DOCUMENT RESUME

ED 202 890 TM 810 313

AUTHOR Alkin, Marvin C.; Stecher, Brian M.
TITLE Evaluation Information Use by Site Level Decision Makers: Methodological Issues.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE 13 Apr 81
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Analysis; *Decision Making; Elementary Education; *Interviews; *School Districts
IDENTIFIERS *Evaluation Research; *Evaluation Utilization

ABSTRACT

The general goals of the study were: (1) to obtain the decision maker's perspectives on evaluation, how he defines it, how he carries it out; (2) to better understand the role of evaluation in site level decision making; and (3) to identify important characteristics of evaluation and of the school setting that affect the level of evaluation utilization at the school site. The methodology of the Decision Maker Study is detailed, describing the selection of the respondent sample, the elaboration of the interview strategy and format, the training of the interviewers, the collection of the interview data, and the several analysis activities--some complete, others in progress. Multiple strategies for data analysis included directed discussions, selected reviews of written data summaries, human data banking and an independent "quotation analysis." This many-faceted approach gave added validity to the themes and patterns of interactions uncovered in the study. (Author/RL)

EVALUATION INFORMATION USE BY SITE LEVEL DECISION MAKERS: METHODOLOGICAL ISSUES

Marvin C. Alkin & Brian M. Stecher
Center for the Study of Evaluation
UCLA

The research reported here focussed on the prominent school-level figures in evaluation utilization -- site level decision makers -- in an attempt to determine how they are impacted by evaluators and by evaluation information. The study was part of ongoing research conducted under the auspices of the Evaluation Use project at the Center for the Study of Evaluation, UCLA. The general goals of the study were threefold: (1) to obtain the decision maker's perspectives on evaluation, how he defines it, how he carries it out; (2) to better understand the role of evaluation in site level decision making; and (3) to identify important characteristics of evaluation and of the school setting that affect the level of evaluation utilization at the school site.

We realized that, if we hoped to obtain a balanced perspective, it would be a mistake to begin our inquiry by focussing on evaluation per se. Of course respondents would discuss the effect of evaluation on a particular action if we directed our interview probes toward this issue. The real question was, if we

*Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, California, April 13, 1981.

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, 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.

asked them about factors that led to program change, or modification of attitudes about a program, would evaluation be mentioned at all? To avoid "loading" the responses toward evaluation's impact, we needed first to capture the school administrator's natural view of decision making, program change, etc., and then to explore the role that evaluation played therein.

Thus, we asked first the broader question of which factors affected program-related decision making at the school. We attempted to unravel the sequence of events that led up to certain significant program occurrences, and then to determine the relationships between various determinants of the action. We particularly hoped to find out how evaluation did or did not fit within the general school decision making framework.

In this paper we detail the methodology of the Decision Maker Study, describing the selection of the respondent sample, the elaboration of the interview strategy and format, the training of the interviewers, the collection of the interview data, and the several analysis activities -- some complete, others still in progress.

Selecting the Respondent Sample

Selecting a respondent sample required operationalizing what we meant by "school site level decision makers" and deciding upon a sample population of them. The decision maker interviews were intended to complement Daillak's (1980) field study of evaluators in "Metro District," so Metro was the logical place to begin. The Evaluation Use Project's historical concern with the evaluation of specially funded programs naturally directed our attention to "school site level decision makers" connected with such programs.

The next step in selecting a sample from Metro was to define explicitly what constituted a specially funded program and who was considered to be

its "school site level decision makers." Specially funded programs we defined as those receiving supplemental federal or state allocations. From the many such programs extant we decided, however, to limit this study by focussing on schools receiving Title I funding because: first, one could be sure such schools had experienced evaluations, since Title I requires them; and second, the program offers a large pool of schools from which to sample.

We needed to define the term, "school site level decision makers" and to select respondents within the Title I schools. We decided to interview three individuals at each school site, in part to obtain overlapping responses from multiple informants in order to "triangulate" our data (in the jargon of qualitative research) but also in part because Metro's Title I schools seem to have multiple important "decision makers" (Daillak, 1980). The school principal was interviewed in every case. In addition, two other informed persons holding influential positions relating to the school's programs were selected: a person with specific responsibility to coordinate the special program and a staff person involved in the administration of the special program.

We next contacted the Superintendent of Metro. He agreed that the project was worthwhile, committed the District to participation, and directed the Evaluation and Testing Office (E & T) to assist in sample selection. We asked for a preliminary random sample of 28 schools from the population of all Title I elementary schools, though we planned to conduct interviews in only 20 schools. Each of the 28 school principals received a letter from the Superintendent informing him that his school had been selected to be in the sample. The letter briefly described the study, endorsed its purposes, and vouched for the researcher's credentials -- but

it also made it clear that school participation was completely voluntary.

