CSE
MONOGRAPH
SERIES
IN
EVALUATION

4

EVALUATION AND DECISION MAKING: THE TITLE VII EXPERIENCE

CENTER FOR THE STUDY OF EVALUATION
UNIVERSITY OF CALIFORNIA • LOS ANGELES

EVALUATION AND DECISION MAKING: THE TITLE VII EXPERIENCE

CSE MONOGRAPH SERIES IN EVALUATION

Editorial Reviewer for Number 4
Jason A. Millman

SERIES EDITOR
Marvin C. Alkin

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles
Los Angeles, California 90024

EVALUATION AND DECISION MAKING: THE TITLE VII EXPERIENCE

Marvin C. Alkin, Jacqueline Kosecoff, Carol Fitz-Gibbon, and Richard Seligman

Center for the Study of Evaluation
University of California, Los Angeles, 1974

CSE MONOGRAPH SERIES IN EVALUATION

NUMBER

1. Domain-Referenced Curriculum Evaluation: A Technical Handbook and a Case Study from the MINNEMAST Project

   Wells Hively, Graham Maxwell, George Rabehl, Donald Sension, and Stephen Lundin $3.50

2. National Priorities for Elementary Education

   Ralph Hoepfner, Paul A. Bradley, and William J. Doherty $3.50

3. Problems in Criterion-Referenced Measurement

   Chester W. Harris, Marvin C. Alkin, and W. James Popham (Editors) $3.50

4. Evaluation and Decision Making: The Title VII Experience

   Marvin C. Alkin, Jacqueline Kosecoff, Carol Fitz-Gibbon, and Richard Seligman $3.50

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF CHARTS

FOREWORD

The provision of information to decision makers has been a central concept in evaluation theory. This monograph, a product of the Evaluation Theory Program at the Center for the Study of Evaluation, explores the extent to which evaluations of bilingual education projects (Title VII) appear to have an impact on decision making at the local level, representing project management, and at the federal level, representing project sponsors. We were interested not only in the impact evaluations have had in the past, but also in the future use of evaluation. We considered what modifications might be required in the way evaluations are conceptualized and mandated, in order to increase the impact of evaluation and thus render the evaluative process more useful. We did not attempt to evaluate the Title VII bilingual education program or to assess the effectiveness of the local projects.

The monograph has been structured to accommodate the interests of various readers, some of whom may not wish to peruse the entire document. In the first chapter the planning and implementation of the study are described. From a nation-wide sample of 42 projects we examined evaluation and audit reports, and collected information directly from project directors. The summary statistics from these data bases are presented in the second chapter. These two chapters comprise Part I of the monograph, which provides the reader with an understanding of the basis upon which rest the two subsequent parts of the monograph.

In Part II we describe our attempts to explore quantitatively the relationships that might exist between the characteristics of evaluations and their impact on the decision making of both managers and sponsors. The evaluation characteristics were grouped into such factors as "evaluation comprehensiveness" and "evaluation context" (Chapter III) and these factors were then employed in multiple regression analyses in which indicators of federal and local decision making were the dependent variables (Chapter IV). Some interesting relationships, and even more interesting lack of relationships, are discussed in these chapters.

Part III presented a more qualitative, conceptual analysis and synthesis of the data base. Chapters V and VI, which comprise Part III, will be of particular interest to those concerned with evaluation theory. The recommendations (Chapter VI) are, we believe, of great importance for those principally responsible for the planning and/or mandating of evaluations.

MCA

Chapter I
CONTEXT AND METHODOLOGY OF THE TITLE VII STUDY

Context of the Study

The Need for the Study

At the Center for the Study of Evaluation (CSE) we have long considered the major function of evaluation to be the provision of timely information to decision makers. Underlying this point of view is the presumption that the information provided must also be useful. If evaluative information is not useful, if it does not serve the needs of decision makers, then evaluation has lost its justification.

However, when we considered the question of whether the information provided by evaluations was actually used by decision makers, it became apparent that many subsidiary questions had also to be addressed. Among these questions were: Which kinds of evaluative information are useful and which are not? Which kinds of information could be useful but are nevertheless rejected? Which decision makers actually use evaluative findings to guide decision or policy making and which ignore them or use them only as a matter of form? Under what conditions is evaluation useful and under what conditions is it not useful?

In order to study some aspects of this complex problem of the utility of evaluation for decision makers, we focused on the impact of the evaluations of federally funded educational programs on decision makers at two levels: (a) the local level (the decision maker at this level being the project director), and (b) the federal level (the decision makers at this level being the federal program monitors and those individuals who make refunding decisions).

The Choice of Title VII

To be able to study the impact of evaluation on decision making we needed a large sample of projects with evaluation reports available. It was desirable that there be some homogeneity in the projects under study. An obvious source of such projects was the Elementary and Secondary Education Act of 1965 (ESEA). A large number of federally funded educational programs are located under the various titles of this act and are generally administered by the state departments of education subject to federal approval. However, the 1967 amendment to this act established two additional programs, Title VII (Bilingual Education) and Title VIII (Dropout Prevention) that were to be operated under direct grants from the Federal Government to local school districts. The absence of an intervening state level of administration seemed likely to facilitate the selection

of a sample, and the increasing emphasis on evaluation in these programs lent them an additionally promising aspect. Evaluation of each local project was mandatory, and the guidelines were considerably more specific and comprehensive than those in earlier programs.[1]

Further, for the first time under ESEA legislation, an independent educational accomplishment audit was a major requirement for accountability. The *Title VII Manual for Grantees and Applicants* (HEW, 1971) establishes that:

> All projects must provide for an independent educational accomplishment audit of the project, to apprise school officials of the validity of their own evaluative processes and data (p. 11).

