evaluators recognized

---

<sup>1</sup> The classical experimental design has also been designated the "agricultural-botany" design. This terminology is not accidental. Many of the statistical and experimental techniques used in educational evaluation were originally developed (e.g., by Fisher, 1935) for use in agricultural experiments. Parlett and Hamilton (1975) concisely describe the most common form of agricultural-botany type evaluation: It "involves the assessment of the effectiveness of an innovation by examining whether or not it has reached required standards on prespecified criteria. Students—rather like plant crops—are given pretests (the seedlings are weighed or measured) and then submitted to different experiences (treatment conditions). Subsequently, after a period of time, their attainment (growth or yield) is measured to indicate the relative efficiency of the methods (fertilizers) used. Studies of this kind are designed to yield data of one particular type, i.e., "objective" numerical data that permit statistical analyses. Isolated variables like IQ, social class, test scores, personality profiles, and attitude ratings are codified and processed to indicate the efficiency of new curricula, media, or methods" (p. 142).

---

This article is a revision of the author's paper, "The Evaluator's Perspective: Evaluating the State Capacity Building Program," which was presented at the 1978 Annual Meeting of the American Educational Research Association, Toronto, Canada, March 30, 1978. The NTS evaluation of the State Capacity Building Program is supported with federal funds from the National Institute of Education, Department of Education; however, the contents of this article do not necessarily reflect the views or policies of the National Institute of Education, or the Department of Education.

I wish to thank Jolinda K. Decad, John Egermeier, Eugene C. Royster, John F. Steege, and A. Jackson Stenner for their helpful suggestions in preparing this manuscript, and Pauline Matson for her careful typing of the manuscript.

that to follow the experimental model faithfully would be too difficult to design and to implement (Evans, 1969). But qualitative designs were shunned as too imprecise to produce either information of value to decisionmakers or information to establish program accountability:

There is almost universal agreement among evaluation researchers that the randomized controlled experiment is the ideal model for evaluating the effectiveness of public policy. If there is a Bible for evaluation, the Scriptures have been written by Campbell and Stanley. (Rossi & Wright, 1977, p. 13)

Following the release of preliminary results of a Westinghouse Learning Corporation-Ohio University evaluation, which indicated that "the long term effect of Head Start appears to be extremely weak,"<sup>2</sup> a major controversy raged on how best to evaluate federal programs. The Head Start evaluation had employed a quasi-experimental (*ex post facto*) design, which, until the results were released, had been heralded as yielding the best evaluation information one could hope to obtain (Rossi, 1972). With repetitive announcements of similar "weak" results by other evaluations, decisionmakers began to realize that knowing whether a program worked or did not work would not be enough. For program improvement purposes, decisionmakers would also need to know why a program did or did not work (Mason, 1973; Pincus, 1980).

Dissatisfaction with the methodologies in vogue resulted in a search for alternatives. Studies employing qualitative methodologies emerged as a solution (Murphy, 1974; Parlett & Hamilton, 1975; Stake, 1978) and the studies' designers began to defend their approach:

Some people assume that natural inquiry is unstructured, undisciplined, and unsystematic; let it suffice to say that such thoughts are unfair, inaccurate, and sim-

plistic. Natural inquiry is no less rigorous than traditional experimental research, it is simply different. The paradigm of natural inquiry is comprehensive in scope, demanding in design, and requires a set of honorable skills that even some rigid experimentalists lack (although they might be embarrassed to admit it). (Wolf & Tymitz, n.d., p. 1)

In the 1970s, qualitative evaluation designs for federal programs began to achieve the respect once reserved only for classical experimental designs, but the controversy on how best to evaluate federal programs raged on.

Until recently, two camps were firmly entrenched in a methodological battle.<sup>3</sup> One camp extolled the virtues of hard, generalizable data (e.g., Campbell & Stanley, 1966; Page, 1978) and the other extolled the superiority of deep, rich observational data (e.g., Eisner, 1979; Stake, 1978). In principle, evaluators agreed on the appropriateness of combining methods, but in practice, evaluators seldom consciously integrated methods; instead they were likely to employ the methods of one of the two camps. Reasons for concentrating on one methodology included limited time, money, and skills. The antagonism between the two camps stymied the utilization of both quantitative and qualitative methods in one study, with the result that,

<sup>3</sup> Interestingly, a similar polemic existed in the field of sociological research between the proponents of qualitative fieldwork and the advocates of survey research. In contrast to the evaluation field, the early preeminent methodology was qualitative fieldwork (as exemplified by such classics as the Hawthorne studies, the Middletown volumes, and the Chicago studies of deviant groups). Following World War II, survey research became the dominant methodology and fueled a debate between the proponents of older field methods and advocates of newer survey techniques. The debate in the social research field was first mediated by the assertions of Trow (1957) and Zelditch (1962) that the nature of the problem dictates the method to be applied. In discussing the broader field of research in the social sciences, Webb et al. (1966) suggested the use of several methods or "triangulation" as an alternative to the single method approach to research. Sieber (1973), stimulated by the ideas of John D. Ferguson, was perhaps the first to document the benefits to be gained by the integration of fieldwork and survey methods in sociological research. Sieber's arguments, although overlooked by evaluators, are as applicable to the evaluation field as they are to the field of sociological research.

<sup>2</sup> President Richard M. Nixon, Economic Opportunity Message to the Congress, February 19, 1969, as quoted in Williams and Evans, "The Politics of Evaluation: The Case for Head Start," *The Annals of the American Academy of Political and Social Science*, September 1969, 385, 118.

until recently, actual examples of "integrated evaluations" were scarce.<sup>4</sup>

By the late 1970s, the debate began to shift beyond qualitative versus quantitative to dealing with the issues involved in integrating methods (e.g., Cook & Reichardt, 1979). In the 1980s, the goal continues to be the design and implementation of evaluations that will provide adequate information for policymakers' needs (see Pincus, 1980), but the means to achieve this goal are changing. Alternative approaches to evaluation that synthesize methodologies from a variety of disciplines have been described and analyzed recently (see Smith, 1980). Additional documented examples of studies that use both qualitative and quantitative methods are still needed. By examining a recently completed evaluation, this article attempts to illustrate how methods may be integrated during design, data collection, and analysis.