We began contacting schools by phone in the sequence of school names on the list received from the District. Almost every principal contacted expressed a willingness to participate, and we halted sample selection once 20 principals had committed themselves and their schools to the study. Then we augmented this sample with two additional schools, selected from those serviced by the compensatory education evaluator studied by Daillak (1980) in the companion study to this research.[1] Thus, ultimately, 22 schools participated in the study.

Interview Strategy and Format

Strategy

Beyond this focussed concern of examining evaluation practice and evaluation influence as seen by school-site decision makers, we also sought to place evaluation's contribution to the school's working environment in an appropriate context. Therefore to truly take the local informants' point of view, we needed to obtain their portrayal of some significant recent occurrences in the life of the program and what forces or considerations affected these occurrences. Then, after hearing the informants' naturally given account of these matters, we would probe for what evaluation did or did not have to contribute, and how great a role evaluation actually played.

The rationale for this guiding strategy was fairly clear. If we had opened the interviews with direct questions about evaluation, we might have been "leading the witness" to overemphasize issues treated in the evaluation and evaluation's significance to program operation. Instead, we had the interviewee identify significant program occurrences first and discuss evaluation later.

Why the phrasing "significant occurrences" rather than, for example, "significant decisions" or "significant concerns?" The term, "decisions," seemed far too narrow - and even unrealistic. (See for example, Weiss 1980). 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. Occurrences was a compromise - but an advantageous one. A "significant occurrence in the life of the program" was more tangible and more likely to be something that our three informants at each school could all recognize, 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.

Format

From the foregoing discussion, we can derive two basic qualifications for our choice of interview method. One is that the interviews need to elicit, with sensitivity and minimal distortion, the respondent's accounts of significant occurrences in their programs and of factors and events impinging on these occurrences. The other is that we did, nevertheless, have an agenda of specific interests which we wished to explore with the interviewees as we discussed both the nominated occurrences and their programs' recent evaluations. To meet these qualifications, we employed a "topic-centered" interview format.

The Topic-Centered Interview. Initially, we considered using a structured interview format with subjects being allowed open-ended responses, but we rejected this choice as too rigid to capture the diverse range of stories we expected to hear from our respondents. In its place, we selected what we term a "topic-centered" interview format. Such a format places

a modest amount of structure on the interviewer -- by outlining in a "topic guide" the topics to be covered -- but leaves specific questions and probes to the discretion of the interviewer. The respondent is almost unfettered, except as the interviewer may take steps to 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) discusses much the same method in his description of the use of an "interview guide":

> 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. (Patton, p. 200)

The brevity, indeed the almost skeletal quality, of the topic guide underlines the key ramification of using such a format: interviewer training must be comprehensive and thorough. The training inculcates in the interviewers the rationale and purpose of the interviews; explains exhaustingly detailed kinds of information which should be sought under each topic; and prepares the interviewers for the verbal interaction they must establish to secure meaningful, high-quality data. The guide, then becomes simply a set of cues to the interviewers, calling out the substructure of trained skills.

Selection and Training of Interviewers

The Research Team

The "interviewers" were drawn from a group of advanced graduate students in Research Methods and Evaluation at UCLA enrolled in a graduate seminar on Evaluation Utilization in the spring of 1980. All students participated in a five-week training sequence.

The selection of "interviewers" was based on a number of factors. First, it was important that the interviewer have some direct school experience. Actual work in a school setting for an extended period of time gave our interviewers a background for understanding nuances and subtleties of school-related decision making and provided a knowledge framework within which to pose questions. Second, we wanted to select, based on the principal investigator's observation, those trainees displaying the highest general maturity and interpersonal skills and the greatest interviewing skill. Most of the group performed at a high level on all dimensions; both those selected as interviewers and those selected as validators (The role of validators will be discussed in a subsequent section.) were very well qualified. On these bases, five interviewers and five validators were selected.

Interviewer Validator Training

Staff training involved four phases: (1) development of understanding of the evaluation utilization research; (2) development of general interview skills; (3) development of familiarity with the Metro District Administration, organizational structure, and terminology; and (4) development of familiarity with the specifics of this particular project and its procedures.

As a first step in training, the members of the research team were

provided with a common background on utilization research. All read and discussed Using Evaluations: Does Evaluation Make a Difference? (Alkin et al., 1979). All had read Michael Q. Patton, Utilization-focused Evaluation (1978) as part of their earlier training. In addition, trainees read other articles on utilization, including major pieces by Carol Weiss, Nathan Caplan, Larry Braskamp et al., Jane David, etc.