It was anticipated that the independent educational accomplishment audit would provide us with another source of perceptions of the evaluation and of the program, and that the audit itself might have an impact on decision making.

The study team was faced with a difficult choice in regard to selection of the program for investigation. On the one hand, the selection of multiple programs for examination (e.g., Title VII *and* Title VIII) might have increased the generalizability of the findings about evaluation and decision making. Selecting multiple programs, however, would have introduced an additional factor into the data. Similarly, since audits have not generally been usual, we might detract from the generalizability of the findings (because what we would have is the effect of evaluations *when outside audits are present*). On the other hand, the presence of the audits might provide an additional vital source of data about the evaluations. These areas of trade-off were noted, and the decision was made to select *either* the Title VII or Title VIII program for the study.

Title VII (Bilingual Education) was chosen in preference to Title VIII (Dropout Prevention) largely because of the greater number of local projects in the Title VII program and because of the greater variety of these projects in terms of size, geographical location, and funding levels. Moreover, the fact that Title VII projects were more likely to be conducted in traditional classroom settings suggested that the operation and setting of these projects would be similar to the kind of situation in which educational evaluations are usually conducted. The choice of Title VII turned out to be a particularly happy one. The cooperation and enthusiasm of those administratively responsible for the direction of the Title VII program were most welcome.[2]

---

[1]Throughout this monograph we will refer to and make distinctions between federal programs, local projects, and instructional programs. In this context, specifically, we will refer to the *Title VII program*, the *local project* that is funded under Title VII, and the particular *instructional program* implemented by the local project.

[2]We particularly appreciate the cooperation of Dr. Albar Peña, Chief, Bilingual Education Branch, U.S. Office of Education, and Miss Elizabeth Keesee, the Assistant Branch Chief, who served in these positions when this study was conducted.

The Title VII Program of ESEA

The Bilingual Education Act made funds available to local school districts for "exemplary pilot or demonstration projects in bilingual and bicultural education" (HEW, 1971, p. 5). The Title VII projects were designed to meet the "special educational needs of children who have limited English-speaking ability, who come from environments where the dominant language is one other than English, and who come from low income families" (HEW, 1971, p. 1). The program had not only the cognitive goals of increasing students' competence in both their dominant language and English, but also the affective goals of building self-esteem and pride in both cultures.

School districts that submitted successful proposals for Title VII projects were approved for five years of operation. (Funding for the projects was made on an annual basis.) As part of the proposal the local district was required to describe its project in terms of four component areas: instructional program; acquisition, adaptation, and development of materials; staff development; and community involvement. The descriptions included overall project objectives as well as objectives for each functional component for the five-year period and for each year of the project's operation.

The first major component and the prime focus of each Title VII project is the instructional component. One purpose of a Title VII project is to improve students' performance in both English and a second language by using both languages during classroom instruction. There are three approaches on which the instructional program may be based. The first alternative is that instruction be given in approximately equal amounts in both English and the dominant language of the target population. A second alternative is that instruction be given entirely in the dominant language of the children in the target population. (This alternative has been employed in those cases where the target population comprises children who are monolingual speakers of languages other than English.) The third alternative for the instructional component is the use of English as a Second Language (ESL). Regardless of the particular approach taken, however, the basic goal is to develop the knowledge and use of two languages.

The second major component of a Title VII project involves the acquisition, adaptation, and development of instructional materials. At the time the Title VII program first began it was found that there were virtually no instructional materials appropriate to bilingual education. In the case of Spanish bilingual instructional programs, for instance, the only available materials were those produced in Spanish-speaking countries such as Mexico, Spain, and Cuba; teachers soon found that these materials were not appropriate for the students in their classes. Consequently, it was necessary either to adapt instructional materials already developed in English to the language of instruction in a particular bilingual instructional program

or to begin developing new materials. As Title VII projects developed, the emphasis on curriculum development was reduced, since it was found that this task was more appropriately conducted by commercial publishers, regional educational laboratories, specially funded curriculum development projects, and other such agencies.

The third major component of Title VII projects is staff development. At the outset of Title VII it was rare to find teachers who were adequately trained in methods of bilingual instruction. Part of each Title VII project accordingly, was devoted to inservice training for instructional staff. This continues to be a mjaor emphasis in the Title VII program.

The fourth major component of Title VII projects is concerned with involving community members in the schools. In those schools involved in Title VII projects, community advisory committees were formed and efforts were made to enlist the involvement of members of the community in the planning and operation of the bilingual instructional programs.

From the very outset Title VII projects were required to have evaluations as well as independent audits. In the case of the evaluations, the local school district could either assign an internal evaluator or contract with an external agency for evaluation services. The USOE guidelines for Title VII required that behavioral objectives be stated for each of the major components of the Title VII project. The objectives were to be stated in terms of "product" or "process." The product objectives typically referred to the behaviors that students (or teachers or parents or administrators, depending on the particular component of the project) would exhibit upon completion of the year's activities. The process objectives, on the other hand, were to be concerned with the conditions that must exist in order for the product objectives to be attained. Specific requirements and formats were developed for listing all project objectives, as well as for indicating the manner in which the attainment of each objective was to be measured.