Just as a standard definition of evaluation does not exist (cf. Ross & Cronbach, 1976; Stenner, 1974), the terms "quantitative" and "qualitative" have not been uniformly demarcated. Here evaluation is defined as a "process of clarifying a set of information needs, and collecting, analyzing, and reporting information to alleviate those needs" (Stenner, 1974, p. 4). Qualitative methods are defined as relatively open-ended methods employed during design, data collection, and analysis. They include "the more episodic, subjective pro-

cedures, common to the case study" (Stake, 1978, p. 6) and other "soft" methods of "naturalistic inquiry" (Guba, 1978). Qualitative methods range from observation and other unobtrusive forms of interaction in natural settings (Webb, Campbell, Schwartz, & Sechrest, 1966) to open-ended probes, ethnographic techniques, case studies, and structured interviews. Quantitative methods are defined as relatively closed-ended methods. Included are survey instruments, closed-ended questionnaires, indices and scales, achievement and aptitude tests, and randomized, controlled experiments (Campbell & Stanley, 1966; Cooley & Lohnes, 1976).

### The State Capacity Building Program and Its Evaluation

Prior to discussing the benefits and costs associated with an evaluation that attempts to integrate methodologies, it is important to describe the specific program and evaluation that will be used to illustrate the integration of methods.

Through the National Institute of Education (NIE)-sponsored State Capacity Building Program<sup>5</sup> (established in 1975 and still operating), state education agencies (SEAs) are awarded 1-year, renewable grants of about \$100,000 each to support the development and eventual institutionalization of statewide dissemination systems for making current educational knowledge and practices accessible to administrators and practitioners. According to the NIE program announcements, such systems are to be comprised of three generic components: (1) an information resource base containing the knowledge or knowledge-based products clients need, (2) linkages to connect the resources with the people who could benefit from them, and (3) a component to coordinate the various activities needed so local educators could use the system for school improvement.

<sup>4</sup> The polarization has also hidden the fact that a continuum of methods might actually exist, and that each camp actually employs, although not necessarily documents, methods along the continuum. Another way to reconceptualize what might be a false dichotomy is to expand Guba's (1978) wave conceptualization of naturalistic inquiry. According to this conceptualization, the naturalistic investigator cycles through a series of observations that are directed first at discovery and then at verification. Through the wave of discovery and verification, the investigator expands and reduces the domain of inquiry using a mixture of naturalistic strategies. The wave conceptualization can be expanded to include qualitative and quantitative methods and reshaped into a spiral (or series of decreasing waves) to represent closure or convergence. (Guba does discuss convergence in his monograph, but does not incorporate this concept in his wave theory.) In the spiral conceptualization, qualitative and quantitative methods interact synergistically throughout the course of the investigation or until closure is reached.

<sup>5</sup> For a more extensive description of the program and its evaluation, readers are referred to the following documents: *Improving State Dissemination Efforts: An Evaluation of NIE's State Dissemination Grants Program, Final Report* (Madey, Royster, Decad, & Baker, 1980) and *The State Dissemination Grants Program: 1979 State Abstracts* (Madey, 1980).

From these generic components, states are expected to develop specific systems, customized to their own contexts, which extend or adapt existing structures for enhancing dissemination services; and as a collaborative effort between NIE and the states, timelines are established by mutual agreement.

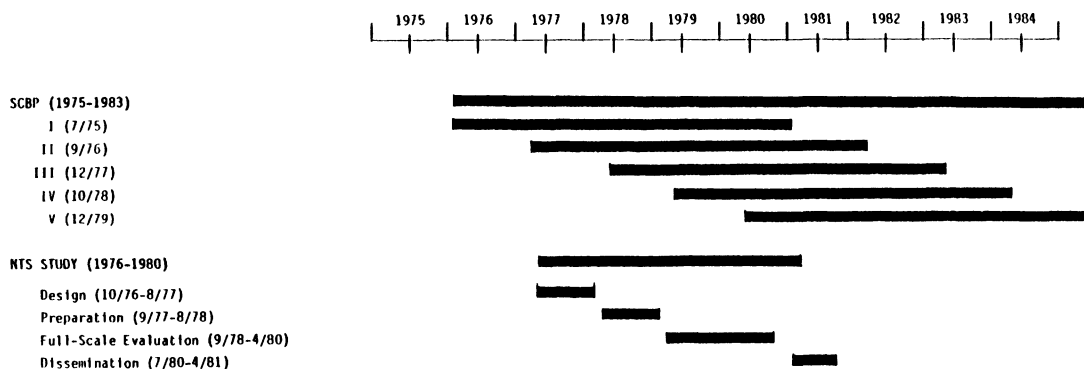
Under the sponsorship of the Research and Educational Practice unit of NIE's Program on Dissemination and Improvement of Practice, NTS Research Corporation conducted a study of the first 4 years of operation of the State Capacity Building Program—1975–1979 (Madey, 1980; Madey, Royster, Decad, & Baker, 1980; Royster, Madey, & Decad, 1980). The purpose of this study was not to evaluate the success of specific capacity building projects, but to identify factors that facilitate or impede SEA efforts to build and institutionalize statewide dissemination systems. The NTS study was intended to develop an understanding of how federal and state policy might promote capacity building for this program and for future capacity building programs.

Given a program so dynamic, NTS felt that restricting its evaluation to strictly quantitative methods was apt to result in limited findings, however “hard” they would have been. A similarly unilateral commitment to qualitative techniques was apt to produce numerous instances that would have been inapplicable to general

policy decisions. An eclectic, integrated approach, however, was eminently suited to an evaluation of this kind, in which program goals were innovative, program guidelines were nonprescriptive to accommodate wide variation among the participants, project development was expected to be evolutionary and incremental, and in which study audiences—and their respective information needs—could differ dramatically. (That is, NIE program staff members, state project staff, legislators, and concerned personnel of state, regional, and local education agencies each made decisions at different levels and required different information in different formats at different times; see Hayman, Rayder, Stenner, & Madey, 1979.)

