DOCUMENT RESUME

ED 214 978 TM 820 213

AUTHOR Daillak, Richard H.; Alkin, Marvin C.
TITLE Qualitative Studies in Context: Reflections on the CSE Studies of Evaluation Use.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
REPORT NO CSE-159
PUB DATE 81
GRANT NIE-G-80-0112
NOTE 140p.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Case Studies; *Field Studies; *Interviews; *Research Methodology; Research Problems; School Districts
IDENTIFIERS Center for the Study of Evaluation CA; *Evaluation Utilization; *Qualitative Studies

ABSTRACT

Qualitative research methodologies employed by the Center for the Study of Evaluation (CSE) Evaluation Use Project over the last six years are reviewed. The report is neither a paean to qualitative methods nor an attack upon them. It represents a self-examination by CSE of its attempts to apply qualitative techniques to an important, and reasonably complex, educational research problem: the study of evaluation information use in local schools. All of these qualitative methodologies have proven useful, although they have not been without difficulties and limitations. In the report, the research methods are described and critiqued. The study methodologies include case studies, evaluation field study, and user interview survey. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Center for the Study of Evaluation

UCLA Graduate School of Education
Los Angeles, California 90024

ED214978

QUALITATIVE STUDIES IN CONTEXT
REFLECTIONS ON THE CSE STUDIES OF EVALUATION USE

Richard H. Daillak
Marvin C. Alkin

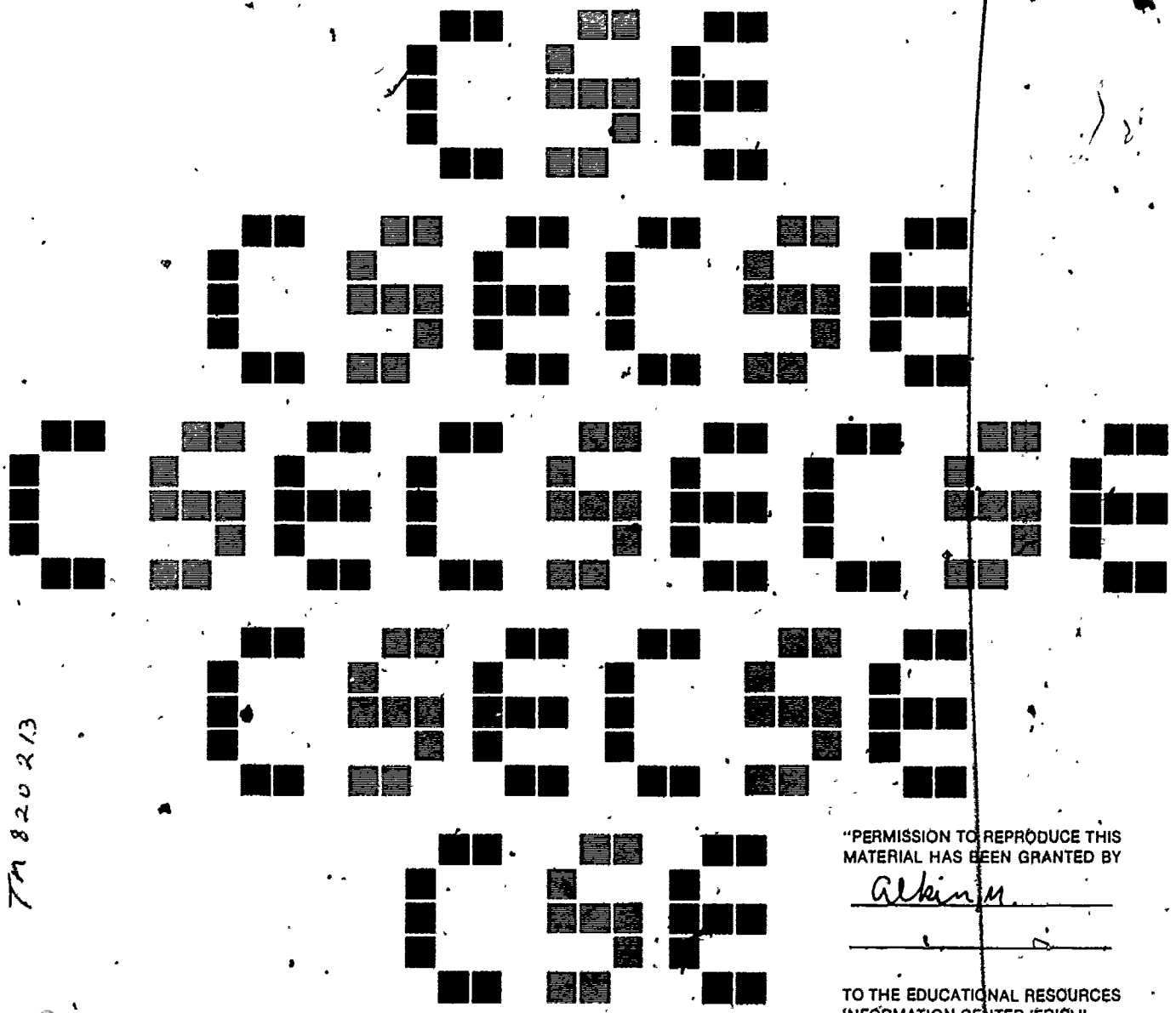
CSE Report No. 159
1981

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy



TM 820 213

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Alkin M.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

QUALITATIVE STUDIES IN CONTEXT:

REFLECTIONS ON THE CSE STUDIES OF EVALUATION USE

Richard H. Daillak
Marvin C. Alkin

CSE Report No. 159
1981

EVALUATION USE PROJECT
(Grant Number NIE-G-80-0112 P-4)
Marvin C. Alkin, Project Director
CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California - Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, U.S. Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

Chapter 1

Introduction

Purpose of the Report

In this report, we review the qualitative research methodologies employed by CSE's Evaluation Use Project over the last six years. The report is neither a paean to qualitative methods nor an attack upon them. It represents a self-examination of our best efforts to apply qualitative techniques to an important, and reasonably complex, educational research problem: the study of evaluation information use in local schools. On the whole, all of these qualitative methodologies have proven useful, although they have not been without difficulties and limitations. In the report, the research methods are described and critiqued.

There are dual reasons for such a review. First, the time seems right for a methodological reassessment. Enthusiasm for qualitative methods burgeoned in the late 1970's and has ebbed a bit, since then. Some researchers have taken qualitative methods to task based upon discouraging experiences (e.g., Miles, 1979). The CSE experience has been less discouraging, although not without its lessons. It seems appropriate, and potentially helpful to others, to share these experiences.

Second, the CSE Evaluation Use Project has had the

opportunity to apply several qualitative research strategies. The Project's use of interview-based case studies, participant observation, and semi-structured interview surveys constitutes an uncommonly broad experiential base, which others may find useful to examine.

The Experience Base

Since 1975, the CSE Evaluation Use Project has completed three major studies of local schools' use of evaluation information. The first study -- actually a set of similarly conducted qualitative case studies -- examined evaluation in five ESEA Title I (compensatory education) or Title IV-C (innovative) programs. Repeated, open-ended interviews were conducted with the evaluators and with administrators and staff of the five programs, and an evaluation history of each program was developed. In addition, the interviewees were asked to explain the uses made of the evaluation findings. A narrative report on each program (i.e. "case") was critiqued by the interviewees and other reviewers. After all five case studies were completed, a cross-case analysis was prepared (Alkin et al., 1979).

Thereafter, two further studies were conducted, this time within a single urban California school district (Metro Unified School District). One, the Evaluator Field Study (Daillak, 1980), used participant observation and interviews to trace the work of three on-staff evaluators

within Metro. Over the 1979-80 school year, the evaluators' work was observed, and institutional influences upon their work were investigated.

The other study, the User Interview Survey (Alkin et al., 1980), administered a loosely-structured one-hour interview to sixty-five school evaluation "consumers" (principals and their subordinate administrative staff) within twenty-two schools. The interviewees were questioned about recent program planning or decisionmaking activities, and about their attitudes towards evaluation. Written summaries of the (tape-recorded) interviews were prepared by the interviewer and by an independent tape-listener. These summaries then were used to develop several analytical papers.

The study methodologies are reviewed in three chapters to follow. A brief epilogue summarizes our experience with the methods.

The Problem: Evaluation Underutilization

There is substantial evidence that program evaluation findings are seldom used by local school decisionmakers, such as district staff, program directors, or building principals (David, 1978; Kennedy et al., 1980), despite the fact that evaluation activity is required of many federal or state categorical aid programs. Local school districts generally comply with the letter of the evaluation requirements imposed upon them. They conduct the student

achievement testing required by funders, and they annually assess each program's success in meeting its prespecified objectives. But schools do not seem to attend to or use the evaluation information they generate, at least not significantly.

Research has attempted to document the precise extent of local evaluation use, to identify explanations for use or disuse, and to determine ways to increase evaluation use. Some researchers have tried to improve evaluation research methods in order to make evaluation data objectively more accurate and, they hope, more useful. Others have examined the interaction between evaluation and client systems, believing that evaluation use is decided more by the decisionmakers' attitudes and intentions than by objective properties of the information itself.

Our research has emphasized the latter approach, focusing primarily on social, political, and organizational explanations for evaluation utilization. To assess complex socio-organizational interactions, we have relied heavily on qualitative studies of evaluation settings rather than employing more reductionistic approaches, such as quantitative or simulation-based research.

Chapter 2

Evaluation Case Studies

RESEARCH DESCRIPTION

In 1975, when prior research on evaluation use was scant, it seemed important to examine evaluations, in an exploratory way, in order to discover the factors that might account for evaluation information use (or disuse) by local school decision makers. If a few plausibly important factors were to emerge from these initial studies, then they might be investigated in later research. For example, if evaluation report formating appeared to affect information use, then later research might examine the merits of alternative procedures for reporting an evaluation's findings to decision makers.

Choosing the Case Study Method

It seemed logical to favor either survey or field research methods for the initial, exploratory studies. Neither field experiments nor laboratory simulations appeared appropriate. Field experiments seemed premature. Paper-and-pencil simulations could have been instituted more quickly, at lower cost. Program circumstances, evaluation findings, and reports all could be simulated. By introducing systematic variations into the simulated evaluations, one could explore the effects of proposed new evaluation methods. However, we hesitated to try building

a simulation for a phenomenon about which so little was known.

Surveys and exploratory field studies remained as plausible research strategies -- and both had their advantages. Questionnaire and interview surveys could reach larger numbers of respondents, providing a more representative view of evaluation. Exploratory field studies -- case studies, in-depth interviewing, participant observation, and related methods -- could provide more detailed research data, however.

For our purposes when the initial research was planned, case studies based on detailed interviews seemed to be the most desirable approach. The argument for exploratory field studies has been made often before:

> Sociologists usually use ...[field research] when they are especially interested in understanding a particular organization or substantive problem rather than demonstrating relations between abstractly defined variables. They attempt to make their research theoretically meaningful, but they assume that they do not know enough about the organization a priori to identify relevant problems and hypotheses and that they must discover these in the course of the research. (Becker, 1958, 652-653)

Investigating evaluation situations through naturalistic field research thus had much to offer, but the problem of deciding just what shape the field research effort should take remained to be decided. Should there be participant observation, or would interviews be the better approach? How should the sites be found, how long would we

spend with them, and what, if anything, should we look for? These questions were resolved as described below.

Case Study Procedures

As is typical in qualitative field research, exact procedural details varied from case to case. Nevertheless, the major procedural steps and their sequence were quite similar across the cases.

Overview

The case studies were primarily interview-based and retrospective (i.e., focusing on a previously completed evaluation). Several in-depth interviews were conducted with the operational staff and the evaluator of the program selected for study. To supplement the interviews, documentary evidence, such as program proposals, evaluation reports, and so forth, were reviewed. The research sought to discover the way that an evaluation, and the findings it produced, had fit into the program's total operations. A primary concern was to discover whether an evaluation had influenced program activities and why it had the influence (or lack of influence) it did.

In several cases, one or more year-long evaluations had been completed and a follow-on evaluation was in progress. In these cases, the interviewer constructed an account of the completed evaluations in the usual fashion (described below) and also updated this account at intervals to reflect contemporaneous developments.

The final product of each case study was a narrative case report summarizing the study findings. Program context was described; one or more evaluations were recounted; evaluation influence, if any, was described; and the case was briefly analyzed, focusing on plausible explanations for the observed degree of evaluation influence.

After all five case studies were completed and reported (a matter of two to three years' work, since the studies were begun at staggered intervals), an integrative analysis was prepared. This reanalysis isolated a number of common factors -- that is, ones recurring across several cases -- which seemed to encourage, constrain, or in some other way affect evaluation information use in the cases studied. Similar factors were clustered until a final framework of factors influencing evaluation utilization was produced. The case study narratives and the framework were published together as the concluding product in the initial research phase (see Alkin et al., 1979).

Procedures for the Individual Case Studies

Site Selection. Only local school programs receiving ESEA Title I (compensatory education) or Title IV-C (innovative program) funds were selected. The choice of local ESEA Title I and IV-C programs reduced the potential program diversity somewhat, yet still provided an abundant program pool from which to choose sites. In addition, ESEA evaluations (required as a condition of

funding) account for much evaluation activity in local educational agencies.

Because only a small number of studies were to be conducted (five in all, selected one at a time over two years), elaborate efforts at randomized sampling were not undertaken. Instead sites were identified either through contacts in the public schools or from among the respondents to an earlier mailed survey which the Evaluation Use Project had conducted in California school districts. The major criteria for site selection were: (a) the existence of a Title I or IV-C program in at least its second year of operation, (b) geographic proximity, (c) willingness of site personnel to participate, and (d) a subjective assessment of the Program's suitability for the study.

A program which met the first criterion would necessarily have gone through at least one full annual evaluation cycle; thus, the interviews could retrospectively trace the impact of the completed evaluation(s) and also examine any ongoing evaluation activities. The second criterion, geographic proximity, meant simply that the program had to be within reasonable traveling distance of our Los Angeles base -- effectively, within the Southern California area. The host's willingness to participate voluntarily was a third, crucial criterion. We had no power to force ourselves upon any

school site. Potential hosts had to be persuaded to cooperate with the research. As a consequence, some amount of self-selection was inevitable in the case study.

The final criterion, a subjective judgment of "suitability," was complex. As the cases were selected, an effort was made to diversify the sample within the broad parameters set by the other criteria. We tried to select programs of varying types: two were Title I, three were Title IV-C; three were in large, urban districts, two in smaller, suburban districts; two cases involved internal evaluators who worked full time within evaluation units of their school districts; in two other instances the internal evaluators were district employees who were not part of an organized evaluation unit; and in the last case, the evaluator was an outside consultant. Besides striving for diversity, we sought to avoid programs which might be misleadingly atypical -- for example, a district's showcase effort, treated markedly differently from its other federally funded programs.

Initial Contacts with Site Personnel. First discussions typically were with the selected program's evaluator. (In only one case was the program director contacted first). Typically, the first contact was a phone conversation in which the proposed study's purposes and methods were outlined. In each case, the person contacted agreed to further discussions, usually through one or more

personal meetings but also occasionally involving an exchange of letters. These initial discussions quickly drew in the program director, the evaluator, and perhaps a school official (such as a principal or district administrator), as well as CSE project members (typically, the Project Director and the interviewer).

The initial personal meetings introduced the research to program administrators. We tried to project a sincere, nonjudgmental interest in their program and its evaluation. Anonymity was promised to the participants: in any research reports, pseudonyms would be substituted for the names of cities, districts, schools, programs, and individuals, and other identifying elements would be disguised in ways which would preserve anonymity without significantly distorting the facts.

None of the sites initially contacted declined to participate. Several factors probably contributed to this high participation rate. Selecting sites through professional acquaintances certainly played a key role, as did the reputation of the senior researcher and the sponsoring agencies.