Methodology of the Study

Selection of the Sample—Federal Ratings

Initial approval for the CSE study of Title VII evaluations was obtained from the Associate Commissioner, Bureau of Elementary and Secondary Education. The Commissioner referred us to the Director of the Division of Bilingual Education who arranged for a senior staff member from CSE to meet with monitors of the Division's Bilingual Education projects. At this meeting the nature and purpose of the study was described and the CSE staff member provided forms on which each monitor was asked to designate nine projects from those under his supervision. These nine projects were to be projects in operation during the 1970–71 school year and were to consist of three projects that the monitor considered "outstand-

ing," three that he considered "average," and three "below average." No guidelines were provided for making these distinctions since we were interested in the monitors' subjective perceptions of projects and the possible relationships between these and the evaluation data and audit data.

It should be noted that this procedure provided an initial sample with one dependent variable (federal monitor's perception of the project's overall quality) already "built in." Since there were seven federal monitors supervising distinct groups of the 129 bilingual projects in operation, it was expected that having each monitor select nine projects would yield a sample of 63 projects. This expectation was not realized. Several of the designated projects had not actually been operational during 1970–71, the year selected for study. In other projects reports were unobtainable, in spite of a search of the Washington files for evaluation and audit reports. In cases where the evaluation report was present and only the audit report unobtainable, the project was retained in the sample; if a final evaluation report for the 1970–71 year was unobtainable, the project was dropped. The final sample consisted of 42 projects, 39 of which had audit reports in addition to the evaluation report. More projects had been lost from the "average" category than from the "outstanding" and "below average" categories, yielding a final sample of 42 projects, 15 of which were rated "outstanding," 12 "average," and 15 "below average." These 42 projects were from locations in 17 different states. Two states had a concentration of projects: California (17 projects) and Texas (six projects). These concentrations reflect the bilingual populations in the two states and the fact that the most usual non-English language in the bilingual instructional programs was Spanish.

Each project was assigned an identification number used in all subsequent analyses; no individual project is identified in this monograph.

Translating Materials into Data

In order to compile and summarize the information contained in the project reports, two instruments were constructed: an Evaluation Data Sheet and an Audit Data Sheet. These forms are described in the following section and are reproduced in Appendix 1. Three research assistants acted as raters for each project. As a project's evaluation report was read the rater completed an Evaluation Data Sheet. The rater then read the audit report for the same project, completing the Audit Data Sheet while doing so. During the reading of the reports, raters were not aware of the federal monitor's assessment of the project or of the project's funding or refunding level. Such information had been collected independently and kept secure. For each project, two raters independently completed the Evaluation and Audit Data Sheets. Their responses were then compared and points of disagreement marked. Using this information and turning back to the project's evaluation and audit reports as necessary, the third rater then

acted as adjudicator. Intractable difficulties relating to data interpretation were referred to the Principal Investigator for resolution. The final adjudicated ratings from the Evaluation and Audit Data Sheets were used in subsequent data analysis. The data from these instruments provided most of the more than 250 independent variables.

Before describing the content and rationale of the Evaluation and Audit Data Sheets, one further source of data derived from the projects' evaluation and audit reports should be mentioned. Each rater was asked to keep a notebook while reading these reports. These notebooks later became a valuable source of representative quotations and anecdotal data.

The Evaluation Data Sheet

The Evaluation Data Sheet consisted of items grouped under five headings: (I) Characteristics of the Evaluator(s); (II) Evaluation Activities; (III) Content and Format of the Evaluation Report; (IV) Results of Evaluation; and (V) Rater's Reaction to the Evaluation.

The first part of the Evaluation Data Sheet dealt with evaluation personnel. Was a single person acting as evaluator or was evaluation a team effort? Were evaluators regular employees of the school district being funded (i.e., internal evaluators) or were their services contracted for in order to perform only the specific evaluation of the current project (i.e., external evaluators)? Another kind of item under this heading dealt with specifying the qualifications of the principal evaluator.

The second part of the Evaluation Data Sheet was concerned with evaluation activities. This was divided into two sections: en route (formative) and summative activities. In keeping with the CSE evaluation model, the questions regarding en route evaluation activities directed the rater to note the presence of implementation and progress evaluation in the evaluation reports. Methods of sampling and the kinds of measurements used were to be noted. Under summative evaluation activities the components of experimental designs used were to be noted as well as the extent to which the design was implemented. The rater was to indicate whether the evaluation report did or did not make reference to a number of areas, such as community/parent relations, cost information, staff training, etc. The measurement methods and statistical procedures employed were also to be noted in this section.

Having categorized who the evaluator was and what he did, the third part of the Evaluation Data Sheet dealt with the nature of the evaluation report he produced. How were data presented? Were all objectives assessed? The CSE raters were required to judge the completeness of the interpretations and descriptions in the reports. It was also expected that the raters would be able to ascertain some basic facts about the Title VII projects from the evaluation report, such as the nature of the target community, the schools involved in the projects, and the ethnic groups

and grade levels of pupils. Questions related to this information were included in the third section.

The fourth part of the Evaluation Data Sheet required the raters to judge whether the evaluator considered the project to have been "successful," "neutral," or "unsuccessful" in such areas as cognitive and affective outcomes and staff and parent/community relations. These items represented CSE raters' interpretation of the evaluator's judgments of the various aspects of the project.