Quantitative and qualitative evaluation methods, therefore, complemented each other importantly throughout the study, from design through data collection to analysis. The evaluation was comprised of four phases: (1) a design phase (October 1976 - August 1977) devoted to describing the program (Madey, Mojkowski, & Strang, 1977; Strang, 1977), clarifying and translating the program's goals into measurable variables, and developing a design, appropriate instrumentation, and data collection and analysis procedures for the study; (2) a preparation period (September 1977 - August 1978), which included initial fieldwork in 23 project sites, some descriptive reporting (Madey, Haenn, & Strang, 1979)

FIGURE 1. *Timelines for the State Capacity Building Program (SCBP) and the NTS Study*





refinements in the study design (Madey, Haenn, Everett, & Strang, 1978) and approval of a forms clearance package (Haenn, Madey, Everett, & Strang, 1978); (3) the full-scale evaluation (September 1978 - April 1980), which included two waves of quantitative data collection (Fall 1978, in 33 project states; and Fall 1979, in all 50 states) and an additional wave of qualitative data collection (Winter 1980, in five project states); and (4) a dissemination phase (July 1980 - April 1981) in which the study's findings and implications (Decad & Madey, 1981; Madey, 1980; Madey et al., 1980; Royster et al., 1980; Royster & Madey, 1981) were shared with policymakers, researchers, and practitioners. Figure 1 summarizes the four phases of the NTS study, highlights major data collection periods, and indicates how the NTS study fits into the overall time period of the NIE program.

### Benefits Accrued in Integrating Methods

An evaluation of the State Capacity Building Program illustrates the benefits obtained by the interplay of methods.<sup>6</sup> Both the contributions of qualitative to quantitative methods and the reverse sequence are discussed. Although the methods are not always interchangeable, each method can be strengthened by using intrinsic qualities of the other.<sup>7</sup> The benefits to design, data collection, and analysis for both sequences are examined.

<sup>6</sup> I am especially indebted to Sieber (1973) for the ideas presented in this article regarding the interplay of methods and for the examples from other studies.

<sup>7</sup> Sieber (1973) is not entirely correct in his interpretation (p. 1340) of Webb et al.'s (1966) argument on behalf of multitechniques. Webb does not argue that the methods are entirely interchangeable as Sieber states. Webb is careful to add that using multiple techniques could "supplement" as well as "cross-validate" a specific technique (p. 2). While Sieber's interpretation of Webb's argument is inaccurate, his own thesis is complete and fairly concise: "There are areas of informational overlap between methods, but there are also large areas of information which can be gained only by a particular technique. If each technique has an inherent weakness it also has an inherent strength unmatched by other techniques. The opinions held by a large population can be measured only by survey techniques; the un verbalized normative pattern of a small group might be measurable only by observation" (p. 1337).

### Contributions of Qualitative Methods to Quantitative Methods

*Design.* Qualitative methods can enrich quantitative designs by improving both the sampling framework and the focus of the overall design. Exploratory interviews and observations can serve either to confirm or to rationalize a sampling framework (Sieber, 1973). Other qualitative tools can be used to pinpoint and establish the priority of information needs.

1. *The sampling framework.* In the NTS evaluation, qualitative methods shaped the sampling framework at both the system and individual levels. Initial fieldwork revealed that the study design should include all states in the program as this would be the only way to adequately capture and describe the variations among states. Initial fieldwork also revealed that the inclusion of states without capacity building projects would provide additional information (in the form of comparison groups) about the effects of the NIE program. Resource constraints, however, prohibited this fieldwork from dictating the original sampling framework for the NTS evaluation. In the final phase of the evaluation, a limited amount of quantitative data was collected from nonprogram states, with the design of the instrument evolving from prior experiences in the field. Individuals were also included or excluded from the sampling framework using information obtained from qualitative work. For example, during the initial site visits it became apparent that surveying all project-affiliated linkers would be a costly and inefficient method of obtaining information about linkage structures and processes within the state. In addition, such a survey could threaten a project's attempt to institutionalize existing linkages when the connections between the project and linkers were informal, as was often the case (see Madey, Royster, Decad, & Baker, 1980). Thus, in the final design, only a sample of linkage agents was interviewed.

2. *The evaluation design.* Qualitative methods also were used to focus the overall evaluation design. The information-based evaluation (IBE) model was used to define and specify information requirements of the evaluation, and policy implications analysis (PIA) was designed to en-

liven the forum through which information users expressed their information needs.

Recognizing that failure to involve the information users in the design phase drastically lowers the probability that the right questions will be posed and answered, NTS developed and used the IBE model (Stenner, 1974). The overriding consideration of this evaluation model is to supply information to individuals in decisionmaking roles. The model's two reference points are the information users and the information needs as expressed in the evaluation questions. Using this model with NIE and the states, priorities were established in information categories (i.e., resource base, linkages, leadership, federal role) and information user categories (i.e., NIE program staff, state project directors). The evaluation resources were then allocated to meet these identified priorities.

To extend and formalize the IBE approach to delineating information domains and information users, NTS developed PIA (Madey & Stenner, 1980; Stenner & Madey, 1977), a program evaluation tool that creates an active forum through which information users can express their information needs. PIA accounts for the policy context surrounding an evaluation and makes more explicit the information requirements of key information users. PIA elicits reactions to hypothesized evaluation findings to determine: (1) whether each finding fits within the expected scope of the evaluation; (2) whether each finding corresponds with expectations; (3) the extent of each finding's policy relevance; and (4) the extent to which each finding is communicated appropriately. PIA also asks for policy actions that might result from the findings, what level or type of information would be needed to justify the findings, and whether additional information would be needed to justify the policy actions. PIA explicates information requirements of policymakers and program participants. In the NTS study, the PIA method was used with NIE staff, state project personnel, and advisory council members to better understand the policy context within which the evaluation was commissioned, and to identify the types of information that would be needed to enhance the decisionmaking process (Madey

& Stenner, 1980). Results were used to guide the study team in developing the final evaluation design. For example, the preliminary study design focused primarily on management alterations to improve projects in the *current* program. PIA results indicated that to be useful to policymakers the evaluation would also need to address such issues as how program designers and implementors could help facilitate "institutionalization," or continuation, of project functions after the NIE grant ceased and what policymakers could learn to apply to *other* dissemination efforts. NTS attempted to incorporate these information needs into the final evaluation design. In addition, findings and recommendations related to institutionalization and future dissemination programs were included in the final evaluation report (Madey et al., 1980).