Data Collection. None of the meetings with subjects were tape-recorded. Written notes were taken, and the interviewer expanded and elaborated these notes after leaving the interview situation. An effort was made to keep the interviews flexible in format and informal in

tone. Initial interviews were particularly wide-ranging. Later, as the interviewer became more familiar with the case details, the interviews often focused on questions prepared in advance by the interviewer.

In the first meeting, the interviewer requested copies of major program documents, including funding proposals, program progress reports, reports of previously completed evaluations, and any other documents the program staff felt would be informative. The documents provided an "official" description of the program and its evaluation, and they were used as background for subsequent interviews.

After the introductory meeting(s), a round of private interviews was conducted with the site program director, the program evaluator, the principal, and any others actively involved in program decision making. In these interviews, the informants were encouraged to describe the program and its evaluation in their own words. Usually, the interviewer simply indicated our broad interests -- the history, context, and events of the program and its evaluation, and the evaluation's influence upon the program -- and then the informant guided much of the conversation, with the interviewer asking clarifying questions.

The evaluation literature did suggest that certain factors would influence evaluation use: the background and training of the individuals involved; the evaluation's purposes and procedures; personal interactions during the

evaluation; and the manner in which evaluation findings were communicated. If these factors were not covered in the informant's spontaneous remarks, the interviewer raised them later in the interview.

No single interview could fully cover all the topics of interest. Each interview had to be limited to an hour or two in length, and, besides, the need for clarification often became apparent only later, after interviewing other informants. Therefore, key informants were usually interviewed several times. The multiple interview sessions appeared to increase the rapport between the interviewer and the informants.

Interviews with the evaluator and the program director frequently indicated potential new interviewees -- for example, a particularly influential teacher, principal, or counselor. These new informants, when interviewed, sometimes suggested contacting others. Thus the interviews "snow-balled".

Case Analysis. The multiple interviews provided a fairly detailed description of each case. The informants' differing vantage points generally complemented one another, each filling in a part of the total picture of evaluation in the program being studied. Event descriptions could be "triangulated" (Guba, 1978) by comparing different accounts of the same actions or events. Conflicting descriptions were not a common problem; the few

apparent conflicts were resolved through followup contacts with those involved.

Once the set of interview data appeared complete -- as evidenced, for example, by considerable redundancy in the data received -- the interviewer prepared a first draft case report. Several steps were then followed to refine this draft and insure the final report's accuracy.

First, the CSE Project Director reviewed the event descriptions for consistency, completeness, and plausibility. Frequently, this review raised questions that the existing data could not resolve. To provide answers, further personal interviews or telephone discussions were held with the informants, and the draft report was revised. This cycle of review, supplementary data collection, and revision was repeated until a draft satisfactory to the CSE Project was developed.

The case report draft was next circulated to the interviewees themselves for comment. A report copy was given to each key informant (evaluator, school project director, etc.), and a personal interview was conducted after allowing the informant one or two weeks to review the report.

The informants were extremely helpful reviewers. They identified factual errors and suggested alternate data interpretations. Where necessary, additional data were collected to thrash out controversial points. Then, based

upon all the information available, the case report was edited into final form. The final report conformed to each informant's perspective on some points and differed on others. All in all, it represented the Project's best effort at a balanced and accurate case presentation.

As a final effort at fairness, and to give readers a better sense of the study's accuracy, the key informants then were asked to respond to an open-ended questionnaire assessing the final case study report and general issues pertaining to evaluation utilization. Their responses were excerpted and lightly edited (with their approval) to produce narrative memos, appended to each case report. Thus, the last word was given to the major characters in each case.

Comparative Analysis of the Case Studies.

By 1978, five cases had been individually analyzed. It seemed desirable to review all the cases in order to identify common forces underlying evaluation use.

The comparative analysis began with a review of the five case reports and an attempt to reduce their detailed narrative data to more manageable capsule summaries. Project staff members independently developed a set of "utilization concept-cards" for each case. That is, a separate 3x5 card was prepared for each critical context factor, event, situation, or participant in each case. Criticality was judged subjectively, based on the

importance that the factor or event seemed to have in determining the way the evaluation was performed and the way its findings were used.

Next, the Project team met to discuss each case, comparing and merging the concept-cards they had developed. The merged cards were then arranged schematically on a bulletin board to illustrate the flow of events and influences that seemed to characterize each case. Finally, the schemas for the five cases were compared in a search for analogous elements or patterns.

Related elements did appear across the cases. For example, in most of the cases there were card entries which related to evaluator credibility. Together, the Project staff discussed, grouped, and labeled the card groupings until, ultimately, about ten broad categories emerged. The categories formed the first version of a framework for analyzing evaluation situations. The category labels captured the terms useful for explaining why the evaluations studied had been influential or had had little effect.

Shifting from the concept-cards to a new format of taped group discussions, the Project members refined the tentative framework, in the process reconfiguring some of the original categories. These intensive discussions occupied several weeks. Gradually, the analysis stabilized, first at the category level and then at the

level of component elements (termed "properties", following Glaser & Strauss, 1965). The final version of the framework is given in Appendix A).

A paramount concern was to link the framework tightly to the cases to avoid imposing any preconceptions on the data. Much of the group discussion centered on whether the data supported the categories as then stated, and whether some other formulation would conform more closely to the case findings.

CRITIQUE

Review of the Case Study Purposes

The research's initial assumption was that evaluation was not very influential. This assumption reflected prevailing opinion, as expressed in many published complaints about evaluation underutilization (e.g., Guba, 1969; Mann, 1972). With the Evaluation Case Studies, the Project hoped to explain why school program evaluations had influenced program planning and decisionmaking so minimally. In the process, it hoped to discover ways that evaluators and program managers might increase evaluation's usefulness.

Canvassing the literature, one could compile a long list of potential causes of evaluation underutilization, some organizational, some procedural, and some interpersonal (c.f. Patton et al., 1975). But there was little evidence underpinning this potpourri of causes, because, prior to 1975, evaluation influence and use had received little study.

The case studies were to explore school settings in order to discover which, if any, of the hypothesized causes of evaluation disuse appeared significant, to identify any other important causal factors, to document evaluation strategies used by practitioners, and to determine how extensively evaluation information was used in several

programs. It seemed likely that case studies, producing detailed descriptions of a few school evaluations, would further these ends economically and expeditiously. It was expected that subsequent studies -- for example, experiments employing in-basket written simulations -- would refine the case studies' conclusions. The case studies' detailed descriptive data could, in fact, be used to develop more realistic simulations. The case studies' exploratory nature, underlined by the expectations for followup work, partially relieved our apprehensions about employing qualitative methods, which were then rather uncommon in educational studies.

Research Design Decisions

Several design decisions shaped the research. One was the fundamental decision to employ case studies rather than some other research strategy. Thereafter, a decision was made to examine already completed evaluations, i.e., to conduct retrospective studies, rather than to follow the unfolding events of a current evaluation. Finally, the decision was made to collect primarily qualitative data (mainly, narrative accounts of events) through intensive, open-ended interviews with a few local evaluation participants.

Case studies were not the only research option considered. Simulation studies -- employing written situation portrayals, mock reports and memoranda, etc. --

were considered as a means of experimentally investigating the factors influencing evaluation use. Based upon advice and our own judgment, simulation studies were deferred, first temporarily and then indefinitely. Initially, not enough was known about school program evaluations to be able to develop simulations in which one could place confidence. Later, after the case studies were completed, it appeared that interactions among evaluators and clients were crucial to utilization, and, because these interactions resisted laboratory simulation, plans for such studies were set aside.

Questionnaire surveys were also considered. Indeed, a small, pilot survey was administered to members of a California evaluation society. But evaluation influence and information use seemed to elude measurement in the questionnaires, and few interesting results emerged. An initial program of case study research seemed the most desirable option. The case studies would permit conversations with evaluation participants, allowing for intensive questioning about evaluation information use and its contributing causes. Insights developed in the case studies could target later research toward the most productive issues.

Practical and theoretical considerations combined to yield a <u>retrospective</u> case study approach. Information use could be assessed fully only at the conclusion of an

evaluation, after all the evaluation findings had been reported. The literature hinted, in fact, that an evaluation's full impact might not be felt until several months after the evaluation report's release (Davis & Salasin, 1975). Consequently, a contemporaneous study of an evaluation, from its inception until the final stages of utilization, could take substantially more than a year to complete. A retrospective study, however, could begin several months after the evaluation's completion, when utilization might be fully apparent, and quickly construct a history of past events. The retrospective approach promised faster results, at lower total cost. These advantages to the retrospective strategy carried the day, even though it was clear that retrospective data would not be as complete as that available in a contemporaneous study.

Once a retrospective approach was chosen, many other details fell into place. Observations were ruled out, except as current observations might shed light on past events. Documentary evidence would be examined, but could be expected to be rather limited. Participants' recollections were the major resource from which to develop the case data. Although recollections could be tapped through questionnaires or fixed-protocol interviews, informal open-ended interviews seemed to be the most expeditious means to detailed case descriptions. Moreover,

such interviews let the interviewees use their own words and their own organizing themes, rather than forcing them to fit their descriptions to a predetermined framework. In an informal interview, interviewees could introduce issues which otherwise might have eluded attention. And interviews could fully explore subtle uses of evaluation information, such as when evaluation strengthened a decisionmaker's opinion about a program.

The research began with an assumption that evaluators and decisionmakers were key participants in evaluation, and with the expectation that their actions largely would determine an evaluation's effectiveness. In each case, therefore, the evaluator and the principal local-program decisionmaker were interviewed. One or two influential others (usually administrative subordinates to the primary decisionmaker) were also interviewed if they appeared to influence the evaluation's design or the use made of the evaluation results.

Methodological Assessment

Descriptive Detail

A descriptive report, ranging from forty to eighty double-spaced typewritten pages in length, was prepared for each case. The reports followed a common pattern. Community and school context were briefly described. The special program receiving evaluation was described next, in somewhat greater detail. Attention was given not just to

describing the program's nature; program history was also reported, including a description of the program's origins and rationale.

Next, the program's evaluation context was briefly described, including the program's evaluation requirements and the evaluator's perceptions of his or her role vis-a-vis the program. A more detailed description of the actual evaluation work, narrated chronologically but without precise dates, followed. This included reportage of meetings between the evaluator and school staff, evaluation procedures, and evaluation findings. The case reports concluded with a description of any uses made of the evaluation findings, and a synthesis of the participants' explanations for use (or lack of use) of each major evaluation finding.

The case reports were written primarily in third-person narrative form. Quotations were few, because none of the interviews were tape recorded and the interviewer had been unable to enter many verbatim quotations in the handwritten notes. Documents, primarily evaluation reports, had been examined, and portions of their contents (e.g., evaluation descriptions and test results) were integrated into the narrative case reports. However, documents were not directly reproduced in the reports. The scarcity of direct quotations detracted somewhat from the vividness of the case reports.

Additional quotations or document excerpts would have provided more visible supporting documentation.

Obscured by the third-person narrative format was the fact that there were only three or four primary interviewees for each case. Given the small number of interviewees, the case studies' detail was high, an indication of the intensity of the interviews. Missing from the studies, however, was the breadth of coverage that is possible when many persons are interviewed. For example, detailed data was not available on classroom teachers' perspectives on evaluation. Little data was collected on the school environment, apart from that which seemed directly relevant to evaluation work. District-wide attitudes and perceptions were not explored in any detail, and there was little to place the studied evaluation in its full district context.

Descriptive Accuracy

There was strong reason to believe that the case studies descriptions of evaluation events were factually accurate. In large part, this is because the case reports concentrated on events in which the interviewees had directly participated, often jointly. The study methods -- repeated interviews, probing for details, cross-checking facts with the other interviewees, and submitting the draft report to the interviewees for review -- all worked to verify the factual descriptions of evaluation events.

Descriptions of evalution use were more difficult to verify. How can one be certain that evaluation information -- or any other information -- was "considered" by a decisionmaker? It is possible that the program staff who were interviewed exaggerated evalution use in order to respond to our obvious interest in instances of use. One safeguard, however, was that the evaluators gave their own views on utilization, which stood as a partial check on the program staff's assertions.

Explanations for events were provided in abundant detail in response to our persistent probing in the interviews. Such explanations are inherently difficult to verify. The interviewer was partially able to judge an explanation's adequacy based upon its consistency with other data and upon an assessment of the interviewee's candor and perceptiveness. Also, explanations were integrated into the case report, particularly in the final section on evaluation use, and were available, therefore, for inspection by the other interviewees and by external reviewers. These reviews were a check on the explanations' plausibility, although they were not proof of their accuracy. Often, the explanations were of events which had occurred weeks or months prior to the interviews, adding another element of uncertainty.

Unquestionably, there were threats to the case studies' accuracy, as the preceding discussion indicates.

What is difficult to convey, however, is the extent to which these concerns were assuaged during the research. The interviewees were generally very cooperative. In the course of the research, a positive relationship between interviewer and subjects usually developed, and there were strong indications that the interviewees' remarks were candid and sincere. These indications, vivid to the interviewer but elusive to describe, gave us confidence that the case studies accurately described evaluation as it appeared to the interviewees. Of more concern was whether the interviewees' understandings were adequate to explain events.

Nature of the Cross-Case Analysis

The cross-case analysis categorized the individual studies' explanations for evaluation information use (or disuse). Each individual case study, in turn, had grounded its analysis of evaluation utilization in the explanations for use or disuse provided by the interviewees. The cross-case analysis generalized from the interviewees' explanations, by identifying similar explanations across the cases and labeling the generic factors or processes which seemed to be represented. And, as a result, the cross-case analysis did not introduce entirely new analytical concepts.

The categorization (or analytical framework, as it was labeled in Alkin et al., 1979) did not furnish a predictive

theory of evaluation use. That is, the categorization did not specify how the identified factors or processes would affect evaluation information use, either singly or in interaction. Instead, a factor's presence in the categorization simply indicated the factor's potential significance. For example, the interrelationship between school site and school district was listed as a factor which could affect evaluation use, but the probable effects of specific types of site-district relationships were not indicated.

The cross-case analysis's most significant contribution was that it identified, based upon empirical data rather than speculation, a limited set of influences upon evaluation use. The analysis was one step toward a more complete theory of evaluation utilization, with more thorough investigation of the identified factors being the logical next step.

Project discussions of the case data went further than the published analysis in identifying the most important influences upon utilization. Our private conclusion was that the evaluator's and the school client's commitment to making evaluation useful was vitally important. It appeared, too, that evaluation information was most likely to be used when the evaluator had established a consultative relationship with a local school decisionmaker. Relationships in which the evaluator acted

as an auditor, monitor, or detached critic appeared less successful. Through close consultations, an evaluation could be designed that would address actual decisionmaker interests, which could be quite different from the interests that might be inferred from official program plans. And in consultations throughout the evaluation, the evaluator could involve the local client in decisions affecting evaluation design, data analysis, and interpretation. The result appeared to be more meaningful evaluation information, in which the local client had developed a sense of ownership.

It appeared that much evaluation disuse was caused by the absence of this consultative evaluation approach, either because evaluators had chosen other roles or because policies or other external constraints interfered. The most influential evaluators studied had evidenced this adaptive, consultative style rather than working solely from formal program descriptions and routine evaluation procedures.

In summary, the published cross-case analysis categorized factors affecting evaluation utilization. Because these factors were drawn from the individual case analyses, the cross-case analysis had high face validity for readers who had found the individual case studies to be plausible. Our private analyses had gone a step further. We assigned primacy to a subset of the categorized factors,

namely the initiative shown by evaluators and administrators and the use of a consultative evaluation approach. These conclusions were supported by regularities in the case data: instances of information use appeared to be associated with special individual effort and close consultation between evaluator and local decisionmaker.