The interview training sessions were conducted by Harold Levine, Department of Education, UCLA, an anthropologist and expert on interviewing. His presentations involved lectures, and videotapes and discussion of model interviews. To familiarize the trainees with the Metro context, Richard Daillak gave a presentation about the organization of the Metro Evaluation and Testing Office, the activities commonly engaged in by the evaluators, and the kinds of assessment commonly found in the schools. A glossary of common school terms, particularly as related to special programs and their evaluation, was presented to each trainee.

The final, and major, phase of the training focussed on the detailed procedures of the study. A number of materials were developed specifically for this project. The interview topic description contained descriptive materials on the meaning and scope of the different interview dimensions. The mock interview narrative consisted of a complete facsimile transcript of the interviewer portion of an interview. The topic guide was a one-page summary of the main topic headings and areas of interest, and was designed for interviewers to use in the actual interviews. The data collection and reporting forms will be described in greater detail in a subsequent section on data aggregation.

When they were sufficiently familiar with the topic guide and with interview techniques, trainees reviewed and discussed transcripts and tapes

from the pilot interviews. Alternative lines of inquiry were suggested to the trainees, alternative ways of phrasing questions were considered, etc. The relative importance of each area of interest was discussed, and suggested allocations of interview time were provided. Furthermore, the principal investigator, aware of the difficulty of keeping the interview "on-target", constructed a repertoire of conversation-directing probes and phrases.

Before conducting the first school interviews, each interviewer went through one additional hour-long simulation. A member of the study team played the role of school decision maker, mimicking the cooperative -- but often disorganized -- responses that had been encountered in the pilot interviews. During the interview, the surrogate decision maker made notes of the interview questions, of the technique, of successful and unsuccessful strategies, and content material which the interviewer had failed to obtain with his or her particular questioning. After the conclusion of the interview, the two discussed the experience in detail and the "decision maker" suggested areas for improvement.

The Field Interviews

Interview scheduling was completed by telephone two or three weeks before the field interviewing was to begin. A period of one hour was set aside for each interview, with 10 to 15 minutes between interviews for the researchers to make notes and refresh themselves. We also indicated that we desired a quiet place in which to conduct the interviews. While not all principals were able to indicate during this initial phone conversation which of their staff would be available at which hour, none had any difficulty assuring us that the arrangements would be made by the date of the interview. Almost without exception, the school personnel we dealt with were cooperative

and willing to go out of their way to fit into our schedules.

The interviews were conducted without major problems. The interviewers generally made sure, in advance, that an appropriate location would be available for each interview. We thought it important that the interviews not take place in a public place; not only could distractions interfere with the conversation, but respondents might find it difficult to answer candidly while their peers were within earshot.

Each of the interviews was taped recorded,and these tape recordings were overt. At the beginning of each conversation, the interviewer indicated that he was planning to make a taped record of the interview to ensure accuracy in the study and to facilitate future analysis. There were no objections to this, although a few of the respondents asked that the machines be turned off occasionally while they made certain comments. In each of these few instances, the respondent commented about another individual at the school, and they did not want the person's name recorded.[2] Aside from these instances in which personalities were involved, there were no irregularities or surprises in the interview process.

Data Analysis

The aggregation of field data is one of the most difficult tasks for those who conduct naturalistic research. Hours of interviews and pages of notes must be systematically summarized into a usable form. A balance must be struck between maintaining the richness of detail afforded by the naturalistic data and reducing data sets to a manageable and comparable form. A number of procedures have been tried by different researchers to accomplish this task. Alkin, Daillak and White (1979) presented a multi-stage data aggregation strategy in Using Evaluations. In the current study we

are guided by that approach, while at the same time we vary its specific procedures because of the nature of our data.

The data analysis proceeded through several phases, including: developing the initial data base, validating and expanding the data base, various data synthesis activities, and a final unrelated synthesis effort which we have called "quotation analysis." Each of the procedures will be discussed below.

Developing the Initial Data Base

To our view, the critical point in the analysis of qualitative data is developing the initial written data base. Knowledge and insights gained from previous research enabled us to focus the interviews on five specific topic areas. This simplified data aggregation by providing a logical framework for interview and validation summaries. The same road map that guided the interviews guided the summaries. Each interviewer used the interview tape and his or her notes of the conversation while recording comments within each of the initial topic areas of the summary form we developed. The second step was to listen to the tape again. The interviewers selected and recorded as many direct quotations as were necessary to capture the significant information and perspectives embodied in the comments of each respondent.[3] Thus, the final summary form contained five or six pages of narrative comment on the respondent's point of view (organized by topic area) and up to an additional five or six pages in relevant direct quotations. The summary form, along with a second summary form to be discussed shortly, became the initial data base for subsequent analysis.