**Data Collection.** Useful information about the receptivity, frames of reference, and attention spans of respondents can be obtained from observations and exploratory interviews preceding the development of quantitative survey instruments (Sieber, 1973). When systematic pretesting is not undertaken because of time or financial constraints, it is still possible to obtain valuable information by using qualitative methods prior to the full-scale data collection phase. Qualitative pretesting can serve two purposes: (1) it may result in a better instrument package, an objective of all design work; and (2) the rapport stemming from qualitative pretesting may smooth the way for more elaborate, time-consuming, and sometimes threatening aspects of full-scale data collection.

In the NTS evaluation, informal pretesting of the Client Assessment Package (CAP), a set of five machine-readable instruments to assess the process of seeking and using information and assistance for school improvement (Madey & Everett, 1978), served both purposes. Qualitative pretesting during the design phase resulted in significant improvements in the original instrument package. Certain items (e.g., client's race) were eliminated and others were refined (e.g., follow-up assistance) in order to reduce respondent burden and eliminate potentially offensive items; in addition, several previously excluded topics (e.g., satisfaction with personal services

provided) were identified as important aspects of client assessment and the instrument package was modified accordingly. The informal pretesting also created a rapport between the individual project staffs and the evaluation team, which facilitated later implementation of the instrument package.

**Data Analysis.** Examples from the NTS evaluation of the State Capacity Building Program illustrate the six ways (Sieber, 1973) that qualitative methods can aid in the analysis of quantitative data. First, the conceptual framework that guides the analysis can be derived, in part or in whole, from qualitative work. Second, in areas where methods yield information overlap, certain quantitative results can be verified by results obtained using qualitative methods. Third, qualitative data can be used as the basis for selecting survey items to be used in index construction. Fourth, external validation of empirically generated constructs can be obtained by comparison with observation data. Fifth, case studies can be used to illustrate statistically derived models. Sixth, clarification of ambiguous and provocative replies to individual questionnaires can be obtained by reexamining field notes.

1. *Conceptual framework development.* To conduct a thorough and comprehensive evaluation, there must be an organizing framework that builds on the information needs, relevant theory, and knowledge of the program. A conceptual framework is a necessary step in evaluation design (Scriven, 1974; Stenner, 1974; Stufflebeam, 1974). Its function is to act as a data organizer to guide instrument development and data analysis.

Six information sources, some of which were qualitative, provided heuristic bases for framework development. First, a considerable amount of descriptive information was collected during the design phase, particularly from the earliest funded states. Second, additional state visits clarified staff perceptions concerning the perceived differences and similarities among projects. Third, completion of an expanded literature review characterized the state of the art in dissemination, knowledge production and utilization, organizations, and implementation theory. Fourth, NTS conducted three workshops during

which state personnel reviewed the study design, instrumentation, and data collection plans. Fifth, experts in the areas of dissemination and implementation were used as consultants and contributed to the development of the organizing framework. Finally, NIE received two draft reports, and suggestions for changes were incorporated in the final conceptual framework. Each of these sources contributed to the conceptual framework (Madey et al., 1978) that subsequently guided instrument development and data analysis.

2. *Verification.* It is sometimes possible to validate quantitative findings by referring to information obtained using qualitative probes and interviews. In the NTS study, survey findings were verified by reference to subsequent fieldwork. Tabulation of the project director questionnaires revealed that some projects reported difficulties in coordinating their dissemination activities with program-specific education programs (e.g., vocational education). It was felt that the continued fragmentation of the dissemination components of federal programs might explain this survey finding, but because a few projects had developed dissemination systems that included a comprehensive set of programs, the evaluation team felt uneasy about resting their case on the survey finding alone. During later interviewing, several directors were asked about their relationships with the dissemination components of specific federal education programs. They believed that the continued fragmentation at the federal level impeded the development of comprehensive dissemination systems within states. Because this viewpoint was expressed with great conviction even by those informants who had made progress in coordinating dissemination efforts at the state level, the field interviews lent plausibility to the survey finding.

Qualitative verification of a quantitative survey instrument usually focuses on selected items or variables as in the above examples. It would be possible, albeit tedious and costly, to extend such verification to an entire instrument.

3. *Index construction.* Indices for use in analysis of quantitative data can be constructed from qualitative observations, probes, and interviews. In the NTS evalu-

ation, items for a quantitative instrument to measure "capacity building" were derived from discussions with project directors during a series of workshops. For example, the variables used to define the construct "institutionalization" included "provision for state funding of project activities after the grant ends," "contribution of other funds to a coordinated dissemination function," "agencywide budgeting for dissemination," and "state legislation dealing specifically with dissemination."

4. *External validation.* A statistical index can be validated by reference to qualitative work. In the NTS study, quantitative measures obtained using the "capacity building" instrument were enriched by using qualitative methods. For example, the measures of "institutionalization" was verified by using informal interviews with several project directors. During the interviews, individuals were told where their project had been ranked on the construct and then asked if the score coincided with their perceptions. Where discrepancies were discovered, additional probing was used to determine whether the mismatch was due to a weakness in the instrument or misinterpretation by the individuals interviewed.

5. *Case study illustration.* The use of case studies to illustrate models or statistical and historical types is a common practice in the social research field (Gordon, 1957; Kahl, 1953; Komarovsky, 1962; Sieber & Lazarsfeld, 1966). Recent evaluations have used this approach also (Pelavin, Hayward, Orland, & Pelavin, 1980). In the NTS evaluation of the State Capacity Building Program, case studies were used to illustrate several approaches to capacity building; the approaches, or models, were derived using quantitative methods.