Generalizability of the Findings

A traditional defense of the studies' generalizability is difficult. Personal contacts frequently were used to locate potential case sites, and subjects' participation was entirely voluntary, hence vulnerable to the biasing effects of self-selection. Although there was an effort to introduce diversity into the sample by selecting different types of projects and various school district environments, greater efforts at scientific sampling could have been made. Yet, even if sites had been randomly selected, five cases would constitute a miniscule statistical sample from which to generalize findings.

These difficulties are common to case study research, which frequently relies upon small samples, with sites selected, in part, for their convenience and cooperative response (Bogdan & Taylor, 1975). The fact is that case studies' generalizability seldom rests on the same grounds as that of quantitative surveys or experiments, if for no other reason than small sample size. But case study findings can apply to other situations, as reader

response to such studies attests. To the extent that others do find the research meaningful and relevant, then there is evidence of generalizability.

The published report of the case study research (Alkin et al., 1979) has received favorable reviews for its realism and for the insights the studies provide into evaluation practice. These reviews are perhaps the best evidence that the findings are generalizable beyond the five cases studied.

Points of Concern

In hindsight, there were aspects to the case study methods which merit some discussion, either because they represent practices we might change, based upon today's knowledge, or because they had ramifications which might be overlooked.

Avoidance of Tape Recording. The interviews were not tape recorded, primarily to avoid inhibiting the interviewees. Although the methodological literature is divided on the issue of recording, recent Project experience, in the User Interview Survey, suggests that taping need not interfere with data quality. In the User Interview Survey, interviewees adjusted quickly to the tape recorder, and they appeared to speak candidly. We feel, now, that the case study interviewees would have adjusted similarly.

Nevertheless, tape recordings are not an unmixed

blessing, even if interviewees are unintimidated by the process. Tape handling, transcription, archiving, and transcript reviews quickly create a sizable logistic burden. Had the case study interviews been taped, however, much more detailed data would have been available from which to prepare the case reports. On balance, we would tape if we had the decision to make today, but we would expect a more time-consuming and expensive study.

Narrow Study Scope. In each case, three or four persons were the primary informants, i.e., the persons with whom the interviewer spoke repeatedly and at length. (Depending on the circumstances, a few others were sometimes interviewed, albeit briefly.) The evaluator was always a primary interviewee. The other primary interviewees -- typically, project directors, building principals or on-site program coordinators -- were selected for their degree of involvement in the evaluation effort or for their significance as program decisionmakers. In general, interviews did not extend upwards to administrative or technical staff at the district office, or downwards, to teaching staff. The interviews, themselves, focused on evaluation work at the program or building level, with a major topic being the evaluation events (meetings, conversations, etc.) in which the interviewees had participated.

Because the interviewees were few and the interviews

were so often focused on evaluation events, limited data was collected on the evaluation's wider context either within a school or within the host school district. It would have been helpful to learn more about the evaluator's tactics with other programs; customary evaluation practices within a district; teachers' attitudes towards and use of evaluation; the primary concerns of school or program personnel during the years studied, regardless of whether these concerns directly related to evaluation; typical decisionmaking procedures in the schools studied; district administrators' inducements (if any) to schools to use evaluation information; and district policies on evaluation and instruction.

Some of the preceding data could have come from existing interviewees. Other data could best have been collected through interviews with persons above or below the thin stratum of school personnel we contacted. At the time the studies were conducted, it was not clear that this wider data would be useful. But, were we to repeat the studies, we would expand the interviewee pool to include informants at a broader range of levels within the schools.

We now would be inclined toward studies with a broader set of informants and broader initial scope. One can gain a more thorough understanding of a school's internal dynamics and external environment by contacting many informants and by spending more observational time within

the school. A broader study, however, requires broader clearances. In arranging for the case studies, every effort was made to keep a low research profile. In retrospect, it might have been better to try to negotiate a broader charter for the research, even though more districts might have refused to participate.

The Studies' Emphasis on Explaining Use. The case studies placed greatest emphasis on explaining the the conditions underlying instances of successful evaluation use. A major hope, after all, was to develop recommendations leading to greater evaluation use. In addition, it seemed intuitively plausible that evaluation disuse was simply the "flipside" to evaluation use, that is, that evaluation disuse would be explained adequately by the absence of those factors which led to use.

In retrospect, it now seems more appropriate to conceive of instances of evaluation use as foreground elements set against a background of predominant disuse. The case study analyses satisfactorily explained how the instances of use came to pass -- through personal initiative and consultative evaluation, primarily -- but they provided less satisfactory explanations for the widespread disuse of evaluation. Why should personal initiative or exceptional evaluation activities be required? Why should their absence lead to disuse? The case studies did not move to these more fundamental

questions, in part because we did not then recognize the need to seek explanations for both use and disuse. Today we believe that the two phenomena need to be considered separately.

The Role of Existing Theory. The case studies relied minimally on existing theory. Instead, the study analysis sought to develop its own grounded theory (Glaser & Strauss, 1965) of evaluation utilization, deduced entirely from the interview data. We were largely satisfied with this strategy in the case study phase of the research.

Our later efforts have drawn more heavily on extant theories of school organizational behavior (some of which were only just emerging during the years covered by the case study project) to help explain the lack of school attention to evaluation data. We feel today that a case study approach which attempts to draw upon relevant theory is quite desirable. At a minimum, existing theory can inform data collection, specify variables of interest, and suggest alternative interpretations of events. Analysis should be grounded in and thoroughly substantiated by case data, and it should not bend the facts to fit presuppositions. But existing theory, properly used, is a wellspring of analytical concepts.

Summary Review

The Evaluation Case Studies accomplished their

intended purposes at a modest total cost. Even though relatively few informants were interviewed, the case studies yielded descriptions of local school evaluation that were far more detailed than any others then available.

The studies indicated that personal initiative and consultative, client-focused evaluation were associated with greater evaluation influence. This finding corroborated contemporary research (Patton et al., 1975).

The research methodology allowed school personnel to describe evaluation work in their own words, and it permitted an examination of school perceptions of evaluation's influence, strengths, and deficiencies. Through open-ended interviewing, the studies uncovered quiet, incremental uses of evaluation information, uses that might not have been detected by other means.

The principal improvement that could have been made in the case studies was simply to broaden the research -- to interview more school personnel, to spend more time in the schools, and to explore a wider range of issues. Had the scope of the studies been expanded, we might have learned more, more quickly, about the organizational and contextual influences upon evaluation. Nevertheless, the case studies, as conducted, were a substantial success.

Chapter 3

Evaluator Field Study

RESEARCH DESCRIPTION

The Evaluation Case Studies produced the analytical framework outlined in Appendix A, and isolated features of evaluation situations which seemed especially significant for evaluation utilization. The elements of the framework varied considerably in potential manipulability. Some, such as the evaluation's mandatory requirements (element 1.2 in Appendix A) or the relationship between the school and central district administration (5.1), were typically beyond an evaluator's control. But others, especially those related to the evaluator's approach, could be purposely manipulated. The research report (Alkin et al., 1979), clearly implied that evaluators should take into account factors such as those listed in the framework. However, it stopped short of formulating a recommended evaluation approach.

Within the Evaluation Use Project, however, possible recommendations were being considered. One key to increased local evaluation use appeared to be a consultative evaluation approach, one deliberately oriented to the needs of specific local program managers. In addition, it seemed that successful evaluators had made a commitment to care most about local users. Although they

carried out the officially mandated evaluation tasks with integrity, their emphasis was on the local consequences of evaluation.

The case studies suggested that a consultative, user-focused approach would stimulate local evaluation use. Yet there were important questions left unresolved by the CSE case study data. For example, were there important organizational prerequisites to be met before a consultative evaluation approach could be effectively implemented? The case studies had focused on the evaluators' work with programs, not on the organizational environment, so this question could not be answered readily. Assuming that consultative evaluation could reasonably be recommended, was it applicable to all the programs with which the evaluator might work? Might some programs rebuff the evaluator?

The Need for a Field Study of Evaluators at Work

A detailed study of school program evaluators at work was needed to examine the evaluation work environment and the nature of current evaluation efforts. Even at a purely descriptive level, such a study would contribute to the still limited database on evaluation practice. In addition, by working with the evaluators for an extended period of time, the researcher might come to see the organizational panorama as the evaluators saw it. One could explore with them the encouragements and constraints

that the organization placed upon their work, factors that might affect any attempts at consultative evaluation. Moreover, the evaluators' interactions with program managers and staff could be directly observed. One could discover whether evaluators were already seeking to adapt their work to local needs. If they were, one could see how their initiatives were received.

Field Study Methods

Overview

The Evaluator Field Study examined evaluation work as it unfolded and as it appeared from the perspective of the school program evaluator. The research relied primarily on on-the-spot observations and informal discussions with the evaluator and colleagues.

Three school program evaluators within the same urban district were observed as they carried out various evaluation activities. The fieldworker entered the study site -- a school district evaluation office -- and spent an initial orientation period "tagging along" with the evaluator selected for the study, observing and discussing his or her work but also concentrating on becoming familiar with the work setting. Thereafter, each evaluator's work was traced in more deliberate fashion, focusing on the stream of decisions and events surrounding various work tasks. Critical events were observed as they occurred, discussed with the evaluator, and followed to resolution.

From the observations and discussions, an evaluator's view of events was constructed.

Selection of Evaluators and Study Situations

Criteria for selecting evaluator-subjects were based primarily upon a conception of the kind of evaluator for whom the CSE Evaluation Use Project wished to develop recommendations. It seemed most appropriate to study in-house evaluators rather than external consultants. A consultative evaluation approach required extensive interaction between the evaluator and program personnel. In-house evaluators seemed more suited to this time-consuming role.

As a second criterion, subjects should be involved with program evaluations -- not simply district testing or the like. Third, if more than one evaluator was associated with a program, the the preferred subject was the senior evaluator, since he or she would be most likely to make decisions about the evaluation's focus. Finally, subject cooperation should be voluntary, not coerced.

Resources and time dictated that no more than two or three evaluators could be studied and that should be picked from no more than two districts. The choice between a single district study and a two district study was problematic. Concentrating all the effort in a single district would allow the greatest coverage of evaluation work in the district. Splitting effort between two

districts would allow cross-comparisons of organizational environments. As it happened, chance events determined the ultimate outcome: a single-district study.

The research began with a two-district strategy. One was to be an urban district, the other a smaller suburban one. Effects of the difference in size, structure, and community environment were to be explored. The search for two districts proceeded along fairly informal lines, as is usually the case in small-scale qualitative studies (Bogdan & Taylor, 1975). Geographic proximity and the existence of a likely contact person were important considerations. Metro Unified School District, the urban district, was a clear choice based upon these criteria. The other site, Santa Lucia (ADA approximately 20-30,000), was a suburban bedroom community.

Work began in both districts, but the Santa Lucia component of the study proved unsatisfactory. Contrary to initial understandings, the subject in Santa Lucia normally did little program evaluation work. The research field visits, in fact, were prompting him to consider initiating new evaluation work that would not have occurred otherwise. With regret, the Santa Lucia involvement was terminated. It was impractical to locate a replacement district at that point in the school year, and therefore all the remaining work was concentrated in the single urban district, Metro Unified.

Metro Unified School District

Metro Unified School District was promised anonymity in this research effort, so exact details of Metro's size, organizational structure, student or community population can not given. Metro was a large urban district in California. It had taken some recent steps toward decentralization, but decision making on important issues reputedly was still highly centralized. Its service population included students from all socio-economic strata and racial-ethnic backgrounds. Like several other California districts, Metro was engaged in an evolving integration effort. In addition, its bilingual service programs were rapidly expanding in response to demographic changes and state legislative requirements. Metro maintained an Evaluation & Testing Office, which handled evaluation related to compensatory education programs, coordinated district-wide achievement testing, and conducted special studies.

Initial Contacts in Metro District

In early May 1979, the CSE Evaluation Use Project Director, contacted the Metro's Evaluation & Testing director (an acquaintance), described the purposes and methods of the study, and inquired about the district's willingness to cooperate. The Metro Evaluation & Testing (E & T) Director was almost immediately receptive to the field study idea.

The E & T Director arranged for a meeting with Mrs. Elaine Bowman, an evaluation staff member fitting the evaluator-subject criteria. Few other staff in the E & T Director's immediate offices were assigned as program evaluators; most either held administrative positions or worked on special activities rather than program evaluation. Mrs. Bowman, however, was a recent addition to the staff, hired expressly to serve as the evaluator of the newly created Preschool Language Program for School Success (PLPSS).

The fieldworker met with the E & T Director and Mrs. Bowman in their offices to discuss the proposed research. The E & T Director remained supportive, and Mrs. Bowman seemed relatively unperturbed at the idea of being observed. The fieldworker accompanied Mrs. Bowman through the months of May and June, observing her work on the PLPSS evaluation.

In the fall, after the summer hiatus, contact was reestablished with the Evaluation & Testing Office. Mrs. Bowman had left Metro District (for a better position), and the fieldworker was directed to Mrs. Carrie Jenkins, an evaluation staff member within the Office's Compensatory Education Evaluation Unit. Mrs. Jenkins had an intriguing dual role as a "evaluation adviser" for a group of Title I schools and as the sole evaluator assigned to a special district-funded child service program. It was the latter

role that first attracted the Project's attention to Mrs. Jenkins. However, as the research progressed, Mrs. Jenkins' Title I duties took center stage in the research, and they were the focus of the final research report.

Contact was made by phone with the Compensatory Education Evaluation Unit's supervisor, who responded cooperatively -- probably because he received assurances from higher Evaluation & Testing Office administrators. (Certainly, the previous spring's smoothly accomplished research activities greased the wheels for this second entree.) The supervisor, in turn, conveyed his approval to Mrs. Jenkins, and the fieldworker called Jenkins to arrange a personal meeting. Mrs. Jenkins agreed to participate, perhaps with some trepidation -- although the research experience assuaged her concerns.

Meanwhile, it was discovered that the PLPSS was being evaluated once again by a new evaluator, Ms. Diane Grimes. Phone calls to the E & T Director and the Assistant Director (Ms. Grimes' supervisor), secured their approval for a further study of the 1979-80 PLPSS evaluation. Ms. Grimes and the fieldworker met in January 1980 and immediately began the new PLPSS research.

Thus, ultimately, data was collected on three evaluators from Metro Unified School District.

The Research Process

The research details varied, quite naturally, from

evaluator to evaluator. Here the general pattern of the research is described.

The very first days in the field were devoted to getting acquainted and getting one's bearings. Typically, the first meeting with the evaluator subject was something of a two-sided interview: the evaluator assessed the fieldworker's intentions, purpose, and style, and the fieldworker learned as much as possible about the evaluator's work and personal style. Of course, this mutual sizing-up was not completed at the end of the first meeting. Generally, the first few meetings had as part of their purpose the goal of acquainting the fieldworker with the evaluator's work situation.

For example, in the PLPSS case, visits to program classrooms were quickly scheduled. And in the case of the Comp Ed evaluator, the fieldworker was soon scheduled to observe a routine Title I liaison visit and to take a tour of the child service program with which the evaluator also worked.

During these initial excursions, the evaluators served as guides to the workplace -- suggesting interesting things to experience and making introductions to colleagues. Much of the initial time was also spent establishing the research relationship. Of course, the fieldworker was simultaneously becoming more familiar with the evaluator's job responsibilities and current activities.

With the Comp Ed evaluator, Mrs. Jenkins, the fieldworker usually spent about a half day for each observational session. This corresponded to the way Mrs. Jenkins usually scheduled her visits to the schools. Interspersed with these half day sessions were occasional meetings at Jenkin's office, to catch up on events that had not been directly observed, to plan future observations, or just to talk.