The fifth part of the Evaluation Data Sheet asked for more subjective judgments from the rater: From the evaluation report, what impression did the rater receive of the quality of the project, of the evaluator's attitude toward the project, and of the usefulness of the evaluation report for analyzing project effectiveness?

The Audit Data Sheet

The Audit Data Sheet consisted of items grouped under the following five headings: (I) Personal Characteristics of the Auditor; (II) Methods Used by Auditor; (III) Perceived Quality of Evaluation Procedures; (IV) Scope of Audit Report; and (V) Rater's Reaction to Audit Report.

The first part of the Audit Data Sheet was concerned with the auditor. The three items in this part concerned the auditor's professional affiliation, highest degree, and major field of study.

The second part of the Audit Data Sheet dealt with the auditor's methods. Items in this second part were intended to ascertain whether an evaluation model had been followed and required the raters to report whether various data sources (interviews, tests, questionnaires, etc., administered to various possible groups such as students, parents, staff) were or were not used by the auditor.

The third part of the Audit Data Sheet was concerned with the quality of evaluation procedures, and required the raters to judge the auditor's opinion of the quality of about 30 evaluation activities, such as the general design utilized, identification of goals, accuracy of data, format of the evaluation report, and formative activities.

The fourth part of the Audit Data Sheet was concerned with the scope of the audit report. Items in this part concerned areas of agreement between auditor and evaluator and the presence of content that might be expected to occur in audit reports, such as consideration of alternative strategies for the evaluation, discussion of the evaluator's recommendations, comments on the relationship between the evaluator and project director, cost information, and recommendations for future modifications. The issue of recommendations was probed in some detail, and raters were required to indicate whether or not recommendations were made by the auditor in some 16 areas, such as formative evaluation activities, project management, and data collection techniques.

The fifth part of the Audit Data Sheet asked for the raters' perceptions of the auditor's statements and the raters' reactions to the audit report. How did the auditor appear to judge the project? How did he judge the evaluation? How useful was the audit report for analyzing evaluation effectiveness?

It should be clear from the above descriptions of the data sheets that not only was a considerable amount of information codified by the use of data sheets but also several items were necessarily judgmental. The use of two raters and an adjudicator for each data sheet was some protection against arbitrary and unrepresentative judgments, but the subjective nature of some items must nevertheless be acknowledged.

To standardize the rating procedure as much as possible the following preparatory activities were undertaken. Raters met to agree upon standards for response categories in each item of the Evaluation and Audit Data Sheets. This activity included reviewing the items in terms of the feasibility of categories and then rating evaluation and audit reports (ones not included in the sample) to test the consistency of our raters.

The Project Director Questionnaire

Subsequent to an initial analysis of the data from the evaluation and audit reports (described below under "Data Analysis"), a Project Director Questionnaire (reproduced in Appendix 1) was constructed and mailed to the 42 persons who had been the project directors in the year 1970–71. This questionnaire had several purposes. Information that had proven to be unobtainable from evaluation and audit reports was requested (in particular, the professional qualifications and affiliations of the evaluator were frequently missing in evaluation reports), and verification was sought for some items that had been uncertain (e.g., internal or external evaluator status). Perhaps most important, however, the questionnaire assessed the impact of the evaluation upon the project director, one of the key local decision makers. This impact was considered as a dependent variable.

The cover letter mailed with the questionnaire gave project directors the option of replying by mail or by telephone interview and all of our follow-up efforts were conducted by telephone. Despite the fact that many persons who had been project directors in 1970–71 were now in other positions and, indeed, were located in many different parts of the country, persistent attempts to reach these persons by telephone finally produced a 100% response rate from the 42 project directors in our sample.

Data Analysis

The dependent variables were measures of the impact of evaluation upon decicion making at the two levels previously mentioned, federal and local. With regard to federal decision making the raw data consisted

of funding levels for fiscal years 1970, 1971, and 1972 and the rating of projects by federal monitors. At the local district level raw data dependent measures consisted of responses from project directors to the questionnaire items that probed the perceived usefulness of evaluation.

The independent variables were drawn from over 250 items from the Evaluation and Audit Data Sheets described above. As such, they represented characteristics of the evaluator, the evaluation, methodology and findings, and other such variables considered as possible correlates of decision-making utility.

Summary descriptive statistics based on the data sheets are reported in Chapter II as well as a discussion of various highlights. The complete listing of summary statistics as derived from an available computer program (Statistical Package for the Social Sciences) is provided directly on the reproduced copies of the data-collection instruments found in Appendix 1.

A number of "new variables" were derived from existing items by summing variables that could be appropriately grouped; for example, the presence or absence of information about various aspects of the context of the projects. Intercorrelations among all variables were computed. Before extending correlational analysis to examine regression equations the raw data were transformed in two ways. First, in order to make the correlations more comparable and less subject to the effects of skewed distributions, many variables were dichotomized close to the median. Second, indices were developed to represent several descriptive features of evaluation reports, such as comprehensiveness and sophistication. Each index was a linear combination of several correlated variables representative of the evaluation characteristic. This procedure was a substitute for factor analysis, a procedure that would have been inappropriate with $N=42$ and number of variables over 250. Indices were also developed in a similar manner from the dependent variables representing decision-making utility. In addition, a "Most Representative Variable" (MRV) was selected from each cluster of variables that had comprised the index. As a result of these activities, therefore, the data were reduced to seven indices characterizing various aspects of evaluation that were to be used as independent variables. Moreover, the dependent variables were now two indices, one reflecting federal decision making and one local decision making. For all indices, both dependent and independent, there was a single variable designated as "Most Representative." The indices and the MRV's will be presented in subsequent chapters.