6. *Clarification.* A final contribution of qualitative work to the analysis of quantitative data involves the clarification of puzzling but provocative replies to written questionnaires. In the NTS study, both telephone and on-site interviews with state project directors were devoted to clarifying quantitative data puzzles. In the survey of both project and nonproject states, one respondent left blank the series of questions dealing with linkers and several other respondents provided curious, possibly significant, but far from clear an-

swers. Subsequent telephone follow-up with all such respondents revealed the difficulties inherent in existing definitions of linkers, a theme which was reinforced during the site visits and which became an important finding in the final report. Similarly, one capacity building project director wrote an ambiguous reply to a closed-ended question regarding the institutionalization of the project's functions with ongoing SEA dissemination efforts. In the course of fieldwork, this director was asked to clarify her answer. She explained that activities once associated with the project lost their project-designation when the project's functions were integrated with ongoing agency efforts, but this was part of the plan to institutionalize the project's functions without creating a turf battle in the agency. Thus, tracing the impact of the NIE grant became difficult as the project evolved. Her clarification illuminated the problem of adequately examining the effects of such projects on agency dissemination activities.

### *Contributions of Quantitative Methods to Qualitative Methods*

Quantitative techniques can contribute to qualitative methods in many ways. The contributions of such techniques as surveys and questionnaires to interviews and case studies are perhaps less appreciated than the reverse. In some instances it would seem to be "methodologically obligatory" for qualitative methods to take advantage of quantitative techniques (Sieber, 1973). As in the previous section, illustrations are derived, whenever possible, from the NTS evaluation.

*Design.* Quantitative methods can improve the design of a primarily qualitative study by identifying both representative and unrepresentative cases at either the system or the individual level. Studying representative cases serves the goal of generalizability; studying unrepresentative cases facilitates refinement of models and theories.

In the NTS study, quantitative methods were used to identify projects that represented different approaches to capacity building. Then indepth case studies were conducted in five states to enrich the quantitative findings. By integrating the findings from both approaches, the final eval-



uation report attempted to describe how SEAs develop and institutionalize state dissemination systems and to pinpoint factors which help or hinder such efforts.

**Data Collection.** At least three contributions of quantitative work to the collection of qualitative data are possible (Sieber, 1973):

1. *Leads.* When quantitative work has been conducted prior to qualitative work, the replies to survey instruments can provide leads for later interviews and observations. The need to ask routine background questions during an interview is eliminated when such questions have been included in a previously administered survey. In the NTS evaluation, guidelines and probes for the interviews conducted during the case studies were derived from items obtained using the survey instruments. For example, some of the probes regarding institutionalization of project functions were derived from the project director's questionnaire.

2. *Overlooked respondents.* Quantitative data can also provide information about informants or subjects who were overlooked initially. Hypothetically, if the data collected using the Client Assessment Package consistently revealed that the most important person in facilitating use of an information service was the local librarian, and if the NTS study had not examined the role of the local librarian at the time of such a discovery, then such a study might be recommended to understand further how individuals use information and assistance.

3. *Correction of the elite bias.* One pitfall in qualitative data collection is an "elite bias" in the selection of informants and in the evaluation of statements. If an evaluation gravitates to the elite of a social system in the course of the qualitative work, the results might be biased. Studying only the elite is a distinct possibility in an evaluation for four reasons. First, the elite or "gatekeepers" are often an evaluator's initial contact with a group. Because first impressions are often lasting, the views of the elite may color the evaluator's objectivity through the ensuing study. Second, if the elite individuals are respected in the larger social system, the evaluator may neglect association with other contacts. Third, the elite individuals may be more articulate than other members of the

group, and thus appear to be more knowledgeable. Fourth, an evaluator who spends more time collecting information from the members of the elite may give more weight to their viewpoints than to those of other group members.

The utility of integrating methods to correct the elite bias is illustrated by using a study that initially set out to examine the structure of two suburban school systems but developed into a study of school boards, superintendents, and leaders of high school teachers:

Prior to looking at the results of the survey, I predicted the proportion of teachers who would respond in particular ways to survey questions. I then compared my predictions with the actual responses. It became obvious when observing these comparisons that I had unwittingly adopted the elite's version of reality. For example, I overestimated the extent to which teachers felt that the administration accepted criticism... Similarly, I had assumed that the teachers were more satisfied with evaluative procedures than was in fact the case... I also overestimated the rank-and-file support for the leaders of the teachers association, with whom I had spent a good deal of time. In short, I had fallen prey to the elite bias, despite recent training in the dangers of giving greater weight to prestigious figures as informants. The survey not only constrained me to see that my qualitative data collection procedures had been faulty, but also provided the opportunity to learn about an entire stratum which I was now aware of having glossed over in the fieldwork, namely, the elementary teachers. Apparently, the elite bias had operated also in my preference for secondary teachers, who are the more esteemed both in the profession and the community. (Sieber, 1973, p. 1353)

**Data Analysis.** Quantitative methods can contribute to the understanding of qualitative observations in at least three ways: (1) by correcting the holistic fallacy, (2) by verifying qualitative interpretations, and (3) by casting a new light on field observations (Sieber, 1973). Of all the synergistic effects gained by the interplay of methods, these contributions are perhaps most overlooked.

1. *Correction of the holistic fallacy.* The "holistic fallacy" is the "tendency on the part of field observer to perceive all as-

pects of a . . . situation as congruent" (Sieber, 1973, p. 1354). This fallacy was a possibility in the evaluation of the State Capacity Building Program. For example, an articulate state project director who felt that he/she was doing an outstanding job of developing a computerized resource base might have given the impression that the other required components (e.g., linkage, leadership/management) were also developing well, when in fact quantitative methods (e.g., the Client Assessment Package) might have revealed that few clients actually had access to the information resource base and even fewer were capable of using the information in the form provided. Although this is only hypothetical, the sensitivity of the NTS evaluators was heightened to the holistic fallacy in their site visits.

2. *Verification of qualitative interpretations.* Verifying observations collected during informal visits is possible using quantitative instruments. In the NTS study, hypotheses crystallized during the design visits were explored more systematically during full-scale data collection. For example, it appeared that those states with the most numerous specialized dissemination programs had more fragmented dissemination systems; the NTS quantitative results confirmed this. Similarly, it seemed as if strong project leadership would facilitate development and institutionalization of a coordinated SEA dissemination system; again, quantitative results confirmed the hypothesis.