With Mrs. Bowman, the PLPSS's first evaluator, the fieldworker was able to observe many of the evaluator's activities during the six weeks in May and June 1979, when the bulk of the evaluation took place. Bowman was too busy, then, for lengthy interviews but conversed at length with the fieldworker on the way to and from meetings and school visits.

In contrast, Ms. Grimes' full year evaluation of the PLPSS required a different research approach. Grimes put in long stretches of desk work in the office. This activity was difficult to observe unobtrusively. The fieldworker settled on a routine of frequent short interviews and phone calls to check the progress of the evaluation. In addition, evaluation "events" were observed, for example, the evaluator's classroom observations, testing activities, interviews with administrators, and other trips to the schools.

Table I summarizes the data on "contact hours" (spent

interviewing and observing) and number of field visits. Telephone conversations are not reflected in these figures.

Table I
Contacts with Evaluators

Evaluator	Program	Time Frame	Contact Hours	Field Visits
Bowman	PLPSS	5/79-6/79	38	11
Grimes	PLPSS	1/80-6/80	46	15
Jenkins	Comp Ed	11/79-6/80	70	20

Depending upon Jenkins' and Grimes' schedules, some weeks during the 1979-80 school year were very busy, involving several sessions with each evaluator; other weeks were quiet. The decision to directly observe any given activity was a joint one. Some activities the evaluators recommended as likely to be interesting; others the fieldworker selected from the evaluator's upcoming schedule, based on the contribution they might make to the unfolding analysis of each case. The evaluators were given veto power over the fieldworker's suggestions, a power almost never exercised.

It was sometimes possible to take notes during the actual observations -- for example, during observations of meetings where other participants were making notes. In any event, detailed field notes were prepared immediately following each observational session.

The Using Evaluations framework (Appendix A and

Alkin et al., 1979, Chap. 9) identified situational characteristics to be observed. As the research unfolded, yet another list of potentially significant topics began to take shape. These were formulated and systematized in a series of draft topic lists (e.g., Appendix B).

Working hypotheses about the important features of the evaluation work were developed as data were collected. These working hypotheses could then be mapped against new data and refined, modified, or replaced.

The Research Report

Observations and discussions with the PLPSS and Comp Ed evaluators continued through the end of the school year. Following a complete review of the field notes, a draft of the research report was prepared in July and August. The three evaluators critiqued that draft in interviews with the fieldworker. Thereafter, a final report was prepared. It described the evaluation work carried out during the observational year, and it also discussed organizational characteristics which appeared to impede evaluation influence.

Review of the Field Study's Purposes

The Evaluator Field Study was a successor to the Project's case studies. It was intended to supplement the case studies by collecting more detailed information on evaluation's organizational context and by allowing direct observation of evaluation activities.

The case studies had suggested that evaluation was most successful when the evaluator worked consultatively with local program administrators, attempting to address their interests and to involve them actively in evaluation planning and data interpretation. Personal initiative also was important to evaluation success, particularly evaluators' and administrators' joint initiative to create a consultative relationship, as well as administrators' resolve to use systematically collected information in decisionmaking.

But, was it realistic to advocate consultative evaluation and personal initiative as answers to the problem of evaluation underutilization? Would organizational realities allow consultative evaluation to be implemented on a broad scale? Could individual evaluators or program administrators introduce this style of work on their own initiative, or would institutional changes be required? Answers to these questions required

more data on evaluation's organizational context and on evaluators' interactions with school personnel. The field study's on-the-spot observations of evaluator work were intended to collect this data.

There was a subsidiary motive for the Evaluator Field Study. Quite simply, it seemed wise to look first-hand at school evaluation work. Although the case studies had gone well, there were hazards to basing analysis so heavily on interview data. The Evaluator Field Study, with its observational work, was to be a check on the case studies' findings.

Research Design Decisions

Because the case studies had suggested steps individual evaluators might take to increase evaluation usefulness, the field study sought to examine closely the work of individual evaluators. This suggested a study focused on individual evaluators rather than a more general study of an evaluation unit's activities. Sample size would have to be small. most two or three evaluators could be studied in depth by the sole fieldworker.

The evaluators could have been selected from either the same or different school districts. There were arguments for either strategy. As stated in the preceding research description, chance events influenced the ultimate research design, in which three evaluators were studied, all from the single district, Metro Unified.

As in the case studies, research access was negotiated through personal contacts, and subject participation was voluntary. Participant observation requires extensive subject cooperation over a long period of time (in this instance, more than a year), making random site or subject selection quite difficult. An effort was made, however to avoid school districts noted for uniquely progressive evaluation practice. A more typical setting was sought, instead.

Standard participant observation methods were employed. Perhaps the only major methdological adjustment was that the office work being studied required more interviewing than had originally been planned (see Points of Concern, below).

Methodological Assessment

Descriptive Detail

The field study report described the subject evaluators' work and characterized school attitudes towards evaluation. Two of the evaluators studied worked full-time with a single program (the Preschool Language Program for School Success). Their work with this program was described chronologically, much as evaluation has been described, in the case studies. Program history and context were sketched, and evaluation, from initial design to final reporting, was detailed.

The field study report did provide quite detailed data

on the circumstances, events, and deliberations that affected the PLPSS's evaluation design. Compared with the evaluation case studies, more information was presented to explain why the evaluation had taken the shape it had. Some of this information was collected through conversations with the evaluator; some, through direct observation of planning meetings and similar events.

PLPSS evaluation findings were described. First-hand observations and discussions allowed a comparison between what the evaluators said about a program, either to the fieldworker or to clients and colleagues, and what they wrote in their final evaluation reports. Finally, some data was also collected on evaluation's apparent impact on the program.

The third evaluator-subject was one of several "evaluation advisers" for the district's ESEA Title I programs. This evaluator provided technical assistance to Title I schools rather than directly evaluating the schools' educational effectiveness. The description of the evaluator's work therefore was organized thematically, around the several varieties of technical service which were provided.

Throughout the study report, vignettes from the fieldwork were used to illustrate situations encountered by the evaluators or to provide more concrete and vivid detail on critical evaluation events. The use of vignettes was

analogous to the use of direct quotations in reports of interview studies.

In the description of the PLPSS work, portions of the annual PLPSS evaluation reports were quoted. These reports and other pertinent documentary materials were excerpted at length in appendices to the field study report. (Segments of these documents were excised so as not to disclose our host's identity.)

The vignettes, narrative descriptions of critical events, and excerpts from documents together furnished evidential support for the field study's descriptions of the Metro school environment. This amount of descriptive detail was available, in large part, because the field study examined contemporaneous events. Observations could be made, participants could be questioned as events transpired or soon thereafter, and miscellaneous memos and documents could be collected conveniently.

The field study report contained few direct quotations from participants, however. Conversations had not been tape recorded or transcribed by hand on the spot, so verbatim quotations were exceedingly difficult to capture. Paraphrases based upon the fieldnotes were more readily constructed.

In general, the Evaluator Field Study provided depth of detail rather than breadth of coverage. That is, activities in which the evaluators participated were

covered in great depth: observations, discussions, document reviews, and semi-formal interviews with the evaluators let us scrutinize these activities closely. Other activities which might affect the evaluator, but in which the evaluator did not directly participate, were not as well explored. For example, evaluation policy set by higher-level district management affected the evaluators. The field study explored the evaluators' perception of that policy, but policy-setting by management was not directly investigated. Similarly, activities internal to a school could affect evaluation, but schools were not an independent focus of the research. The field study, therefore, truly hewed to the evaluator's perspective on events, with resultant advantages and disadvantages.

Descriptive Accuracy

Field study data came from observations of the evaluators at work, from a great many conversations with the evaluators, and from a much smaller number of group discussions involving other school personnel, the evaluator, and the fieldworker. From this database, data could be substantiated to varying degrees.

Events directly observed by the fieldworker were well documented, of course. And the fieldworker could compare his own understandings of these events with the interpretations provided by the evaluators or others.

Other data dealt with events or circumstances which

were not directly accessible to the fieldworker: comments on history or current events which the fieldworker had not witnessed; assertions that a practice was standard, customary, or mandatory; or discussions of policy, expectations, or job requirements. This information could be examined for internal consistency and plausibility, and, since such topics arose frequently, there were many opportunities to ask for clarification or greater detail. But there were limits to the confidence one could place in this data.

Data quality depended greatly on the relationship established between the fieldworker and each evaluator. With the two evaluators whom the fieldworker followed in the 1979-80 school year, a solid working relationship gradually developed. By year's end, these evaluators appeared comfortable with the research, candid, and forthcoming. We were completely satisfied with their cooperation. The (PLPSS) evaluator who was studied for seven weeks in late spring 1979 was cooperative on most matters, but less open than the others. Importantly, from the many hours of observation and conversation, the fieldworker gained insights into the evaluators' interests, predispositions and biases -- thus, allowing more informed assessment of the evaluators' comments.

Without question, additional data from a wider set of informants -- school-based personnel, the evaluators'

colleagues and supervisors, and higher level district personnel -- would have aided the study. There would have been more opportunity to explore the contextual matters that evaluators described but that the fieldworker could not verify independently. Because the fieldworker did not wish to dilute the study's tight evaluator focus and his own close identification with the subject-evaluators, these other informants were not interviewed.

Taking all these facts into account, our best judgment was that the study yielded an accurate view of evaluation as it was known to the evaluators and to an observer (the fieldworker) operating from the evaluators' perspective. But it was also clear that the study offered a situated view of events, one limited by the fieldwork's emphasis on the evaluators' perspective.

Nature of the Field Study Analysis

There were district differences between the field study analysis and the evaluation case studies' analyses. The case studies' analyses were constructed from interviewees' explanations of evaluation influence in their programs -- explanations requiring a high level of inference and synthesis on the interviewees' part. The cross-case analysis then generalized from these explanations.

The field study analysis was constructed from lower-inference data: observations, discussions of

individual events as they occurred, judgments of the typicality of an attitude or procedure, etc. This raw detail was assembled into a description of the evaluators' work and a summary analysis of aspects of the school context that seemed to impede evaluation influence: loose administrative control of classroom instruction, resistance to outside intrusion into the school or classroom, and general hostility towards evaluation.

The field study analysis compared the findings regarding Metro with extant data and theory on school organizational behavior. From the literature, it appeared that the organizational characteristics observed in Metro were common in public schools. According to some theorists, these characteristics fit into a stable organizational pattern which was functional for school survival and growth.

The crux of the field study analysis, therefore, lay in matching the study data to existing theories of school organizational behavior. Theory, in turn, predicted obdurate resistance to evaluation activity and evaluation information, a prediction which conformed to the experience of Metro evaluators in their efforts to increase evaluation influence.

Generalizability

The evaluator field study described the work of three evaluators in one district, Metro Unified. Two of the

three evaluators worked with the PLPSS, a unique preschool program within Metro, and no claim was made that others pursued similar work.

The third subject was one of several ESEA Title I evaluation advisers in Metro. Conversations with the evaluator and otehrs indicated that the evaluator studied engaged in a representative set of work tasks with the Title I programs. This impression was also supported by the User Interview Survey's findings regarding evaluation services available to school administrative staff.

The Evaluator Field Study also described school staff attitudes towards evaluation and evaluators. Observations of group interactions indicated that the characterizations were broadly accurate. Independent confirmation came from the User Interview Survey, which discovered school attitudes consistent with those reported in the Evaluator Field Study.

The Field Study did not claim inherent generalizability beyond Metro Unified. But the study's findings were consistent with patterns of organizational behavior observed in other school systems, as described in the organizational theory literature. Therefore, there was reason to believe that the Metro findings might be applicable elsewhere.

Points of Concern

Lack of Interim Reporting. Comprehensive analysis

was deferred until the summer of 1980, when a first draft of the study report was prepared. Before then, analytical themes had been discussed informally within the Project, but no written synthesis had been undertaken. In hindsight, it might have been helpful to have written an interim research report, perhaps in March or April. If such a report had been written, subsequent data collection might have been refocused in order to explore more completely the emergent analytical themes.

Fieldworker Fatigue. The mechanics of participant observation were more timeconsuming and fatiguing than had been anticipated. Fieldwork sessions were exhausting because of the concentration required to collect data and simultaneously maintain a neutral, unobtrusive role. Field note preparation was a lengthy process, even when the notes were dictated rather than directly typewritten. Note preparation easily required twice as much time as the observation sessions.

Because data collection is so time consuming, it becomes very tempting to cut corners. Analysis is deferred. One procrastinates about coding field notes. More insidious, though, is the way that one may unconsciously simplify and routinize data collection by repeatedly observing the same types of events, talking with the same people, and visiting the same schools.

One needs the opportunity to step back from the

harried life of the data collector in order to think carefully and creatively about the research. In future participant observation studies, we would try to structure more opportunties for reflection and planning, by scheduling planning and analysis weeks, interim reports, and periodic rap sessions with colleagues, for example.

The Problem of the Office Environment. Much of a program evaluator's work takes place in an office setting, and there is a good deal of quiet desk-work or telephone conferencing. It is extremely difficult to observe these office activities unobtrusively. Is the fieldworker to sit nearby? If so, where? What is the fieldworker to do while others in the office are working? How can the fieldworker collect data about telephone conversations without becoming a nuisance or an embarrassment? There are no easy answers to these questions. When the evaluators held meetings or made visits to schools, observations could be made more comfortably. Then, the fieldworker was accepted with equanimity. But in a general office setting, the fieldworker's constant presence was troublesome.

The compromise solution was to observe scheduled meetings and school visits, but to visit the evaluation office setting primarily for brief drop-in calls before or after scheduled observations or for scheduled discussions with the evaluators. Two of the subjects had heavy schedules of meetings and school visits, so considerable

observation work was done. One evaluator had more office time; interviewing was more frequent in her case. A major difficulty with this approach was that the evalator's impromptu meetings in the evaluation office were not scrutinized.

Alternatives to traditional participant observation methods might be more suited to a professional office environment. For example, subjects might complete daily activity logs or they might be asked to dictate a daily journal, with their dictated tapes being collected and transcribed weekly. Observations and interviews could be interspersed with these other data collection methods, as appropriate. It seems important, in any event, to experiment with new methods for studying office work.

The Effectiveness of Person-Centered Ethnography

The Evaluator Field Study was person-centered in that it examined the worklife of selected individuals. Ethnography and participant observation more often explore activity within a physical setting (e.g., a hospital or prison) or social group (e.g., a youth gang), although precedent for person-centered ethnography exists. (e.g., Wolcott, 1973.)

In retrospect, we have some doubts about the efficiency and efficacy of person-centered ethnography. The method is forbiddingly intense. Observer and subject both value their privacy and find constant partnership

difficult. Tacit understandings limiting observational frequency emerge, and the observational work is likely to be less complete than the researchers originally may have envisioned.

To the method's credit, strong working relationships can emerge between the ethnographer and the subjects, which may increase subject candor. To a degree, the observer and subject may become co-investigators. But there are disadvantages, too. The fieldworker's close association with the subjects is very apparent, and the fieldworker may find it difficult to become a confidant to other persons in the setting. Also, it may be difficult to divert research attention to persons or settings detached from the original subjects, should the need arise.

In the Evaluator Field Study, for example, once the research commitment was made to the evaluators it would have been difficult to switch tactfully from evaluator-centered observation. Yet, the incoming data might have justified greater attention to school-based evaluation consumers, to the evaluators' colleagues, and to top administrators in the district office. In the field study, therefore, we encountered a problem similar to that encountered in the case studies, namely a sense that the informant pool could have been expanded usefully. In the field study, however, the initial research specifications (viz. person-centered ethnography) made it quite difficult

to include additional informants. If the research had been focused on the topic of evaluation linkages with school decisionmaking, rather than on the evaluators *per se*, then an expansion of the study to include these additional informants might have been accomplished more easily.