Employing this condensed data base, regression equations were then computed to examine the relationships between decision-making utility and evaluation characteristics in the sample under study. Although we would have preferred to divide the sample in two in order to run a cross validation of the regression equations, this procedure was not deemed

appropriate in view of the sample size. We realized that these regression equations capitalize on chance variations in the data, but they nevertheless represent the strength of association between dependent and independent variables located by the procedures described in this chapter. Stepwise regressions were used in each case in an attempt to predict decisions (e.g., funding level) and perceptions of the utility of evaluations.

Summary

The six chapters of the monograph are set forth as follows. Chapter I has described the context and methodology of CSE's Title VII study. As previously mentioned, summary descriptive statistics appear in Chapter II. The development of indices for the independent variables is described in Chapter III. In Chapter IV the indices representing the dependent variables of decision-making utility are described and the results of the regression analyses are reported. Chapter V discusses the utility of evaluation information and those elements of the decision context to be considered in conducting the evaluation. In Chapter VI recommendations are made that are intended to clarify the distinctions between formative and summative evaluation.

Chapter II

DESCRIPTIVE DATA

This chapter presents descriptive data on the Title VII evaluation and audit reports as well as information on the evaluators and auditors themselves. Some indication of the impact of the evaluation data on local project directors is also included. These descriptions are the results of the analyses of the 42 evaluation reports, the 39 audit reports, and questionnaires completed by the 42 project directors. These documents were detailed in Chapter I.

Evaluation Report Data

Information from the evaluation reports was coded by the CSE raters onto the Evaluation Data Sheet (Appendix 1). The sections following discuss the characteristics of the evaluator, formative evaluation activities, summative evaluation activities, and the final evaluation report.

Evaluator Characteristics

The first part of the Evaluation Data Sheet attempted to answer the question: Who conducted the evaluation? This question, unfortunately, could rarely be answered from a reading of the evaluation report. Even when CSE raters were permitted to go back to original project proposals, information about evaluator characteristics was incomplete for many projects. The Project Director Questionnaire, described in a later section of this chapter, filled in the gaps somewhat; that section provides information on the characteristics of the evaluator, whether he was external or internal, his professional affiliation, degrees held, etc.

Formative Evaluation Activities

Formative, or en route, evaluation is concerned with monitoring and improving an instructional program while it is being carried out. In this regard, formative evaluation may be distinguished from summative evaluation which is concerned with the status of an instructional program as a completed entity. In conducting formative evaluation activities, the evaluator must consider two questions:

Has the program been implemented in accordance with the program plan? (implementation evaluation)

Is the program meeting its stated objectives? (progress evaluation)

In the sample of 42 project evaluation reports used in this study, evidence of evaluation procedures designed to verify program implementation was noted in 25 of the cases (60%). Table 2.1 shows the procedures that were employed in the implementation evaluations. The most common techniques noted by the CSE raters were program "observation" and "interviews," both of which seem to be appropriate methods of determining the extent to which an instructional program has been implemented as planned. Many of the evaluation reports are subject to criticism, however, for their failure to describe how the observing and interviewing were actually conducted.

TABLE 2.1

Implementation Evaluation Procedures Employed by Evaluators (N = 25)*

Procedure	N	%
Program Observation	24	96
Interviews	12	48
Miscellaneous School Records	8	32
Checklists	7	28
Review of Staff Reports	7	28
Questionnaires	3	12

*Only 25 of the 42 Evaluation Data Sheets indicated the use of implementation evaluation.

Procedures for conducting progress evaluations were noted in 22 of the 42 evaluations (52%). Table 2.2 shows the procedures that were employed in determining, during the course of the school year, the extent to which the instructional program was achieving its objectives. These data were obtained in response to the item "Indicate if the procedure was used to monitor program progress . . ." Once again, "program observation" was the evaluation procedure noted with the greatest frequency. However, in contrast to the report on implementation evaluation, both checklists and staff reports were frequently used by the evaluators in monitoring the progress of the programs toward their stated objectives. It is of some interest to note that testing, with either locally-developed or standardized instruments, seemed to be employed in relatively few cases as a measure of progress.

We should note here that the distinction between an activity conducted for formative, as opposed to summative, purposes is not immediately obvious. For the purposes of our classification the following convention was observed by the raters. An activity was classified as formative only if there were evidence in the evaluation report that the resulting information was put to use by local decision makers during the year. Evidence of

this use was derived from statements in the evaluation reports such as "The high mean scores on pretest indicate that students had already mastered much of the material and we therefore decided to add more difficult objectives to the curriculum."

Data from a reading of the final reports suggest that some formative evaluation was employed in slightly more than 50% of the cases. As we have noted, evaluators were somewhat vague when it came to specifying the actual procedures that were employed in formative evaluation. Of even more concern to the present investigation, the information contained in a great many of the evaluation reports does not enable one to know how the data collected for formative evaluation purposes were *used to improve the program.*

TABLE 2.2

Progress Evaluation Procedures Employed by Evaluators (N = 22)*

Procedure	N	%
Program Observation	16	73
Checklists	16	73
Staff Reports (written or oral)	14	64
Documentation	8	36
Locally developed Tests	6	27
Nationally normed Tests	3	14
Interviews	3	14

*Only 22 of the 42 Evaluation Data Sheets indicated the use of progress evaluation.