3. *Casting new light on qualitative observations.* Quantitative findings can sometimes clarify an inexplicable or misinterpreted field observation. Serendipity is not limited to qualitative work. The analysis of quantitative data often yields unanticipated observations and concepts. It even has been argued that surveys are uniquely qualified for the measurement of unanticipated concepts (Schenkel & Sieber, 1969). Although an example from the NTS study is not available, another study is used to illustrate how quantitative findings can sometimes clarify qualitative observations:

In a study of two suburban districts, it was observed that a smaller proportion of teachers in the larger district turned out to vote in the bond issue election [than in the

smaller district]. When this observation was shared with informants, many explanations were offered. The poor turnout had been tentatively attributed to the alienation of many teachers in the more bureaucratized system. . . . While perusing the distribution of responses to a survey, the investigators noticed with surprise that 39% of the teachers in the larger district resided outside of the district, compared with only 18% in the smaller district. The teachers in the larger district were simply less often legally qualified to vote. The observation of poorer turnout was therefore reinterpreted. (Sieber, 1973, p. 1357).

When unanticipated findings emerge from case studies, they should be shared with the evaluation's clients in an effort to eliminate competing hypotheses and to more accurately explain the phenomenon under investigation.

### *Summary of Benefits Accrued*

In summary, benefits are gained by employing both quantitative and qualitative methods in one study:

The methods are analogous to *zooming in* and *zooming out* with a lens. To the extent that they are reproduced objectively, wide-angle, telephoto, and microscopic views must be *simultaneously* valid, and zooming from different directions merely focuses attention on different facets of the same phenomenon. . . . There are no grounds, logical or otherwise, for calling any view simple. We can start anywhere and zoom in to infinite detail, or zoom out to indefinite scope. (Willems & Raush, 1969, pp. 82-83)

Qualitative methods provide depth to some of the causes behind the changes (or lack thereof) observed using more quantitative methods. Qualitative methods sharpen the focus on and provide additional explanations of the observed relationships. Quantitative methods yield instruments that can be used repetitively with little additional expenditure of resources. Qualitative methods provide an enduring yardstick for measuring change. Table I summarizes in outline form both the general contributions of qualitative methods to quantitative work and the reverse.

In the NTS evaluation, qualitative meth-

ods used during the design phase contributed to the formulation of a conceptual framework and the development of more quantitative instruments. During the data collection phase, concurrent use of both methodologies augmented the information yield. Because the evaluation team was in the field for a portion of the data collection, it was possible to learn about the actual meaning of the questionnaires to respondents. To some extent, it was possible to assess the obtrusiveness of the quantitative instruments and take such qualitative judgments into account during interpretation of the data. In the data analysis phase, findings obtained using the quantitative instruments were clarified and extended using, once again, qualitative methods. By interweaving qualitative and quantitative methods over the duration of the NTS evaluation, regardless of the primary method during any phase, the utility

of the information provided to decision-makers at the federal and state levels increased. Integrating methods in a single evaluation yielded a richer tapestry of findings and a deeper understanding of the program. Decisionmakers seemed more likely to understand, value, and use the evaluation results because they were actively involved in the evaluation.

### **Cost-benefit Considerations of Integration**

Perhaps no evaluation is without potential disadvantages, and one which integrates methods is no exception. While integrating methods does alleviate the criticisms of a single method approach (cf. Mason, 1973; Webb et al., 1966), the evaluator encounters a new set of problems when quantitative and qualitative methods are integrated within a single evalua-

TABLE I

*Summary of Benefits Obtained by Integrating Methods in Design, Data Collection, and Analysis*

| <i>Contributions of Qualitative Methods to Quantitative Methods</i> |   |
|---|---|
| <i>Design</i>   | Qualitative methods can enrich quantitative designs by improving: <ol style="list-style-type: none"> <li>(1) the sampling framework</li> <li>(2) the overall study design</li> </ol>  |
| <i>Data Collection</i>  | Qualitative methods can enrich quantitative data collection by improving: <ol style="list-style-type: none"> <li>(1) the instrument package</li> <li>(2) instrumentation administration</li> </ol>  |
| <i>Analysis</i>   | Qualitative methods can enrich quantitative analyses by providing: <ol style="list-style-type: none"> <li>(1) a conceptual framework to guide the analysis</li> <li>(2) verification of quantitative findings</li> <li>(3) index construction of quantitative analyses</li> <li>(4) external validation</li> <li>(5) case study illustration</li> <li>(6) clarification of quantitative findings</li> </ol> |
| <i>Contributions of Quantitative Methods to Qualitative Methods</i> |   |
| <i>Design</i>   | Quantitative methods can enrich qualitative designs by identifying: <ol style="list-style-type: none"> <li>(1) representative cases, to serve the goal of generalizability</li> <li>(2) unrepresentative cases, to refine models and theories</li> </ol>  |
| <i>Data Collection</i>  | Quantitative methods can enrich qualitative data collection by providing: <ol style="list-style-type: none"> <li>(1) leads for later interviews and observations</li> <li>(2) information about overlooked subjects</li> <li>(3) correction of the "elite" bias</li> </ol>  |
| <i>Analysis</i>   | Quantitative methods can contribute to the understanding of qualitative analysis by: <ol style="list-style-type: none"> <li>(1) correcting the holistic fallacy</li> <li>(2) verifying qualitative interpretations</li> <li>(3) casting new light on field observations</li> </ol>  |

tion. Formal arguments for quantitative and qualitative methods notwithstanding, the root of the conflict is probably that many studies simply cannot afford to implement both modalities and expect, at the same time, to do either particularly well. The more techniques employed, the more time, people, and money will be needed; consequently, the decision to integrate methods must be based on pragmatic assessments of the projected data payout.