Summary Review

As hoped, the Evaluator Field Study yielded detailed information about the work context of local school evaluation, and the study was a provocative test of the earlier case studies' explanations for evaluation utilization. Confronted by the Metro data, we were forced to temper our faith in consultative evaluation as a remedy for evaluation underutilization. The field study's greatest asset was that it offered this opportunity to witness evaluation in the making.

Our principal regret was that data collection was confined, to a degree, by the evaluator-focused approach that was taken. Greater flexibility to follow the emergent research leads would have been helpful. In any future applications, we would be inclined to experiment with a broader range of data collection activities, although we would not wish to forego the observational work, which was so useful.

Chapter 4

User Interview Survey

RESEARCH DESCRIPTION

The User Interview Survey, described here, examined the views of local school evaluation users: school principals and mid-level staff within schools. The interviews sought to discover the primary issues, decisions, or concerns occupying school staff; to determine if evaluation information was relevant to, or influential in, school decision making related to these primary concerns; to acquire information on the uses to which evaluation information was put; and to obtain staff opinions about current evaluation services.

One motivation for the User Interview Survey was a concern that the previous research might have overstated evaluation's importance in the schools. The initial CSE case studies had focused tightly on evaluation activities; from the studies it was difficult to place evaluation in proper perspective. Was staff attention mostly directed elsewhere and only occasionally turned to evaluation-related matters? When program decisions were made, how much discussion and deliberation took place and what role did evaluation data play? Interviews could address these questions.

In addition, the interviews could reach a larger, more

representative, sample of school personnel than had the case studies. Staff views on evaluation strengths and weaknesses, on evaluation influence, and on potential evaluation improvements could be systematically explored. And by conducting the survey in the same district chosen for the Evaluator Field Study, the evaluators' and the local school users' perspectives could be compared.

Study Procedures

Selecting the Respondent Sample

Defining the Population. Selecting a respondent sample required operationalizing the term "local school evaluation users". Several sampling decisions were quickly determined by the research circumstances. Because the User Interviews were intended to complement the Evaluator Field Study, Metro Unified School District staff would be sampled. However, to cover Metro, an urban district, quite a number of interviews would be required. With the resources available, Metro alone could be studied.

The Project's prior focus on evaluation in specially funded programs directed attention to users connected with such programs. More than this, however, there was the simple fact that Metro did little program evaluation except of specially funded programs. The study was further limited to schools receiving Title I funding. One could be sure such schools had experienced evaluations, since Title I requires them, and the Title I program offers

a large pool of schools from which to sample.

Three individuals were interviewed at each school site, in part to obtain overlapping responses from multiple informants in order to "triangulate" the data but also in part because Metro's Title I schools seemed to have multiple important decision makers (Daillak, 1980). The school principal was interviewed in every case. In addition, two other persons holding influential positions relating to the school's programs were selected. Various school personnel fit this description; a working list of all acceptable job titles was developed.

As one of the two additional interviewees, the staff member coordinating the school's special programs was selected. (This was usually the Title I Coordinator, Special Program Coordinator, or an Assistant Principal.) The final respondent at each school was to be someone involved in administering the special program, although not necessarily at a senior level. In a large school, this individual's duties might be entirely administrative. In a small school, it was often necessary to interview a Resource Teacher, Curriculum Supervisor, or Bilingual Coordinator -- individuals who usually had teaching responsibilities in addition to their administrative duties.

Contacting the School District. The Metro superintendent was contacted. He approved the project,

committed the District to participation, and directed the Evaluation and Testing Office (E & T) to assist in sample selection.

An oversize sample was initially selected, to allow for attrition. Twenty-eight Title I elementary schools were randomly selected, although a final sample of only twenty schools was desired. The district compiled the desired sample, which included schools from all geographic areas of the district as well as schools of diverse size and ethnic composition. Each school principal received a letter from the Superintendent informing him that his school had been selected. The letter briefly described the study, endorsed its purposes, and vouched for the researcher's credentials -- but also made it clear that school participation was completely voluntary.

The school principals were telephoned, reminded of the Superintendent's letter, and asked when a research team member might conveniently conduct the interviews. Initial calls were made in the sequence in which the school names appeared on the list received from the District. Return calls were often necessary, however, in order to reach the principals. All but two of the principals contacted were willing to participate, and sample selection was halted once 20 principals had committed themselves and their schools to the study. This sample was was augmented with two additional schools, selected from among those served by

the Compensatory Education evaluation adviser studied in the Evaluator Field Study. Thus, ultimately, 22 schools participated in the study.

Interview Strategy and Format

Strategy. To place evaluation in context, one must recognize that school-site decision makers have many pressing concerns other than evaluation. Numerous issues compete for staff attention, and it seems likely that many of these escape examination in the typical school program evaluation. Therefore, to take the local informants' point of view, the respondents were first asked to describe recent significant program occurrences. Next, the program's recent evaluations were discussed. Finally, the interviewee was to expatiate on evaluation's general usefulness, strengths, and problems.

The rationale for this strategy was simple. If the interviews had opened with direct questions about evaluation, they might have led the informants to overstate the importance of evaluation-related issues. Therefore, the interviewees were to identify "significant program occurrences" first and discuss evaluation later. Also, hinging much of the interview on specifics -- significant program occurrences and recent evaluations -- helped avoid the generalities that plague abstract discussions of evaluation's Virtues, Faults, and Impacts.

Why the phrasing "significant occurrences" rather

than, for example, "significant decisions" or "significant concerns"? "Decisions" seemed too narrow, and perhaps even unrealistic. Following Weiss (1980), it appeared that local school personnel might have difficulty thinking of themselves as decision makers and identifying decisions they had made. Weiss argues that, in bureaucratic organizations, policy actions often are not decided but rather accrete in a gradual flow of "small uncoordinated steps taken in many offices -- by staffs who have little awareness of the policy direction that is being promoted or the alternatives that are being foreclosed" (p. 382).

If decisions seemed too narrow and idealized, then concerns seemed too negative and issues too much in the realm of ideas or attitudes, possibly not yet actualized. The term, occurrences, was a compromise. A "significant occurrence in the life of the program" seemed more likely to be something that informants could identify, discuss, and analyze. It connoted a change or departure from the ordinary stream of activity in the school -- an opportunity for influence, something that evaluation might (or might not) have affected.

By having respondents discuss these self-selected occurrences, one could investigate several matters: what the respondents felt were important decisions or events; how the school went about handling the occurrences; and how evaluation or other information sources were used to deal

with the occurrences.

Format

Data Collection in Perspective. The Questionnaire-Jawboning Continuum is one construct for thinking about the use of structure in data collection. At the questionnaire end of the continuum, data collection is quite structured. All respondents face the same questions, which appear in predetermined order. In a selected-response questionnaire, idiosyncratic responses are not permitted. The respondent cannot volunteer information beyond that asked for in the instrument, and the data collector cannot tailor the interaction to the individual respondent. While this data collection method offers comparability across subjects, its sensitivity is limited to the choices within the questionnaire.

"Jawboning" defines the other extreme of the research continuum. In jawboning, neither questioner nor respondent is bound by external structure. In fact, the notion that one person is the questioner and the other the respondent, is misleading. Jawboning is more nearly an unstructured conversation between two persons. Neither party has a specific agenda, and the conversation is guided only by its own internal logic. Jawboning is rich in detail, since the participants are free to exhaustively explore any given topic. However, since each conversation takes its own unique course, data from "jawboning" lacks comparability

from subject to subject.

Between the two extremes, there are other data collection options. For example, in a "standardized interview", questions can be carefully scripted, but the interviewee can be allowed open-ended responses. Or, as another option, an interviewer might be allowed to conduct a seemingly freeflowing conversation with the subject, after which the interviewer might complete a structured, forced-choice questionnaire reporting on the discussion.

The Topic-Centered Interview. Initially, a standardized interview providing for open-ended responses was considered. This choice ultimately was rejected as too rigid to fit the diverse "significant occurrences" which the respondents might choose to discuss. Instead, a topic-centered interview format was chosen. Such a format modestly structures the interview -- by outlining the topics to be covered -- but it leaves question phrasing up to the interviewer. The respondent is almost entirely unfettered, except as the interviewer may refocus the respondent's remarks or move the discussion along to other topics. Thus, the topic-centered interview offers great flexibility within a guiding framework.

Patton (1980) terms the same method the "interview guide" approach, and he states its function rather well:

> An interview guide is a list of questions or issues that are to be explored in the course of an interview. An interview

> guide is prepared in order to make sure that basically the same information is obtained from a number of people by covering the same material. The interview guide provides topics or subject areas within which the interviewer is free to explore, probe, and ask questions that will elucidate and illuminate that particular subject. Thus, the interviewer remains free to build a conversation within a particular subject area, to word questions spontaneously, and to establish a conversational style -- but with the focus on a particular subject that has been predetermined. (p. 200)

The Interview Survey's topic guide is displayed in Appendix C. The brevity, indeed the almost skeletal quality of the guide, underlines the key ramification of the method: interviewer training must be comprehensive and thorough. The training, with its supporting materials (Appendices D & E), inculcates in the interviewers the purpose of the interviews, specifies the information which should be sought out under each topic, and prepares the interviewers to secure useful data. The brief written guide simply cues the interviewers, reminding them of the interview structure.

Interviewer Selection and Training

The Research Team. The interviewers were drawn from a group of advanced students enrolled in a UCLA graduate seminar on evaluation utilization. All students received five weeks of training. Five interviewers and five "validators" were selected to be members of the research team, together with the Project director. Those students

who had demonstrated the greatest interviewing proficiency during training, and who had prior public school experience, were designated as interviewers. The validator's role is discussed in a subsequent section.

The Training Program. Training involved four phases: (a) familiarizing students with evaluation utilization research; (b) developing general interview skills; (c) familiarizing students with Metro District administration, organizational structure, and terminology; and (d) training students in the specific purposes and procedures of the User Interview Survey.

All the students read and discussed Using Evaluations: Does Evaluation Make a Difference? (Alkin et al., 1979). All had previously read Utilization-focused Evaluation (Patton, 1978). In addition, they reviewed other articles on evaluation utilization, including works by Weiss, Caplan, Braskamp, David, and others.

Interview skills training was conducted by a UCLA faculty specialist in qualitative methods. Training sessions involved lectures, videotapes of model interviews, practice interviews, and discussions.

To familiarize the trainees with the Metro research context, Daillak, the investigator in the Evaluator Field Study, lectured on the organization of the Metro Evaluation and Testing Office, and onthe evaluation activities

commonly found in the schools. A glossary of common school terms relating to special programs and evaluation was presented.

Finally, the training focused on the detailed study procedures. Supporting materials were developed, including an interview topic description, a mock interview narrative, a one-page topic guide, and summary data reporting forms for interviewers and validators.

The interview topic description explicated the topics to be covered in the interviews (Appendix D). The mock interview narrative was a facsimile transcript of the interviewer portion of an interview (Appendix E). The topic guide was a one-page topic summary, to which interviewers were to refer during the actual interviews. (Appendix C). The summary data reporting forms (Appendix F) will be described in a subsequent section.

The trainees, together with outside experts, helped revise the training materials. In addition, interview procedures were pilot tested in two schools, and training materials were modified, as necessary. For example, improvements were made in topic sequencing, suggested phrasing of interview questions, etc.

The trainees reviewed and discussed transcripts and audio tapes from the pilot interviews. Questioning strategies and question phrasing were considered during these sessions. The relative importance of each topic was

discussed; interview time allocations to the topics were suggested, along with procedures for keeping the interviews on target.

Before his or her first school interview, each interviewer conducted an hour-long simulated interview with a study team member who role-played a school decision maker, mimicking the kinds of responses that had been encountered in the pilot interviews. During the interview, the mock interviewee took notes on the interviewer's questions, techniques, successful and unsuccessful tactics, and on information which the interviewer had neglected to request. At interview's end, interviewer and mock interviewee discussed the interaction, and the interviewee suggested areas for improvement.

The tapes from the first actual interview were used to train interviewers in data summarization. Each interviewer listened to the tapes from this interview and completed the data summary reporting form. The interviewers then met jointly to discuss the summaries, and, based upon the exercise, improvements were made to the reporting forms. Just as importantly, however, group discussion helped to standardize the data summarization process.

The Field Interviews

Scheduling Interviews. The interviews were scheduled by telephone two or three weeks before the interviewing was to begin. Each school was telephoned, and

the proposed interview procedures were discussed with the school's principal. Principals were told that three one-hour interviews were desired at each school. They were asked to identify two other staff members who were school level evaluation users, as defined earlier, and a tentative date and time for the interviews were set.

The three interviews at a school were scheduled consecutively, with a 15 minutes break between each. The schools were asked if they could arrange a quiet location for the interviews. School personnel were very cooperative.

Appointments were reconfirmed by telephone one or two days in advance. If the names and titles of the two subordinate staff interviewees had not been obtained previously, they were obtained during this second telephone contact. At one or two schools, a scheduled interviewee was unavailable the day the interviewer arrived, and the principal had selected an appropriate substitute satisfying the respondent selection criteria. In one of the 22 schools, chance circumstances on the day of the interview resulted in only two, rather than three interviews being conducted.

Conducting Interviews. Interviewing proceeded smoothly. Interviews usually took place in private offices. The first interview at a school was always conducted with the school principal. Before that interview

began, the rest of the day's schedule was always reconfirmed.

At the beginning of each conversation, the interviewers indicated that they wished to tape record the interview to ensure accuracy and to facilitate future analysis. There were no objections to this, although a few respondents asked that the machines be turned off momentarily during the one-hour sessions.

Data Analysis

Developing the Initial Data Base

It was theoretically possible to transcribe the interview tapes and use the written transcriptions as the base data. However, the 65 hours of tape would have produced an unwieldy volume of transcripts and required substantial clerical support. Instead of being transcribed, the interviews were summarized on standardized reporting forms (see Appendix F). The summary forms' structure paralleled the interview topic guide and interview topic description.

As soon as possible after finishing the three interviews at a school, the interviewer completed the summary forms, one for each interviewee. The interviewers referred to their written notes and listened to the interview tape as they completed the summary. Then they listened again in order to hand transcribe quotations that seemed particularly valuable. The final summary form,

therefore, contained five or six pages of narrative, organized by topic area, and as many as five or six additional pages of selected quotations.

Validating the Initial Data Base

Several strategies were employed to increase the interview summaries' accuracy. As described previously, tape recorders were used in each and every interview. There are arguments both for and against interview taping. In its favor, taping: (a) lets the interviewer concentrate on questioning rather than on transcribing the respondent's answers; (b) lets the interviewer focus his or her attention on the respondent and maintain a more natural, conversational interaction; and (c) permanently records the study's raw data.

Using the taped interview record, it was possible to obtain a second, independent summary of each interview and thus help confirm the interviewers' summaries. After the interviewers' summaries had been completed, the interview cassette tapes were given to a "validator." Working from the tapes alone, the validator completed summaries exactly like those used by the interviewer. Validators listened fully to each tape before beginning their summaries; then they replayed the tape while completing the summary forms. Validators also identified and transcribed key quotations, listening to the tape once more, if necessary.

Interviewer and validator summaries were compared, and

a procedure was developed for resolving inconsistencies. A panel, consisting of the interviewer, the validator, and a third research team member, examined both written versions of the interview and, if necessary, listened to the interview tape to resolve disputed issues.

Few differences were found beyond variations in the degree of detail with which observations were reported on the summary forms. There were two or three instances in all in which an interviewer and a validator reported contradictory information. Relistening to the tape recordings easily resolved the differences. The small number of discrepancies and their simple resolution were encouraging indicators of summary accuracy.