Summative Evaluation Activities

As one might suspect, an activity called "summative evaluation" occurred in *all* of the evaluations in the sample. These summative evaluations attempted to make judgments about the successes and failures of the instructional program at the conclusion of the school year.[1] In order to consider the evaluations from the point of view of the "experimental" designs they employed, the final evaluation reports were examined with particular attention to the major design employed for the assessment of cognitive achievement. The results of this approach are reported here in terms of the widely recognized classification of experimental designs provided by Campbell and Stanley (1966).

Campbell and Stanley classify designs as either pre-experimental, quasi-experimental, or true experimental. As shown in Table 2.3, pre-experimen-

[1]We would maintain that since the program had not fully stabilized in most cases true "summative" evaluations did not take place.

tal designs were the most frequently employed in the 42 evaluations studied. Twenty-seven evaluations (64%) used either a one-shot case study or one group pretest-posttest design in the area of cognitive achievement. These pre-experimental designs enable descriptions to be made; they do not enable inferences to be drawn about the effects of the special treatment (the bilingual program) as opposed to the effects of the normal maturation process or the normal impact of schooling.

TABLE 2.3

Designs for the Assessment of Cognitive Achievement In 42 Title VII Evaluations

Design		N
Pre-experimental		(27)
One-shot case study	$X\ O_1$	8
One group pretest-posttest	$O_1\ X\ O_2$	19
Quasi-experimental		(15)
Non-equivalent control group	$O_1\ X\ O_2$	15
	$O_3\quad O_4$	

Note: Subscripts do not necessarily connote different tests. Thus, in the non-equivalent control group design, O_1 and O_3 would represent observations on the experimental and control group on the *same* test.

The remaining 15 evaluations (36%) attempted to employ a quasi-experimental design, the non-equivalent control group, which Campbell and Stanley emphasized as "well worth using" when true experimental designs are impossible (1966, p. 47). This design permits inferences to be made about the effect of the special program with far fewer caveats and competing hypotheses than any of the pre-experimental designs. The strength of these inferences about the effect of the program depends largely upon how comparable the control and experimental groups are. In five of the 15 cases attempting to employ this design, the "non-equivalent control group" was judged by the CSE raters to be so non-equivalent (on pretests) that inferences would not be justified. Thus only 10 of the quasi-experimental designs were successfully implemented.

Table 2.4 shows the project components that were incorporated in the summative evaluation designs. As the data in the table indicate, the major components of a Title VII project as specified in USOE guidelines—instruction, materials development, staff development, and community/parent involvement—were addressed by the evaluations. Consistent with the stance taken by USOE program officials, the greatest emphasis was given to student cognitive achievement. It is interesting to note that

only in a few cases did the evaluators go beyond stated objectives to consider unanticipated outcomes in either the cognitive or affective areas. One of the leading evaluation theoreticians (Scriven, 1972) has argued that evaluators must look beyond program goals to consider the unanticipated consequences of the program. This requires considerable skill on the part of the evaluator, and the added effort required to obtain data on unintended consequences may account for the rarity of such data. It is possible that a positive relationship might exist between the inclusion of data on unanticipated consequences and the overall quality of the evaluation.

TABLE 2.4

Project Components Included in Final Evaluation Reports (N = 42)

Component	N	%
COGNITIVE AND AFFECTIVE STUDENT OUTCOMES		
Cognitive achievement: pre-specified goals	42	100
Affective/attitudinal changes: pre-specified goals	26	62
Cognitive achievement: unanticipated outcomes	8	19
Affective/attitudinal changes: unanticipated outcomes	4	10
COMMUNITY INVOLVEMENT		
Community/parent relations: pre-specified goals	39	93
Community/parent relations: unanticipated outcomes	3	7
STAFF DEVELOPMENT		
Staff training	36	86
Staff performance and attitudes	30	71
Staff relations	14	33
MATERIALS DEVELOPMENT (PROCESS OBJECTIVES)		
Process objectives	31	74
COSTS		
Cost information	10	24

The evaluation techniques that were employed in gathering data for summative evaluation purposes are presented in Table 2.5. The most common technique employed was the administration of nationally-normed tests. The second most common technique, termed "documentation," involved the evaluator's examining project records. These documents, maintained by teachers and other project staff, took a wide variety of forms: records of students' progress toward project goals for the instructional component, records of inservice activities for the staff development component, records of materials acquired, and records of community/parent involvement, usually in the form of minutes of meetings. The remaining techniques identified in Table 2.5 appear to be those that one would

normally associate with a summative evaluation—administration of questionnaires, locally developed tests, checklists, reports of site visits, reviews of project reports, and evaluator-conducted interviews. It is interesting to note that Table 2.5 suggests that standardized tests are a far more common evaluation device than locally developed tests for summative evaluation.

TABLE 2.5

Evaluation Techniques Employed in Gathering Data for Summative Evaluation (N = 42)

Technique	N	%
Nationally-normed Tests	40	95
Documentation	35	83
Questionnaires	34	81
Site Visits	31	74
Locally developed Tests	31	74
Checklists	26	62
Reports from Project Staff	23	55
Archival Data	19	45
Interviews	18	43

Having considered the data-gathering techniques that the evaluators employed in their summative evaluation activities, one might well look at the techniques that were used to analyze these data. The most common form of analysis was the use of simple descriptive statistical procedures such as measures of central tendency and percentages of students achieving one or more of the project's objectives. A second common form of data analysis was one that we designated as "documentation without statistical referents." In this non-quantitative approach reliance was placed upon the interpretation of informally examined data (best known as eye-balling). Although in many cases raw data were made available, no attempt was made to systematically organize and/or explore trends in the data. Without the aid of any statistics the evaluator provided a commentary on the meaning of his data. A third common data analysis technique used hypothesis-related statistical procedures such as t-tests and Chi-square tests. Most often these techniques were employed either in making comparisons between pretest and posttest scores or between "experimental" and "control" groups.