Validating the Initial Data Base

A number of strategies were employed to ensure the accuracy of the interview summaries. In particular, we cite four project procedures which helped to assure data validity: use of tape recorders, use of independent validators, internal verification, and external verification.

We used tape recorders for validation in each and every interview, although arguments have been propounded againstthe use of tape recorders. (They're intrusive, artificial, a mechanical crutch, etc.) However, there are also strong arguments in their favor: (1) they free the interviewer to concentrate more on developing his or her next questions instead of recording the respondent's previous answer; (2) they allow the interviewer to focus his or her attention on the respondent and maintain direct eye contact while conducting the interview; (3) they allow the interviewer to go back to the interview at a later time and listen for things that might not have been readily apparent during the interview; and (4) they serve as a permanent record of the raw data of the study. This last, to our view, constitutes the most important reason for using tape recorders.

After an interview had been conducted at a school and the summary forms completed by the interviewer, the cassette tapes were turned over to a validator. Working from the tapes alone, this person completed a second independent set of summaries. Validators listened to the tapes (and completed their summary forms) in the order in which the interviews took place, playing each through completely before beginning to write and summarize. Thus, an hour was devoted to nothing but listening to the interview and getting an understanding of the situation. Following this, the validators began summarizing the interview according to the topics in their summary topic guides. (Validator's summary forms paralleled those used by the

interviewers.) A second listening of the tape generally produced the remainder of the information necessary for the summary forms. Validators were also asked to identify and transcribe key quotations from each respondent. Frequently this required listening to the tape a third time; occasionally only portions of the tape needed to be reviewed.

The two summaries together (interviewer's and validator's) provide the basis for within-project verification of the data base's accuracy. A step-by-step comparison was made of each pair of summary documents to determine the degree to which there were discrepancies. A high correspondence between the two summaries would allay fears that the data aggregation process might have introduced individual biases or discontinuities.

In fact, while some differences between the validator and interviewer summary sheets existed, the differences were (almost without exception) trivial. After the initial comparisons, there were only two or three instances in all 66 hours of tapes in which the interviewer and the validator reported information which was contradictory. Moreover, none of these discrepancies centered on a focal issue in the interview. Relistening to a portion of several tape recordings provided a simple but satisfactory resolution of differences. As a result, we are confident that our data aggregation process accurately portrayed the interview information.

The validation story does not end here. There is yet another level at which distortions might have been introduced into the data. It is possible that what was actually said in the course of the interview did not accurately reflect the respondent's point of view, perhaps because of the interview artificiality.

To ensure that our interviews actually depicted the understandings of

the interviewees, we employed a second validation process. A sample of respondents was asked to judge the accuracy of the summaries of their own interviews. Five different interviewers were used in the study and the second school visited by each of them was selected for field validation.[4] Thus, the three interviewees at each of these five schools constituted our field validation sample. Copies of the interviewer summary forms were mailed to these respondents. They were asked to review the summaries and note inaccuracies. We asked them, "Do these summaries accurately reflect the events you described?" This field validation process, therefore, gave us a measure of the sensitivity of our interviews. We already know that our summaries accurately reflected what the respondents said. Now we would learn if their words accurately portrayed what they meant about the occurrences.

An analysis of the respondents' comments revealed very few substantive differences with the summaries. In most cases, elaborations and explanations offered by the respondents represented relevant information that had not come out in the interview process. In sum, detailed analysis of the comments affirmed that our interpretations of the events and respondents' points of view were quite valid. This external validation procedure constituted yet another assurance of the accuracy of the qualitative data.

Data Synthesis

A number of steps were involved in synthesizing the data from the initial written data base. Group discussions, tallying descriptive data, and a procedure for eliciting appropriate field data from the initial data base are described below.

Group Discussions. The synthesis of data began with a series of open dis-

cussions among the research team members (the principal investigator, five interviewers, and five validators). The group met weekly following the completion of interviewing and validation to share initial impressions and experiences. Research team members were already beginning to notice patterns among responses in their data subsets which might hold across the complete sample.

After three group meetings, each team member prepared a draft report based on the interviews he/she conducted or validated, and as we hoped, a surprising number of different points of view emerged. To obtain an external critique of the themes which appeared to be emerging from the data, a conference telephone call was arranged with Michael Patton. Members of the research team discussed their initial thoughts with Patton, resulting in a good many insightful and illuminating comments.