The design phase of the NTS evaluation, for example, exceeded the initial timeline by 7 months, a lag between planning and implementation that caused legitimate concern among prospective information users despite the measurably greater richness and applicability of the data that would be collected. Another inconvenience was that for a long time program improvement efforts were derived from information other than study findings. Further, where quantitative techniques tend to be rather aloof from actual program operations, qualitative techniques tend to require a great deal of personal interaction, which can be burdensome for all concerned however genial those interactions might be. Sharing information and collaborating actively with a wider-than-usual spectrum of program participants will unavoidably require more time, personnel, travel, and materials. Extended timelines means more people on salary or people on salary longer, or both. Procedural necessities, such as clearing an unusually comprehensive instrument package through the Office of Management and Budget, may require additional time. Integrated research methods also require evaluators who are or could quickly become unusually adept in both modes. In short, time and resource expenditures will increase across the board.

Of course, resourcefulness, meticulous planning, scheduling, and creative task assignment can mitigate some of these problems. Individual surveys can be economized if informal networks can be identified among respondents, permitting a systematic approach. Among the interviewers, some can be instructed to extend their survey observations into unstructured interviews, or to make certain kinds of observations in the course of their other duties. On balance, however, it appears that

an integrated study can be mounted most economically when the object is compact (e.g., a school system), and where the respondents are clustered in settings that have definite boundaries; but attempting an integrated evaluation of a broad-based, multifaceted, and variably structured program would be self-defeating without adequate time, funds, and personnel.

In response to these rather stringent conditions, however, experience indicates that an integrated approach to program evaluation is the modality of choice where the study seeks to determine not only whether or not a complex program works, but also why and how well it works, and how it (or similar future programs) might best be improved. This inference does not follow automatically from the enhanced information yield; but where the information users are systematically and individually diverse, and the kinds of information required from the study vary widely as a consequence, an integrated approach appears to be the appropriate way to provide the study's several audiences with the policy-relevant information they require. Decisionmakers are more likely to understand and use evaluation findings when they have been involved personally in the study's design and implementation; there is greater assurance that the questions needed for decisionmaking purposes will be asked and answered; and inferences, interpretations, and conclusions are supported with richer and more comprehensive sources than could be expected from a study that used quantitative or qualitative methods alone.

Lastly, an integrated study offers the legitimate possibility of substantial long-term savings because once the study questions are answered, and program policy issues are precipitated by them, much of the data needed should be accessible immediately from an existing and relatively comprehensive information bank. Program officers may find it unnecessary, therefore, to undertake extensive follow-up studies because the support for alternative policy options may already be available; or second-level studies can be mounted economically in the form of reanalyses, or as precisely targeted substudies that can be plugged into analysis frames derived from existing data.



## Conclusions and Recommendations

The adjustments in traditional evaluation designs necessitated by the integration of quantitative and qualitative methods appear to produce another evaluation style. In terms of methods, one plus one equals three. And what's the "three?" The interplay between the two types of methods; the interaction, the synergistic coming together, which creates something that never existed before. As exemplified by the NTS evaluation of the State Capacity Building Program, the integration of quantitative and qualitative methods within a single evaluation can enhance the three major study phases of design, data collection, and analysis. To benefit from both methodologies, an evaluation must be able to incur increased costs in time, personnel, and resources. When these costs are not exorbitant, integrating methods will increase and enhance the information yield and, by so doing, facilitate the process of clarifying and alleviating information needs.

Presently, documented examples of this "integrated evaluation" are too limited to deduce all the general principles to be followed in organizing future evaluations. As more studies undertake an integrated approach to evaluation, those principles will emerge. In the meantime, evaluators should be aware of all available evaluation methods and choose their methods in terms of their value for producing the kinds of information needed in any given study, rather than in terms of some preconceived biases about what constitutes legitimate evaluation methods. In making these choices, evaluators should examine the full spectrum of those available. When existing methods are inappropriate, attempts to develop, implement, and test new methods should be undertaken. When selecting methods and designing a study, evaluators should be receptive to the enormous potential that integrating methods offers for improving studies.

## References

- BORING, E. G. The nature and history of experimental control. *American Journal of Psychology*, 1954, 67, 573-589.
- CAMPBELL, D. T., & STANLEY, J. C. *Experimental and quasi-experimental designs for research*. Chicago, Ill.: Rand McNally, 1966. Originally published in Gage, N. L., Ed., *Handbook on Research on Teaching*. Chicago, Ill.: Rand McNally, 1963.
- COOLEY, W. W., & LOHNES, P. R. *Evaluation research in education: Theory, principles, and practice*. New York: Irvington Publishers, 1976.
- COOK, T. D., & REICHARDT, C. T. (Eds.). *Qualitative and quantitative methods in evaluation research*. Beverly Hills: Sage Publications, 1979.
- DECAD, J. K., & MADEY, D. L. Analysis of linker activities, roles, and types: Findings from the NTS study. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.
- EISNER, E. W. The use of qualitative forms of evaluation for improving educational practice. *Educational Evaluation and Policy Analysis*, 1979, 1(6), 11-20.
- EVANS, J. Evaluating social action programs. *Social Science Quarterly*, 1969, 50(3), 568-581.
- FISHER, R. A. *The design of experiments*. Edinburgh: Oliver and Boyd, 1935 et. seq.
- GORDON, C. W. *The social system of the high school*. Glencoe, Ill.: Free Press, 1957.
- GUBA, E. G. *Toward a methodology of naturalistic inquiry in educational evaluation*. Bloomington: Indiana University, 1978.
- HAENN, J. F., MADEY, D. L., EVERETT, M., & STRANG, E. W. *Request for OMB clearance with supporting documents for the evaluation of the state capacity building program in dissemination*. Durham, N.C.: NTS Research Corporation, April 1978.
- HAYMAN, J., RAYDER, N., STENNER, A. J., & MADEY, D. L. On aggregation, generalization, and utility in educational evaluation. *Educational Evaluation and Policy Analysis*, 1979, 1(4), 31-39.
- KAHL, J. A. Educational and occupational aspirations of "common man" boys. *Harvard Educational Review*, 1953, 23, 186-203.
- KOMAROVSKY, M. *Blue-collar marriage*. New York: Random House, 1962.
- MADEY, D. L. *The State Dissemination Grants Program: 1979 State Abstracts*. Durham, N.C.: NTS Research Corporation, 1980.
- MADEY, D. L., & EVERETT, M. *The Client Assessment Package*. Durham, N.C.: NTS Research Corporation, 1978.
- MADEY, D. L., HAENN, J. F., & STRANG, E. W. 1978 *State Abstracts with an analysis across the states*. Durham, N.C.: NTS Research Corporation, 1979.
- MADEY, D. L., HAENN, J. F., EVERETT, M., & STRANG, E. W. *A conceptual framework for the evaluation of the State Capacity Building Program*. Durham, N.C.: NTS Research Corporation, 1978.
- MADEY, D. L., MOJKOWSKI, C., & STRANG, E. W. *State Reports*. Durham, N.C.: NTS Research Corporation, July 1977.
- MADEY, D. L., ROYSTER, E. C., DECAD, J. K., & BAKER, R. F. *Building capacity for the improvement of educational practice: An evaluation of NIE's State Dissemination Grants Program (Final Report)*. Durham, N.C.: NTS Research Corporation, 1980.
- MADEY, D. L., & STENNER, A. J. Policy implications analysis: A method for improving policy research and evaluation. In C. B. Aslanian (Ed.), *Improving educational evaluation methods: Impact on policy* (Sage Research Progress Series in