Nevertheless, a second quality check was introduced (on a sampling basis). The interviewees at the second school visited by each of the five interviewers were asked to comment of the summaries of their interviews. Copies of the interviewer summary forms were mailed to the respondents, and followup phone calls were made a week after the mailing, reminding respondents to return the summaries with their comments. Ten out of 15 summaries were returned. (It seems plausible that interviewees who disagreed with the summary would be more likely to return comments.) Four of the ten respondents made no corrections. Twenty-six comments were made on the remaining six forms.

The respondents' had few substantive differences with

the summaries. In most cases, the respondents simply took the opportunity to provide additional (i.e., supplemental or corroborative) information not mentioned in the original conversations.

Data Synthesis

Data synthesis entailed group discussions, simple quantitative tallies, and a procedure for "interrogating" the data base.

Group Discussions and Quantitative Tallies. The initial data synthesis began with wide-ranging discussions within the research team (i.e., among the principal investigator, five interviewers, and five validators). The team met weekly after completing the interviews and validation. Team discussions touched upon many topics, including what constituted "significant occurrences" for the respondents, what data sources seemed most important to the interviewees, and what interviewee reactions were to evaluation data.

After three group meetings, each team member drafted a report based solely on the interviews he or she had conducted or validated. Several tentative analytical themes emerged from the reports. These themes then were critiqued in a conference telephone call with an outside colleague (Michael Patton).

Some simple quantitative analyses were carried out immediately. For example, the respondents' job titles were

categorized and tallied. Other tallies provided a quick descriptive look at the kinds of "occurrences" discussed by the interviewees.

Using Case Experts. As the research team discussions progressed, questions became more complex, requiring more time-consuming study of each written interview summary than had been necessary for the sample tallies. To expedite this, each team member assumed responsibility for the interviews from approximately four schools. In effect, each team member became the case expert for about a dozen interviews. As far as possible, team members were assigned interviews with which they were already familiar, namely, interviews they had conducted or validated. The case experts became sufficiently familiar with their assigned interviews that they could quickly recite or locate relevant facts in the data summaries.

At this point, the data analysis discussions began to focus more tightly on a limited number of themes. Each team member took charge of one or more of the tthemes, by questioning fellow team members on relevant points during group meetings or by preparing short written questionnaires to be completed outside of the meetings. Acting as case experts, the other team members marshalled facts and quotations and provided page citations to the written summaries. The "theme leaders" drafted analyses of their chosen themes, which were then reviewed by the full

research team. The most promising drafts were expanded into Project working papers.

Review of the Interview Survey Purposes

The Interview Survey was designed to provide a broader perspective on school staff attitudes towards evaluation, and thereby complement the Evaluator Field Study's work in Metro Unified School District. The Evaluator Field Study was expected to provide only partial knowledge of school-level attitudes towards, and uses of, evaluation information, because the field study was so tightly focused on evaluators rather than school site personnel. To explore school perceptions more fully, we wished to interview at some of the schools served by the evaluators under study. And to place these interviews in proper perspective, we hoped also to interview personnel in other, similar schools.

An interview survey of school-level staff seemed an ideal strategy. With proper design, the survey results would be representative of Metro staff. Appropriate questions could be formulated based upon the case study findings and the Metro fieldwork. And the data from schools served by the field study's evaluator-subjects could be examined for its compatibility with the fieldwork findings.

Research Design Decisions

Although the Interview Survey was to complement the

Field Study it was also to be defensible on its own as a research investigation. Therefore, every effort was made to survey a representative sample of ESEA Title I schools in Metro and to collect comparable data from all the interviewees.

As anticipated, the Evaluator Field Study influenced Interview Survey Design. From the fieldwork, it appeared that instruction was seldom an issue for school-wide discussion, planning, and decision making. Administrative personnel in the schools seemed detached from instructional decision making and, therefore, not always interested in evaluation's findings about instructional effectiveness. In order to explore these indications, school planning and decision making was added as topics for the interviews.

The rationale for the interview guide has been described earlier in this chapter and will not be repeated here. The choice of interview length (one hour) was a compromise: long enough to establish some rapport and to move beyond pat answers to questions; short enough not to inconvenience the interviewees significantly.

Several considerations prompted the decision to tape record the interviews. In a one-hour interview it seemed unlikely that taping would make the subjects much more reserved or cautious than they otherwise would be -- thus vitiating the major argument against recording. In taping's favor was the promise of having a permanent

verbatim record of the interviews. Full transcriptions were too costly, however. As a compromise strategy for extracting quotable material and a written interview record, it was decided that narrative summaries would be prepared by the interviewer and illustrated with quotations selected from the tapes.

Methodological Assessment

Descriptive Detail

The Interview Survey's descriptive detail must be examined from two perspectives. First, one must consider the raw detail developed in the interviews themselves. Second, there was the detail transmitted in the study's analyses.

Raw data comprrehensiveness and detail varied widely from interview to interview and from topic to topic within each interview. Interviewer training had included a model allocation of time to topics, but a review of the tapes showed that the time allocations varied significantly. In some cases, the interviewees appeared to pull the conversation to a certain set of topics; other times, the interviewers appeared to guide the interview idiosyncratically. Irrespective of time allocation, detail varied depending upon the interviewer's skill and the interviewee's talent as an informant. Overall, the interviews covered the desired topics, but not always in the desired detail or proportion.

The study analyses drew upon the raw interview data in different ways. For example, mini-case studies were prepared for two schools (the schools in the sample that were served by the Compensatory Education education adviser from the Evaluator Study). These mini-case studies offered the most descriptively detailed presentations of interview data. They were developed from an intensive examination of the six hours of taped interviews in these schools.

Most of the remaining analyses relied upon the written interview summaries and the case experts' recollections of the subset of tapes each had auditioned. Several working papers were prepared. Two were classifications of interviewees, programs, and the "significant occurrences" discussed by the interviewees. These two papers were prepared from the interview summaries and did not make use of direct interview quotations. A third paper summarized Metro's major evaluation and testing endeavors. This summary was prepared from intra-Project discussions, was entirely narrative, and did not include illustrative vignettes or quotations.

The remaining working papers presented thematic analyses of the interviews. Recurrent themes were analyzed in the working papers and illustrated by quotations from the written interview summaries.

Descriptive Accuracy

There was little opportunity to independently verify

the accuracy of the interviewee's remarks. The interviews were relatively brief. The three interviewees within a school did not all discuss the same events or contextual conditions, reducing the opportunity for triangulation. And there were no follow-up interviews in which to probe, clarify, or cross-examine. As noted in the methods description, however, fifteen sampled interviewees were given the chance to comment on the written summaries of their interviews. This provided a check on the summaries' adequacy but did not insure that the interviewees' descriptions and interpretations were accurate.

Accuracy, in this situation, is best examined in the aggregate. Any description or opinion which recurred in several interviews was more likely to have some factual basis. Behavioral patterns, visible in the interviewees' descriptions of the way they made decisions in their schools, also seemed worthy of confidence.

There was a clear effort to check the data for consistency and plausibility. The analytical themes were discussed at length, too, and efforts were made to search out supportive and nonsupportive data. The case experts major responsibility was to serve as stringent critics based upon data in the interviews they had been assigned.

Nature of the Analysis

As already mentioned, the analysis sought to generalize from the interviews, identifying recurrent

opinions and behavior patterns, or broadly applicable interpretive themes. Four thematic analyses were written. One paper differentiated evaluation activities into three classes: formal evaluation, informal evaluation, and evaluation-like activities. Exact definitions of the classes were not given, but major exemplars of each class were enumerated. The paper's tripartite classification of these activities was "etic," i.e., imposed by the analyst and not indigenous to the respondent group.

A second paper summarized the reported attitudes toward, and uses of, several specific evaluation activities: state program reviews, needs assessments, standardized testing, etc. Information relevant to this paper was available in most of the interview summaries.

Another paper examined the interviewees' explanations for evaluation's influence (or lack of influence) in their schools, along with interviewees' suggestions for evaluation improvement. The author argued that three categories of factors recurred in the interview comments: proximity, competing demands on time, and psychosocial relationships. These categories were explicated and further subdivided. The implications of the subcategories for evaluation utilization were discussed.

The fourth analytical paper proposed that school principals' management styles could be differentiated. Two major styles were discerned among the interviewed

principals, and it was argued that the styles were associated with different patterns of evaluation use. The styles were not defined unambiguously, however.

Generalizability

Sample size and representativeness were more satisfactory in this study than in the case studies or the Evaluator Field Study. Twenty-two ESEA Title I funded schools were sampled randomly from within Metro school district, and sixty-five interviews were conducted. We believe that Interview Survey adequately captured the attitudes and perceptions of Metro Title I personnel.

Generalizability beyond the Metro population cannotserted be asserted rigorously. Many of the issues identified in the interviews have been mentioned by other school researchers, however, suggesting that the findings were not inherently idiosyncratic to Metro.

Points of Concern

Interview Heterogeneity. Recent reanalysis of the interview tapes has indicated that, across the 65 interviews, there was greater heterogeneity in topic coverage and emphasis than had been anticipated. Despite extensive training, interviewers did not always elicit appropriate information, and some interviewers systematically diverged from the intended topic emphases. As a consequence, missing or incomplete data was not an uncommon problem. It may be that this heterogeneity is

almost inevitable when a topic-guide interview is administered by several different interviewers. To correct the problem, one could either change to a standardized interview (i.e., one with a fixed question script) or reduce the number of interviewers. Two or three highly skilled interviewers, confering frequently, might hold to a common interview format more successfully. In future studies we would be inclined to select one of these options. If a panel of several interviewers was necessary, standardized interviewing would seem preferable.

Lack of Quantification. The thematic analyses suffered from a lack of tabulations or cross-tabulations of relevant evidence. For example, one analysis asserted that "competing demands upon time" were often cited as discouraging evaluation utilization, but no indication was given of how many interviewees mentioned this factor. Another paper related two different management styles to different patterns of evaluation. This relationship could have been illuminated by an appropriate cross- tabulation, but was not.

Two circumstances discouraged such quantitative displays. First, on any given topic, many interviews were inconclusive. To take one of the preceding examples, many interviews did not touch upon the question of time demands. One cannot know if these interviewees would have agreed that competition for time was a significant impediment to

evaluation influence. Nevertheless, a tally of the number of interviews which mentioned this factor might have been informative -- except for a second difficulty, the lack of full tape transcripts. A tally based solely upon the interview summaries could easily be in error. Only by examining the interview tapes or a full written transcript could one accurately determine how to tally an interview.

Lack of Full Transcription. In general, the lack of full interview transcriptions was troublesome. It impeded quantitative analysis, as indicated above. Moreover, the typicality of the few quotations which were noted in the summaries (often because of their pithiness or color) may have been overstated. Finally, if transcripts had been available on a timely basis, senior Project staff might have been able to monitor interviewer performance more readily.

Transcription is an onerous burden, however. Other researchers have reported an average four to one ratio of transcription time to tape time (Patton, 1980). Full transcripts would be a superior data analysis resource, but budgeting for transcription is difficult.

"Occurrences" versus "Decisions". In the interviews, respondents were asked to discuss two or three "significant occurrences" in the recent history of the program. The way these occurrences -- program changes, decisions, new activities, new requirements, etc. -- were

handled by the school was investigated in some detail.

. . In hindsight, we would have done better to ask the interviewees to describe recent significant decisions. That change in wording would have focused interviewers' and interviewees' attention on the decision making and planning activities associated with program change, topics which were our major interest. Unfortunately, the interviews often dwelled, instead, on the substance of the changes, and on how arrangements were made to implement new program activities. The need for this different wording might have been apparent if there had been a more extensive pilot test of the interviews, a test in which all the prospective interviewers had participated.

Subjectivity. The topic-guide interviews gave the interviewer considerable latitude, which we know affected topic coverage. One must wonder whether differences in question phrasing and in responsiveness to interviewee remarks could have subtly cued the interviewees and affected not only topic coverage but also the tone and content of the interviewees' responses.

Summary Review

The Interview Survey genuinely seeemed to evoke candid, interesting data from interviewees. During the hour long interview, the interviewees appeared to become accustomed to the tape recording and seemed to speak honestly, and sometimes quite insightfully, about their

schools. At the conclusion of the study, we felt that much had been learned about school decisionmaking and about the role of evaluation at the school level. Moreover, the number of interviews undertaken, and the sampling procedures, engendered confidence that the findings would be representative of Metro Title I school perceptions.

Yet as more rigorous and more quantitative analyses were undertaken, another side to the research experience emerged. It was extremely difficult to rigorously define terms or quantify the data. Holistic analysis was possible, and general analytic themes could be identified and communicated through judiciously chosen illustrations and quotations. But more systematic analyses were frustrated by the slipperiness of the concepts involved, by the imprecision of the language used by interviewees (and interviewers, as well), and by the diversity in coverage from interview to interview. The survey sampling design had held out the promise of useful, straightforward statistical analyses. The reality was more complex, more difficult, and less amenable to quantitative reduction and analysis.

Also, the use of the topic-guide interview approach was instructive. Even though the interviewers had had several weeks of group training, and despite the use of validators and standardized data summary forms, there was substantial variability in technique and emphasis from

interviewer to interviewer. Perhaps significantly, a brief pilot test was not sufficient to reveal this potential for variability.

In all, the Interview Survey convinced us of the power of interviewing, and also of the sizeable challenge in guiding such studies to a successful end. Fortunately, we wished to make an exploratory investigation of behavior and attitudes in Metro, and the research results met our needs quite well. If more stringent hypothesis verification had been the goal, instead, then the survey experience would have been less satisfactory.

Chapter 5

Epilogue

All the research efforts described in this report contributed significantly to our understanding of evaluation information use in local schools. Measured at this bottom line, all the methodologies have been successful. The methods have their differences, however, in the demands they place upon the researcher and in the uses to which they are most appropriately put.

The Evaluation Case Studies and the Evaluator Field Study employed methods which were almost complementary faces of the same coin. Both efforts were case studies, in fact, because the Evaluator Field Study was actually a case study of the evaluators' work. The studies differed in their primary mode of data collection, of course. The Case Studies relied heavily on interviewing; whereas the Field Study favored observation and on-the-spot, informal conversation. Both studies generated narrative datanotes.

The deficiencies of one study were the strengths of the other. The interviews in the Evaluation Case Studies gave us hearsay data, but from several strategically placed informants. The Field Study yielded compelling, first hand observations, but observations anchored to a narrow vantage point.

Neither interview case studies nor participant

observation require that the researcher have a strong prior understanding of the phenomenon under study. So both are suited to exploratory research. One can learn as one goes, although an awareness of alternative theories or conceptual frameworks can be invaluable as one tries to make sense of the data.

The Case Studies and the Field Study complemented each other so well that an amalgam of the two would probably have been superior for our research purposes. The case studies would have benefitted if the researcher had had the opportunity for a brief immersion in the culture of the schools studied. The field study could have profitted from a wider set of informants -- probably as interviewees. It seems foolish to segregate observations from interviews, although team research may be necessary to effectively reach distinct informant groups. The stimulation afforded by team research could also help to counter the routinization mentioned as a hazard in solo fieldwork.

Both case studies and participant observation typically focus on just a few cases or settings, making generalizability a concern. Findings from small samples can generalize, of course, but the problem is convincing a skeptic (or sometimes oneself) that the findings <u>will</u> do so.

The User Interview Survey, in contrast, yielded 65 one-hour interviews at 22 randomly selected schools. But

sample size and sample selection method were only the most obvious dimensions on which the Interview Survey differed from the other studies. Interview planning, instrument development, and interviewer training were all very demanding, requiring much more preparation and lead-in time than had the other studies. The selection of a topic-guide strategy eased the instrument development task somewhat, but it increased interview variability. Consequently, it was more difficult to carry out quantitative analyses which could exploit the method's inherent sampling advantages.