Descriptive Data. Once the research team had identified appropriate categories and specified precise definitions, we compiled and summarized the descriptive data. For example, job titles of respondents, the significant occurrences described by the respondents, and several broader descriptive topics were categorized and coded.

The Human Data Bank. Coding categories for descriptive analysis was a relatively easy task, but the bulk of the analysis required a more complex strategy. Initially, we thought we would search the complete collection of interview and validator summaries for information pertaining to each topic. Unfortunately, this method was extremely slow. Case-by-case review consumed time, and resulted in apparently inefficient duplication of effort.

This is both a blessing and a curse of qualitative research: through repeated review and examination, subleties and nuances are uncovered which would be overlooked in most quantitative studies; on the other hand, it is not as convenient as simply turning quantitative data over to a computer

programmer and obtaining a quick statistical printout.

We developed a compromise technique that fell somewhere in between our initial procedure and the quantitative efficiency just alluded to. It became what one member of the team dubbed "the human data bank." (While the label makes the principal investigator cringe, he has yet to find an acceptable substitute.) Each of the participants, the five research team members and the principal investigator, was responsible for specific schools. Team members became informed stand-ins for the actual raw data for the next phase of analysis.

The analysis team met regularly throughout a three month period. Each person selected one of the themes which had emerged from our earlier work to pursue in greater detail. Each member of the human data bank would comment on the idea based on the information given by the respondents in his or her group of schools. A certain relationship may have been true in three of the four schools for which he or she was responsible, while special circumstances in the fourth might have upset the pattern.

When an idea appeared worthy of more complete investigation, the person leading that inquiry would draw up a questionnaire or organize a series of direct probes which could be put to the "human data bank." Each of the group members would carefully scan their respective summary sheets and prepare a response. They identified specific relevant examples and related direct quotations from the respondents. Most importantly, they also identified code numbers which could be used to locate the information in the summary sheets. The human data bank respondents served as a card catalogue: they offered ready reference to all the important data related to the topic.

The final phase in analyzing the data related to a particular topic was to access the data summary sheets directly. With precise directions and road maps in hand, it was much easier to find the relevant information in data summary sheets. Because of the participation of the other members of the group, not only were the papers that evolved at this stage stronger and more thoroughly thought out, but the process of checking them directly against the raw data was significantly simplified.

Quotation Analysis

One of the major strengths of our qualitative analysis was the multiple validation and quality assurance procedure. Another unique methodological procedure was the independent analysis of the initial data set from multiple perspectives.

Independent from the group meetings, descriptive data aggregation and individual paper writing involving "human data banking," the principal investigator assembled a full set of quotations from the project's initial data base. A careful reading and analysis of these quotations by a staff member not associated with the other activities provided a separate and independent categorization of the data and a data synthesis. Themes developed in the currently reported phase of the study will be cross validated through the final phase of the research. In addition, new ideas and relationships have emerged that will be explored in the months ahead.

Conclusions/Summary

Three important issues had to be addressed in this methodological research. First, we had to select an appropriate research strategy to answer the questions that guided our study. The topic centered interview was

developed to afford the balance between structure and flexibility while maintaining the proper sensitivity to detail required by our subject. Second, careful validation procedures were employed to insure that the interviews accurately represented the events recalled by the respondents and that the written summaries accurately represented the tape recorded conversations. Finally, multiple strategies for data analysis were undertaken, including directed discussions, selected reviews of written data summaries, human data banking and an independent "quotation analysis". This many-faceted approach gave added validity to the themes and patterns of interactions uncovered in the study.

NOTES

[1]One of the schools was one in which Daillak had had substantial observation time in the company of the Metro evaluator, and in which there seemed to be a close relationship between evaluator and site personnel. In contrast, Daillak had secured less information about the other school, and it was selected partially for that reason.

[2]We had indicated to each respondent that all data would be recorded anonymously at the beginning of the interview. But such assurances are not always remembered...or believed.

[3]Though the process was not overly complex, it was nonetheless quite time consuming. The summary form for each one hour interview took 2½ to 3 hours to complete.

[4]We felt that the first set of interviews would not be truly representative. (Nor would the last one.)

REFERENCES

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

Daillak, R. A Field Study of Evaluators at Work. CSE Report No. 154. Los Angeles: University of California, Los Angeles, December 1980.

Patton, M.Q. Qualitative Evaluation Methods. Beverly Hills, California: Sage Publications, Inc., 1980.

Patton, M.Q. Utilization-Focused Evaluation. Beverly Hills, Califorina: Sage Publications, Inc., 1978.

Weiss, C. H. Knowledge Creep and Decision Accretion. Knowledge: Creation, Diffusion, Utilization, v. 1, no. 3, 1980.