- Evaluation, Vol. 11). Beverly Hills, Calif.: Sage Publications, 1981.
- MASON, W. S. *Problems of measurement and the NIE program*. Washington, D.C.: National Institute of Education, 1973.
- MURPHY, J. T. *State education agencies and discretionary funds: Grease the squeaky wheel*. Lexington, Mass.: D. C. Heath and Company, 1974.
- PAGE, E. B. More objective. Debate, Should educational evaluation be more objective or more subjective? with Robert E. Stake, delivered at the annual meeting of the American Educational Research Association, Toronto, March 1978.
- PARLETT, M., & HAMILTON, D. Evaluation as illumination: A new approach to the study of innovative programs (1975). In G. V. Glass (Ed.), *Evaluation Studies Review Annual*, 1976, 1, 140-157.
- PELAVIN, S., HAYWARD, B. J., ORLAND, M., & PELAVIN, D. *An evaluation of the Fund for the Improvement of Postsecondary Education*. Durham, N.C.: NTS Research Corporation, 1980.
- PINCUS, JOHN (ED.). *Educational evaluation in the public policy setting*. Santa Monica, Calif.: Rand Corporation, 1980.
- ROSS, L., & CRONBACH, L. J. (EDS.). *Handbook of evaluation research. Educational Researcher* (essay review by a Task Force of the Stanford Evaluation Consortium), 1976, 5(10), 9-19.
- ROSSI, P. H. Testing success and failure in social action. In P. H. Rossi & W. Williams, *Evaluating social programs: Theory, practice, and politics*. New York: Seminar Press, 1972.
- ROSSI, P. H., & WRIGHT, S. R. Evaluation research: An assessment of theory, practice, and politics. *Evaluation quarterly*, 1977, 1(1), 5-52.
- ROYSTER, E. C., MADEY, D. L., & DECAD, J. K. *Building capacity for improvement of educational practice: An evaluation of NIE's State Dissemination Grants Program* (Executive Summary). Durham, N.C.: NTS Research Corporation, 1980.
- ROYSTER, E. C., & MADEY, D. L. *Relating research to policy issues in the NTS Study of State Capacity Building and Dissemination*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.
- SCHENKEL, W., & SIEBER, S. D. *The design of survey questionnaires: A case history approach*. New York: Bureau of Applied Social Research, Columbia University, 1969. (mimeo)
- SCRIVEN, M. Evaluation perspectives and procedures. In W. J. Popham (Ed.), *Evaluation in Education*. Washington, D.C.: American Educational Research Association, 1974. (Monograph)
- SIEBER, S. D. The integration of fieldwork and survey methods. *American Journal of Sociology*, 1973, 28(6), 1335-1359.
- SIEBER, S. D., & LAZARSFELD, P. F. *The organization of educational research* (Cooperative Research Project No. 1974, USOE). New York: Bureau of Applied Social Research, Columbia University, 1966.
- SMITH, N. L. Federal research on evaluation methods in health, criminal justice and the military. *Educational Evaluation and Policy Analysis*, 1980, 2(4), 53-59.
- STAKE, R. E. The case study method in social inquiry. *Educational Researcher*, 1978, 7(2), 5-7.
- STENNER, A. J., & MADEY, D. L. *Policy implications analysis* (Exercise No. 1, Design for the Evaluation of the State Capacity Building Program in Dissemination). Durham, N.C.: NTS Research Corporation, 1977.
- STENNER, A. J. *An overview of Information Based Evaluation: A design procedure* (Information Based Evaluation Series, Book 1). Durham, N.C.: IBEX, 1974.
- STRANG, E. W. *Building capacity in dissemination: Literature review*. Durham, N.C.: NTS Research Corporation, 1977.
- STUFFLEBEAM, D. L. Alternative approaches to educational evaluation: A self-study guide for educators. In W. J. Popham, (Ed.), *Evaluation in Education*. Washington, D. C.: American Educational Research Association, 1974. (Monograph)
- TROW, M. Comment on participant observation and interviewing: A comparison. *Human Organization*, 1957, 16(Fall), 33-35.
- WEBB, E. J., CAMPBELL, D. T., SCHWARTZ, R. D., & SECHREST, L. *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally, 1966.
- WILLEMS, E. P., & RAUSH, H. L. *Naturalistic viewpoints in psychological research*. New York: Holt, Rinehart & Winston, 1969.
- WILLIAMS, W., & EVANS, J. W. The politics of evaluation: The case for Head Start. *Annals of the American Academy of Political and Social Science*. September 1969, 385, 11-132.
- WOLF, R. L., & TYMITZ, B. *Toward more natural inquiry in education*. Bloomington: Indiana University, n.d. (mimeo)
- ZELDITCH, M., JR. Some methodological programs of field studies. *American Journal of Sociology*, 1962, 67, 566-576.

---

## AUTHOR

DOREN L. MADEY, Director, Corporate Relations, Suqua School of Business, Duke University, Durham, NC 27706. Specializations: Research, evaluation, and public policy.