The thematic, non-quantitative approach taken in the Interview Survey analysis has substantial precedent (e.g. David, 1978; Kennedy et al., 1980). We were satisfied that useful results had emerged. Procedures for thematic analyses are not codified, however, and one hesitates to make the same claims for the generalizability of an analytical theme that one might make for the generalizability of a quantitative result from a sample of similar size. We believe that more attention should be given to the problem of analyzing this type of interview survey.

All three methodologies have, as we have said, contributed to our understanding of evaluation information use in the schools. None of the methods is clearly superior; all have advantages and disadvantages. Just as clearly, researcher judgment must be exercised at many

points throughout the studies. In the research descriptions and the critiques, we have discussed the decision making which shaped the research, and the ramifications of the decisions made. In hindsight, there are methodological decisions which we might make differently today. But none of the choices which were made were disastrous to data quality or validity; their effect was actually quite modest. On balance, we believe that we have been well served by our qualitative explorations of school evaluation activity. From reassessments such as this one, we hope to be able to design even better studies in the years to come.

References

Alkin, M.C., Daillak, R., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills: Sage Publications, 1979.

Alkin, M.C., Stecher, B., Daillak, R., with P. Stern, D. Davis, J. Horowitz, & E. Appel. Interview survey of users: Interim Report. Los Angeles: Center for the Study of Evaluation, University of California, October 31,1980.

Becker, H.S. Problems of inference and proof in participant observation. American Sociological Review, 1958, 23, 652-660.

Bogdan, R., & Taylor, S.J. Introduction to qualitative research methods. New York: Wiley-Interscience, 1975.

Daillak, R.H. A field study of evaluators at work. (CSE Report No. 154). Los Angeles: Center for the Study of Evaluation, University of California, 1980.

David, J.L. Local uses of Title I evaluations. (Research Report EPRC 21). Menlo Park, CA: Educational Policy Research Center, SRI International, 1978.

Davis, H.R., & Salasin, S.E. The utilization of evaluation. In E.L. Struening & M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills: Sage Publications, 1975.

Glaser, B.G., & Strauss, A.L. The discovery of substantive theory: A basic strategy underlying qualitative research. American Behavioral Scientist, 1965, 8 (6), 5-12.

Guba, E.G. The failure of educational evaluation. Educational Technology, 1969, 9 (5), 29-38.

Guba, E.G. Toward a methodology of naturalistic inquiry in educational evaluation. CSE Monograph Series in Evaluation (No. 8). Los Angeles: Center for the Study of Evaluation, University of California, 1978.

Kennedy, M.M., Apling, R., & Neumann, W.F. The role of evaluation and test information in public schools. Cambridge, MA: The Huron Institute, August, 1980.

Mann, J. The outcome of evaluative research. In C.H. Weiss (Ed.), Evaluating action programs. Boston: Allyn & Bacon, 1972.

Miles, M.B. Qualitative data as an attractive nuisance: The problem of analysis. Administrative Science Quarterly, 1979, 24, 590-601.

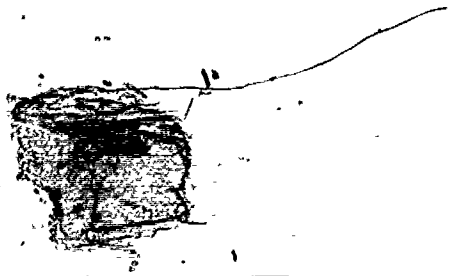
Patton, M.Q. Utilization-focused evaluation. Beverly Hills: Sage, 1978.

Patton, M.Q. Qualitative evaluation. Beverly Hills: Sage, 1980.

Patton, M.Q., Grimes, P.S., Guthrie, K., Brennan, N.J., French, B.D., & Blyth, D.A. In search of impact: An analysis of utilization of federal health evaluation research. Minneapolis: University of Minnesota, 1975.

Weiss, C.W. Knowledge creep and decision accretion. Knowledge: Creation, diffusion, and utilization. 1980.

Wolcott, H.F. The man in the principal's office: An ethnography. New York: Holt, Rinehart, and Winston, 1973.



FRAMEWORK FOR ANALYZING EVALUATION SITUATIONS*

Category 1: Preexisting Evaluation Bounds

- Property 1.1 School community conditions
- Property 1.2 Mandated bounds of an evaluation
- Property 1.3 Fiscal constraints
- Property 1.4 Other nonnegotiable requirements

Category 2: Orientation of the Users

- Property 2.1 Questions or concerns about the program
- Property 2.2 Expectations for the evaluation
- Property 2.3 Preferred forms of information

Category 3: Evaluator's Approach

- Property 3.1 Use of a formal evaluation model
- Property 3.2 Research and analysis considerations
- Property 3.3 Choice of role
- Property 3.4 User involvement
- Property 3.5 Dealing with mandated evaluation tasks
- Property 3.6 Rapport
- Property 3.7 Facilitate and stimulate the use of information

Category 4: Evaluator Credibility

- Property 4.1 Specificity
- Property 4.2 Changeability

Category 5: Organizational Factors

- Property 5.1 Interrelationship between site and district
- Property 5.2 Site-level organizational arrangements
- Property 5.3 Other information sources
- Property 5.4 Teacher and staff views
- Property 5.5 Student views
- Property 5.6 Costs and rewards

Category 6: Extraorganizational Factors

- Property 6.1 Community influence
- Property 6.2 Influence of other governmental agencies

---

* From Alkin, M.C., Daillak, R., & White, P. Using Evaluations: Does Evaluation Make a Difference? Beverly Hills, California: Sage Publications, 1979.

FRAMEWORK FOR ANALYZING EVALUATION SITUATIONS*

Category 1: Preexisting Evaluation Bounds

- Property 1.1 School community conditions
- Property 1.2 Mandated bounds of an evaluation
- Property 1.3 Fiscal constraints
- Property 1.4 Other nonnegotiable requirements

Category 2: Orientation of the Users

- Property 2.1 Questions or concerns about the program
- Property 2.2 Expectations for the evaluation
- Property 2.3 Preferred forms of information

Category 3: Evaluator's Approach

- Property 3.1 Use of a formal evaluation model
- Property 3.2 Research and analysis considerations
- Property 3.3 Choice of role
- Property 3.4 User involvement
- Property 3.5 Dealing with mandated evaluation tasks
- Property 3.6 Rapport
- Property 3.7 Facilitate and stimulate the use of information

Category 4: Evaluator Credibility

- Property 4.1 Specificity
- Property 4.2 Changeability

Category 5: Organizational Factors

- Property 5.1 Interrelationship between site and district
- Property 5.2 Site-level organizational arrangements
- Property 5.3 Other information sources
- Property 5.4 Teacher and staff views
- Property 5.5 Student views
- Property 5.6 Costs and rewards

Category 6: Extraorganizational Factors

- Property 6.1 Community influence
- Property 6.2 Influence of other governmental agencies

---

* From Alkin, M.C., Daillak, R., & White, P. Using Evaluations: Does Evaluation Make a Difference? Beverly Hills, California: Sage Publications, 1979.

Category 7: Information Content and Reporting

- Property 7.1 Substance
- Property 7.2 Format
- Property 7.3 Information dialogue

Category 8: Administrator Style

- Property 8.1 Administrative and organizational skills
- Property 8.2 Initiative

APPENDIX B

Observation Foci

R. Daillak
Rev. 4/80

OBSERVATION FOCI

1.0 Miscellaneous Evaluation and Testing Duties

2.0 Misc. Comp. Ed. Duties
2.1 Liaison Visits
2.2 Testing Services
2.3 Ongoing Planning and Evaluation
2.4 Needs assessment
2.5 School plans
2.6 Mock reviews
2.7 State Program Quality Reviews

3.0 Child Service Program Duties
3.1 Liaison Visits
3.2 State Preschool Evaluation
3.3 MH/SS Survey
3.4 Contextual constraints
3.5 Other data collection activities

4.0 PLPSS Duties
4.1 Pupil testing
4.2 Other data collection activities
4.3 Evaluating staff meetings

5.0 Evaluation Work Tasks
5.1 Learning one's job
5.2 Attending staff meetings
5.3 Planning and designing evaluation efforts
5.4 Preparation activities
5.5 Interactions
- 5.5.1 Exchanging information
- 5.5.2 Making recommendations
- 5.5.3 Building relationships

5.6 Collecting evaluation data
5.7 Analyzing evaluation data
5.8 Reporting, formally or informally
5.9 Giving public presentations
5.10 Representing the E & T Office
5.11 Giving misc. technical assistance

6.0 Special Topics
6.1 Attitudes towards evaluation and testing
6.2 Attitudes towards the evaluator
6.3 Bilingual education issues
6.4 Confidentiality
6.5 Situations marked by conflict
6.6 Evaluator's follow up activities
6.7 The role and functions of evaluation
6.8 Who initiates evaluation work?
6.9 Who is involved? Who participates?
6.10 Evaluation methodology
6.11 Organizational constraints
6.12 Regulations, Requirements, and Policies
6.13 Participants' roles
6.13.1 Client roles
6.13.2 Evaluator roles
6.14 (Open)
6.15 Sensitivity to clients
6.16 The use of forms, recipes, and other formal structures
6.17 Perceptions of evaluation success
6.18 Supervisors and supervision
6.19 The role of testing in evaluation
6.20 Timing and time constraints
6.21 Interpersonal relationships
6.22 "Compliance" issues
6.23 Especially significant individuals
6.24 Use of memoranda
6.25 Efforts to build evaluation capacity
6.26 Parents and the community in evaluation
6.27 "The Schools' Point of View"
6.28 The effects of positive vs. negative findings
6.29 Special evaluator characteristics

7.0 Fieldwork issues
7.1 Rapport
7.2 Fieldworker's asides

APPENDIX C

Interview Guide

Name of Interviewer: _____ School Code: _____

Respondent Code: _____

INTERVIEW GUIDE Title: _____

1. Introduction
   - Who we are interviewing
   - Why ("uses of information in special programs")
   - Confidentiality
   - Appreciation

2. Description of Specially-Funded Programs (Consolidated Project)

3. Duties & Responsibilities

4. "Significant Occurrences in the Life of the Program"
   - Changes (personnel, goals, materials, attitudes, etc.)
   - Rejected Alternatives

5. Factors Affecting Identified Occurrences
   - Description/History
   - Different Influences
   - Resolution Process

6. Role of Evaluation in Identified Occurrences

7. Role of Evaluation in General
   - Administrative Level (Within-school, District sponsored, PQR & mock review)
   - Description
   - Influence on Action & Attitudes
   - Factors Affecting Impact
   - Improvement?

(Repeat if appropriate: 5. Role of Evaluation in Identified Occurrence)

8. Additional Comments

APPENDIX D

Interview Topic Description

Interview Topic Description

(Training Document)

Introduction to the Study

The purpose of this research is to determine the role that information, particularly evaluation information, plays in school level program decisions. It is difficult to ascertain the relative importance of evaluation information directly. Asking about evaluation tends to bias the respondents' recollections towards just those situations in which they did consider information from evaluations. Instead, the school-level decision makers will be asked to identify significant occurrences in the life of the school programs. The situations they select will be analyzed to determine the factors that affected their beliefs and actions. Among these factors may be evaluation.

Hour long interviews will be conducted with school-level administrators, who might be users of evaluation information. These will not be structured interviews with rigid protocols, but naturally evolving conversations guided toward certain carefully selected topics. The topic guide is outlined below. The precise wording of questions asked by each interviewer will not be predetermined, rather it will evolve within the topic framwork as part of the natural conversational style of the interviewer. Similarly, the exact ordering of questions will be an interactive function of many factors, including, for example, the focused or diffuse quality of the respondent's answers, etc.

Model Introductory Remarks

Hello, my name is _____. We are interviewing elementary school administrators to investigate the ways they use different types of information in school planning and administration. We are particularly interested in schools with specially-funded, supplemental programs.

I can assure you that everything we say in this interview will be strictly confidential, and any reports that are written will be completely anonymous. If you do not object, I would like to tape record our conversation. It allows me to capture your thoughts correctly, and makes our work much more accurate. However, if at any time you would like to stop the recording for a moment, please indicate that to me and I will turn off the machine.

I would like to start by asking you for a brief description of the specially-funded programs here at _____school.

Description for Interviewers

Topic Area 1: Specially-Funded Programs in the School

A basic knowledge of the nature and scope of the specially funded programs in each school is necessary to understand the context in which decisions occurred. Initially, only a very general description will be sought; specific details will be elaborated as part of the subsequent inquiry into selected events and occurrences.

Model Opening Question: I think the easiest place to begin is with a description of the program here at _____ school. Can you give me a very brief description of the programs you have here as part of the school's Consolidated Project?

Topic Area 2: User's Position and Responsibilities in the School

We also need to know each respondents duties and responsibilities in the school. In particular their administrative relationship to the school's special programs will be important. At the outset a very general description will suffice. Details will be obtained as specific decisions are investigated later in the interview.

Model Opening Question: Can you give me a general description of your job and what your duties are with respect to the programs you just described?

Topic Area 3: Significant Occurrences in the Life of the Program

This is a crucial question, for the respondent's answer will determine the situations on which the bulk of the interview will focus. Ideally, each respondent will be able to recall significant program decisions in which they participated. Realistically, however, the evolution of a school program is more a matter of

incremental change than formal "decision" events. Thus, each school administrator will be asked to identify two or three events that they believe were "significant occurrences in the life of the program(s)". Subsequently, the interview will focus on these occurrences and the factors that affected the described outcomes.

Model Opening Question: As is said at the beginning we're interested in the way information is used by school administrators. To talk about this I want to identify 2 or 3 particular situations. I would like you to think back over the past two years and try to recall two or three significant occurrences in the life of the program here at _____ school. I realize that this question is somewhat vague, but it is vague on purpose. I want to get your impression of what was importnat rather than mine. Try and recall a few different occurrences that you thought were significant in determining the shape and character of the program during the last two years. For now I'd just like to list two or three such occurrences. We'll discuss the details later,

sub topics:

--changes (personnel, organization, goals, curriculum, materials, activities, attitudes, other milestones, etc.)

--rejected alternatives

--reinforcements in points of views, attitudes

Topic Area 4: Factors Affecting the Specified Occurrences

To determine the relative contribution of evaluation information in the total decision context, the respondents will be

was not mentioned, the question will be specifically asked by the interviewer at this stage in the interview.

Model Opening Question: Did evaluation make any difference in this situation?

Topic Area 6: The Role of Evaluation in General

To this point, evaluation has appeared as a secondary consideration in the interview. The situations identified by the respondents were allowed to define the scope of the discussion. Now, evaluation will be considered in its own right, and the respondent's wider knowledge and contact with evaluation will be investigated.

Model Opening Question: We've discussed _____ and _____ in great detail, and I think I understand the important factors involved in those occurrences (brief elaboration). Dr. Alkin and I are particularly interested in the usefulness of information from evaluations. I'd like to ask you to shift your thinking from these specific situations to thinking about evaluation in general. Will you take a minute to recall the program evaluations that have gone on in the past year or two; then, try to tell me what impact they had on you and on the programs at the school?

sub topics:

--level (within school, district sponsored activities, PQR and "mock review")

--characteristics of the evaluation (formal/informal, content, style, personalities, method of communication, etc.)

--its influence (on actions, attitudes, etc.)

--improving evaluation usefulness

Topic Area 5: The Role of Evaluation in the Identified Situation

After the more extensive discussion of evaluation it may be appropriate to repeat the earlier inquiry into significant occurrences. Certain subtle evaluation influences may have emerged from the lengthier discussion which were overlooked previously.

Topic Area 7: Additional Comments

At the conclusion of the interview, there will be a brief open-ended discussion period. Respondents will be given the opportunity to modify or expand their previous comments and clarify any misinterpretations.