The Evaluation Report

The culmination of the evaluator's activities is the final evaluation report; this report is the product or outcome of the evaluation effort and,

as such, is of considerable importance to this study. No matter how high the quality of an actual evaluation, it will be of little worth if the results are not well presented. With this in mind, the raters began with a consideration of the language used in the reports. Of the 42 evaluation reports examined, 11 (26%) were judged to have been written in highly technical language. That is, in order to read and understand these reports, one would need to have considerable sophistication in the interpretation of research results. Seventeen of the reports (40%) were judged to have been "somewhat technical." That is, some sections of the report could be understood only by the technically trained reader, but most sections of the report could be understood by an individual who had no special background in research and evaluation. The remaining 14 reports (33%) were judged to have been written in non-technical language and, thus, could be understood by the average reader.

The raters found that 25 (60%) of the evaluation reports included data on all of the stated objectives of the specific project being evaluated. This means that 17 of the reports did not report data on some of the project objectives. These omissions were explained by the evaluator in only five cases. There are a variety of explanations that might be advanced for these omissions. At one extreme, one might hypothesize that evaluators tended to leave out those results that would indicate lack of success on the part of the project being evaluated. (If true, this would be a most serious situation, questioning the professional ethics of the evaluator and thus the credibility of the entire evaluation effort.) Alternatively, the omissions might be explained by the fact that the original set of project objectives was reduced in number or that for some legitimate reason data on some objectives could not be obtained (perhaps the tests did not arrive on time). Speculations on this question are, however, just that—speculations. It would have been far more appropriate for the evaluators to have carefully documented the reasons for not including data on the complete set of the project's objectives as derived from the project proposal.

One of the evaluator's most important functions in the final evaluation report as it is presently conceived and required by Title VII is to indicate the project's successes and failures. Table 2.6 presents the evaluators' judgments on the success of each of the components of the projects they evaluated. An examination of this table leads to several important observations. Quite clearly, the data indicate no failures. Evaluators have carefully refrained from specifically indicating that the goals of a particular component have not been attained. Only in the rarest of cases did the evaluator include data that demonstrated the failure of a particular project component. While it is important that evaluators do not seek, for some unwarranted purpose, to highlight project failures, if they are to provide decision makers with useful information they must render a full and complete accounting of what took place over the course of the year.

TABLE 2.6

Evaluators' Judgments on the Success of Each Project Component (N = 42)

Component	Evaluators' Judgment			
	Successful	Neutral	Unsuccessful	Other*
COGNITIVE AND AFFECTIVE STUDENT OUTCOMES				
Cognitive outcomes: anticipated	32	9	1	0
Cognitive outcomes: unanticipated	2	3	1	36
Affective outcomes: anticipated	18	2	1	21
Affective outcomes: unanticipated	2	1	0	39
COMMUNITY INVOLVEMENT				
Community/parent relations: anticipated	34	2	3	3
Community/parent relations: unanticipated	2	0	1	39
STAFF DEVELOPMENT				
Staff relations: anticipated outcomes	10	4	0	28
Staff training	29	5	2	6
Staff performance and attitudes	23	5	2	6
COSTS				
Costs: anticipated	0	6	0	36

*Evaluator provided no information and/or project did not include this component.

As indicated in Table 2.6, the greatest perceived successes of the Title VII projects in the sample occurred in cognitive student outcomes, in the community/parent involvement component, and in the staff development component. Furthermore, the data show that very little attention is given to unintended outcomes of various components and to matters of project cost.

The evaluator's responsibility does not end with the presentation of the results of his analyses. These results must be discussed and interpreted in such a way that the decision maker can use them. The raters found that only 11 (26%) of the 42 evaluations contained "complete" interpretations of the evaluation findings. Eight (19%) of the evaluations contained "fairly complete" explanations, while 22 evaluations (52%) contained only "very brief" interpretations. In Table 2.6 the judgments of project success were not necessarily supported by data. Our raters were reporting explicit judgments; not necessarily judgments accompanied by supporting evidence. In an extreme case the scores from the experimental group were displayed in the main body of the report and judged there as indicative of success of the instructional program; but in an unreferenced appendix, control group scores provided sufficient evidence to cast serious doubt upon the evaluator's conclusions.

Having presented evaluation results along with appropriate discussion and interpretation, the evaluator has yet one more task. Evaluation results are not an end in themselves; they must lead to a *decision*—a decision to drop the project, a decision to expand it, a decision to focus on a different target population, a decision to make some changes in the structure of the program. Accordingly, the evaluator ought to include in his report some recommendations for project modification that grow out of the data that he has collected, analyzed, and reported. Thirty-nine of the 42 evaluators made recommendations for project modification. Table 2.7 shows the project areas in which modifications were recommended; the modifications tended to be recommended in such relatively "safe" areas as staff training and community/parent involvement. In the more critical areas of instructional program design, objectives, and management, few recommendations were offered. When one considers the degree to which the evaluators' recommendations called for extensive changes, the findings are not encouraging. In only *five* cases did the evaluator recommend major revisions in the project.