Model Opening Sentence: Before we conclude, I want to give you an opportunity to make any additional comments about our discussion. Is there anything you feel should be clarified or expanded with respect to the situations you identified, the various factors you singled out or about evaluation in general?

Thank you very much for your cooperation.

Revised 2-15-80

APPENDIX E

Mock Interview Narrative

User Survey

Mock Interview (Questions Only)

"Introductory Statement" ....let's begin"

"A simple place to start would be for you to give me general description of the special programs operating here."

"What I mean is your Consolidated Project. I imagine you are receiving funds from a number of different sources -- maybe Title I, or Miller-Unruh; School Improvement, whatever. I'd like you to tell me which programs are operating here and briefly describe what you are doing with the funds."

I see. You have Title I funds, School Improvement, and Title VII. Describe the Title I project for me briefly?"

"OK. What about Title VII?"

"And how are you using the School Improvement money?"

"Thank you. I think I have a general picture, but let me double check. I'll describe your project to you, and you tell me any ways in which the portrayal is inaccurate: 'The Title I funds are used.....' (brief restatement of project description)"

"I want to talk about some of these program areas in greater detail later, but first I'd like to know more about your particular role at the school. Can you give me a brief description of your job?"

"OK, and what are your responsibilities as far as the special programs are concerned?"

"I see, in general you have supervisorial duties for all three projects? Tell me, do you separate them in your mind, or do you act towards them as if they were just one unified project?"

"Which makes the most sense for me, to discuss them separately -- first Title I, then Title VII, etc. -- or to ask about the project as a whole?"

"Fine. Before we go on let me try to restate what you said. Your primary responsibility, the thing that takes up most of your time, is supervising and planning for the project."

"Thank you. I appreciate the correction. I don't want to oversimplify. Besides planning and supervising, what other kinds of activities are you called on to perform?"

"A sort of jack-of-all trades, then?"

"OK, I think I have enough of an overview to start. As we proceed I may ask you to clarify certain things for me. The next question is probably the most difficult one in the interview, because it is somewhat vague. Not only that, but I'm going to ask you to review the past two years of the project in your mind and pick out certain things for me. I would like you

to think back over the last two years and, from your administrative perspective, identify two or three significant occurrences in the life of the program. Later we'll talk about each of them in detail. Right now I just want a list of two or three occurrences that you thought were significant in deciding the shape and character of the project"

"OK, the State's decision to award you School Improvement funds was certainly significant. I guess I didn't make my question clear enough, because I want to know about events or deliberations or changes that happened within the school's discretion, not outside your control. Can you recall any of this type of occurrence and tell me briefly about it?"

"I realize that it's not exactly clear, but I'm being vague on purpose. I don't want my point of view about what is important to determine your answer. I'd want to learn your definition of what things were important. But I can give you some example of the type of thing that might have occurred."

"I want to know about important changes or modifications of program actions that you took here at the school during the last couple of years. They could be part of developing the project plan, they might be personnel changes, or modification in the project approach, or new materials, or things of that nature. If you think about the school program as it is now, I want to know the significant occurrences over the last two years that determined where you are now."

"That's what I had in mind. When you decided to reduce the number of resource personnel and hire more classroom aides. That's a perfect example of what I'm interested in. Can you recall one or two more such actions that you think were important in the shape of the project?"

"Good, you purchased a new set of individualized math worksheets and student record sheets. Can you think of one more example? It may not actually have been a change, but something you considered and then rejected. A decision not to do something."

"OK, if you don't recall anything else right now, these two examples will be excellent. Later on, if something else occurs to you, please let me know."

"I want to discuss each of these events in more detail. Let's go back to the first event you identified. What I would like you to do is to tell me how it came about that you decided to reduce resource teachers and hire more aides. I want to know all the factors that influenced this decision, both pro and con. Give me a much more detailed description of what happend."

"OK. It was at the time you were planning the annual project application for last year."

"When was that?"

"And what happened?"

"I see."

"When you say that the teachers wanted more aides, how did you know that? I'm not just trying to be difficult, but what I am particularly interested

in is how different people's ideas and wishes were communicated? How was the desire to have aides made known?"

"Let me go back a minute. Who sat on the committee that drafted the new application?"

"So, it was a three member committee with lots of large subcommittees, one for each area. In which part did the decision to go with more aides arise?"

"If it had been talked about for a long time and mentioned at faculty meetings in the past, what were the reasons why people felt this would be a good idea?"

"The teachers at the school were pretty convinced that it would help them do a better job."

"They said that dealing with all the IEP's had really become a headache. Would it be fair to say that the presence of all the LES/NES kids accelerated the need for more assistance with paperwork and small groups?"

"I hear you saying that there was a sort of general groundswell for more aides, and that paperwork was becoming impossible, as was classroom management, with all the LES/NES kids. Was there anything else that influenced this decision?"

"Well, for example, what did people think about the job the resource teachers were doing?"

"I see. There was some support for the resource teachers, but a lot of people didn't care that much? What was going to happen to them?"

"Oh, it wasn't really a question of firing them. How important was it that everyone knew there would be classrooms for them this year, and they wouldn't have to transfer to new schools?"

"Were there any outside influences?"

"By that I mean, at the area level or the district level were there suggestions to go with more aides?"

"But everyone knew that people were doing it a lot?"

"You mean it was sort of a trend in the district and everyone wanted aides?"

"Now it makes more sense to me."

"Let me try to summarize and you correct me: There was sort of a groundswell for aides in other schools, and teachers were really feeling the increasing pressure of classroom management tasks with all the IEP and stuff. The resource teachers were liked but their jobs were dispensible and there wasn't too great a personal loss since they could all still work here, so there was sort of universal desire for more aides. How would you modify that summary?"

"Right."

"One last thing. This seems to have been based on common knowledge and common sense. Was there any evidence that aides would make a difference?"

"What it boiled down to then was you knew you had to do something because the test scores were low, and aides seemed to fit the problem as you saw it."

"Fine, this has been very interesting. I appreciate all the detail you were able to recall. Now let's talk about the other situation for awhile--when the school decided to switch from your old math program to the individualized student inventory program, PRISM. This must have been a more complicated decision. Tell me about it."

"Uh huh."

"I see. Do they have these demonstrations very often?"

"Oh."

"Let me try and review these points one by one in my own words. First, your math test scores were poor. How did you know that?"

"When did you find out each year?"

"How bad were they?"

"Does the district give you any advice, or do you just get the numbers?"

"Does the evaluator give you any suggestions or just explain the test scores?".

"OK, so you knew you had to do something, and you saw PRISM demonstrated at the CMC conference. When you suggested it, did everyone automatically go along with your suggestion?"

"How did you present it to the teachers?"

"You made a case for PRISM. What ideas did other people have?"

"What's HappyMath?"

"Let me interrupt for a second. I don't really need an explanation in that much detail. I'm impressed that you remember this program in such careful detail, however. Let's get back to the actual choice. Weren't there any advocates for the old program?"

"What were their reasons?"

"How did you decide?"

"What did people base their votes on?"

"So PRISM had test results and HappyMath didn't? Were there any other inputs into the choice?"

"Is it true that you had the power to make this decision here at the

school without any comment from the area staff, or downtown or from the state?"

"The district was pushing more individualized programs, but did they suggest PRISM in particular?"

"In what way were they "pushing" for more individualized programs?"

"Individualization seems to be the big thing. Was HappyMath individualized?"

"So it was going to be one or the other, the old program was pretty much doomed by the times?"

"This whole process is very interesting to me. Were there any other factors involved in shaping people's opinions before they actually voted?"

"Were the results from PRISM at other schools good?"

"But the teachers at other schools liked it? How did that word get around?"

"How many teachers do you think went along just because you recommended it?"

"Realistically how influential do you think your particular point of view is?"

"Take a minute to think. Was there anything else that entered into the final decision?"

"The PTA committee preferred PRISM. How much did that matter? If they had preferred HappyMath would it have changed the decision?"

"Anything else?"

"Let me try and summarize. Please tell me if I have left anything out: Your test scores were low and you knew you had to do something. Though the district wasn't suggesting any program in particular, there was a push toward individualization.....etc."

"Good. I don't want to forget the fact that the teachers got to review the materials at the faculty meeting. And, though it was the math sub-committee that made the final recommendation, almost everyone really had an input."

"This has been very useful. Of all the things that went into this situation, which influences would you say were the most important?"

"The direct contact was the key in your mind?"

"Before we leave these two situations I want to ask about one thing in particular--evaluation. In the Math program case you told me that the district test scores really motivated you to make the change. Did evaluation or the evaluators influence that decision in any other way?"

"So, Mr. Bertrand confirmed your enthusiasm for PRISM. That at least told you you were on safe ground. Did he talk to the staff?"

"Then his influence was only indirect, and not too great?"

"What about the classroom aide decision. In what way did evaluation influence that situation?"

"I didn't think so, but I wanted to ask you specifically. Essentially there was no hard data or evidence for aides one way or the other, nor for resource teachers."

"I really appreciate the care with which you are approaching my questions. For the rest of the interview I'd like to change focus a bit. Is there anything else you would like to add before I switch topics?"

"Of particular interest to me is the role that evaluation plays in school program improvement and change. At this point I'd like you to think specifically about evaluation, not just in the two situations we discussed, but in any way that it affects the school. I'm going to ask you to think back over the same two year period and recall all the different contacts you've had with evaluation and evaluators. Then I'm going to ask you something about these contacts."

"First can you summarize for me the ways if any you come into contact with evaluation?"

"You get a regular visit every couple of months from the evaluator. Does that have to do specifically with the Title I program or with the whole Consolidated Project?"

"Besides that visit, how often do you see test scores?"

"Are these contacts helpful to you?"

"In what ways do they influence your decisions?"

"In what ways do they influence the program?"

"What is it about the evaluation that makes these visits so useful?"

"I want to structure this discussion a bit by asking you about three different levels of evaluation activity, and ways to improve each one. First, I'd like to know if there is any evaluation that goes on at the school itself, not district level or area level information, but things you might do yourselves that you consider to be evaluation?"

"Because your budget is so small, you let the district handle most of the formal activities. Are there informal things that are done as part of your school level program?"

"Then let's consider next the district level. What are the evaluation activities that affect your school which the district carries out?"

"Basically then, it's the every other month visits and the test score reports?"

"What kind of effect do these have on the program?"

"So the personality thing makes a big difference. Do you think that these district efforts could be made more useful to you?"

"How would you design the evaluation so it would be more useful to the schools?"

"Well, what information would you like to have?"

"What kind of support might they provide to make things easier for you?"

"If the reports were simpler, do you think people would actually read them?"

"That is a very interesting suggestion. I don't think I've ever heard that idea before, and it makes so much sense."

"Well, finally about the state level evaluation. What about the PQR?"

"Whoa, let me interrupt for a second. I know that these things are an administrative nightmare. I've heard a wide range of comments about their actual benefit--everything from 'totally useless' to 'though they were a burden, there were some important benefits.' How would you characterize the value of PQR to this school?"

"Let's separate the PQR itself from the Mock Review. Does that make sense?"

"All right then, what about the mock review? It comes first doesn't it?"

"What effect did it have on you and the school?"

"OK, let's talk about negative impressions and then the positive ones."

"You really think that, beyond the mere inconvenience, this had a negative effect on the actual school program?"

"OK, what about its other influences. Tell me about some of the things you gained from the mock review."

"Good. What about it made the information useless to you?"

"OK. Anything else?"

"Well then, if you were in charge and could redesign the mock review, how would you change it to make it more useful?"

"I'll note your objections to the length and such. If you could change things to make the information and the interaction more useful, what would you change?"

"I guess it doesn't make too much sense to talk about the mock review without also talking about the PQR. So tell me about it. How useful to you is it?"

"Well, exactly what, as best as you can recall, are the consequences of the PQR?"

"Well, does it affect the program in any direct way?"

"What about people's attitudes or opinions? What changes do you see as a result?"

"Which factors in the PQR interfere with its usefulness the most?"

"Which are its strong points?"

"If you could redesign the whole PQR process, what changes would you make to make it more useful to the school program?"

"Anything else?"

"Good. While this is fresh in your memory, I want to go back to the two situations we discussed earlier and ask you if you now recall any evaluation influences that you didn't recall earlier. In the case of the aides, did any of these various evaluation factors enter into that decision?"

"And in the case of the math program, PRISM, tell me about any other evaluation influences you didn't recall when we first talked."

"Fine. I doubted that you'd missed anything, but I wanted to double check. The subtle factors are the ones we forget the most easily."

"I'm just about through with my questions, but before we conclude are there any additional comments you want to make about the things we've discussed? Is there anything else you want to add about the two program situation or about the different factors you felt contributed to those decisions, or about evaluation in other areas?"

"In that case I want to thank you very much for your cooperation. You have been extremely helpful."

APPENDIX F

Interview Summary Form

Evaluation-User Survey

Interview Summary Form

Name of Interviewer: _____ School: _____

Respondent: _____

Title: _____

Step I. After completing the interview, but before listening to the recording:

1. Based on the complete interview, describe in one paragraph the specially-funded programs operating at this school.

2. In one paragraph, describe the respondent's duties and responsibilities, particularly as they involve the special programs you discussed.

3. In one paragraph <u>each</u>, describe the significant occurrences identified by the respondent and discussed in the interview.

Situation 1:

Situation 2:

4. For each occurrence: A. List (in approximate order of importance) the factors that influenced the final outcome.
   B. Summarize in one paragraph the interrelationships among these factors.

Situation 1:

Situation 2:

5. Was evaluation information a factor in each of these situations? For each occurrence summarize in one paragraph the role of evaluation.

Situation 1: _____

Situation 2: _____

6. In one paragraph each describe the respondents experiences with evaluation at the "school", "district", and "state" levels. Indicate a) type of evaluation, b) its influence/ usefulness, c) factors contributing to its influence, d) ways of improving evaluation.

"school level" activitities: _____

"district level" activities: _____

PQR & mock review: _____

7. Summarize in one paragraph any additional comments that were important.

8. In one paragraph describe the interview context--the salient features of the setting, the participants and the interaction.

Step II: Replay the interview tape. (Set the counter at zero 000 at the beginning of each new side.)

As you listen to the interview:

1. Make additions/corrections to the descriptive paragraphs you wrote in Step I.
2. Select important quotes to illustrate key features of the interview.
3. Write out the quotes on the following pages.
   a. First indicate in a sentence or two what is being discussed immediately prior to the quote, i.e., some context for the remark. If it is an answer to a particular question, give the question.
   b. Write the quote as accurately as you can.
   c. Don't forget to indicate the tape counter reading at the beginning and end of each quote.

Key quotes:

Topic Area _____ Tape Counter at beginning of quote _____ Side A B ?

Context/Question: _____

Quote: _____

Tape Counter at end of quote _____ Side A B ?

Topic Area _____ Tape Counter at beginning of quote _____ Side A B ?

Context/Question: _____

Quote: _____

Tape Counter at end of quote _____ Side A B ?

Topic Area _____ Tape Counter at beginning of quote _____ Side A B ?

Context/Question: _____

Quote: _____

Tape Counter at end of quote _____ Side A B ?

Topic Area _____ Tape Counter at beginning of quote _____ Side A B ?

Context/Question: _____

Quote: _____

Tape Counter at end of quote _____ Side A B ?

Topic Area _____ Tape Counter at beginning of quote _____ Side A B ?

Context/Question: _____

Quote: _____

Tape Counter at end of quote _____ Side A B ?

Topic Area _____ Tape Counter at beginning of quote _____ Side A B ?

Context/Question: _____

Quote: _____

Tape Counter at end of quote _____ Side A B ?

Topic Area _____ Tape Counter at beginning of quote _____ Side A B ?

Context/Question: _____

Quote: _____

Tape Counter at end of quote _____ Side A B ?