TABLE 2.7

Project Areas in which the Evaluator Recommended Modifications (N = 39)

Program Area	N	%
Evaluation Procedures	26	62
Staff Training	24	57
Community/Parent Involvement	18	43
Program Design	16	38
Program Objectives	11	26
Program Management	11	26
Program Implementation	7	17

Project and Evaluation Quality

The final section on the Evaluation Data Sheet concerns the judgments made by the CSE raters. The raters were first asked to describe the evaluator's attitude toward the project as evidenced in the evaluation report. Only five evaluators were judged to have been overly positive (making the project look much better than it actually was). Thirty-four of the evaluators were described as having "positive, but fair" attitudes, while only three evaluators appeared to have been "critical, but fair." None of the evaluators was described as having been "hypercritical."

In regard to the usefulness of the evaluation reports for analyzing project effectiveness, the CSE raters indicated that only six were very useful.

An additional 12 evaluations were described as fairly useful, while 17 were described as being only of limited use. The seven remaining evaluations were thought to have been of no use; that is, the CSE raters *could not,* on the basis of these reports, *make any judgments about the effectiveness of the projects that had been evaluated.*

Following the same procedure used for the evaluation reports, information from each independent educational accomplishment audit was coded on the Audit Data Sheet (Appendix 1). This information is discussed in the following sections.

Audit Report Data

Auditor Characteristics

The most notable characteristic of the audit reports is the absence of information on the auditors themselves. In terms of formal education, all that is known is that one-third held doctorates. Virtually nothing is known either about their field of study or about their professional affiliation (although in most cases where a professional affiliation was given, it was with a private agency). Perhaps this information was not considered essential in a final audit report.

Audit Procedures

Fifty-two audit procedures were enumerated on the Audit Data Sheet. These were derived from a careful examination of the literature on independent audits, including USOE guidelines (HEW, 1971) and the *Handbook for Educational Program Audit* (Morin, 1971). Table 2.8 presents the audit procedures that were employed by at least two-thirds of the 39 auditors. These data were obtained in response to the item "Indicate which sources of data the auditor used as a basis for his report . . ."

As might be expected, one of the two most commonly used data sources was the site visit. (It is worth noting, however, that two projects in the sample did not have an on-site visit from the auditor.) The other most common technique, used by 88% of the auditors, was the review of tests that had been administered by the evaluator. In most cases, however, the auditors did not indicate what criteria they used in judging the adequacy or appropriateness of the tests. Furthermore, it was rare for the auditor to question or challenge the evaluator's selection of instruments. A third data source was the final evaluation report with 36 of the auditors using these reports in conducting the audits. A fourth technique was examination of the raw data gathered by the evaluator; such data included test and questionnaire results. While 36 of the auditors used this source of information, most of the audit reports did not go on to indicate *how* these data were examined by the auditor and what procedures, if any, were used to *verify* the data gathered by the evaluator. There were relatively few

instances where the accuracy of the evaluator's data was questioned by the auditor. Many of the auditors also reported conducting interviews with project personnel, generally with the project director and other staff; interestingly enough, only seven audits reported interviews with students.

TABLE 2.8

Data Sources Employed by at Least Two-Thirds of the Auditors in the Sample (N = 39)

Data Source	N	%
Site Visit	37	88
Evaluator-administered Tests	37	88
Final Evaluation Report	36	86
Raw Data Gathered by Evaluator	33	79
Interviews	29	69

In contrast to the most commonly employed audit techniques, one might also consider some of the *least* frequently employed audit procedures. Table 2.9 shows which audit procedures from among the 52 listed were employed by fewer than one-third of the auditors. Only one of the auditors administered his own tests or questionnaires. Few auditors reported looking at "archival" information such as newspaper clippings, staff correspondence, students' cumulative records, and letters from members of the community. Most of the auditors did not report the use of either the evaluation proposal or the interim evaluation report in their work. This last finding may be related to the fact that our data are based on the contents of the final audit report; one would expect that several of these items might have been treated in the interim audit reports.

TABLE 2.9

Data Sources Employed by Fewer than One-Third of the Auditors in the Sample (N = 39)

Data Source	N	%
Auditor-administered Tests or Questionnaires	1	2
Correspondence with Staff	2	5
Project Financial Records	3	7
Evaluation Contract	5	12
Evaluation Proposal	10	24
Continuation Proposal	9	21
Interim Continuation Report	7	17
Archival Data	12	29

Consideration of the procedures used by the auditors leads to the conclusion that four commonly acknowledged audit practices were employed: site visits, reviews of instrumentation, examination of final evaluation reports, and interviews with staff. The auditors did not administer their own tests, nor did they make extensive use of project archival data.

When reading about educational accomplishment auditing, one invariably comes across the suggestion that part of the audit function should include a re-analysis of some, if not all, of the data that the evaluator has collected and reported. Only in this manner, it is contended, can the auditor truly *verify* the evaluator's findings. Our study yielded few, if any, examples of the re-analysis of evaluators' data by the auditors. Nevertheless, it seems fair to conclude that, in general, auditors were doing the things that auditors are supposed to do. The data suggest, however, that auditors should be more attentive to describing *in detail* the procedures they employ. A number of the auditors *were rather vague about what they did;* there is no room for ambiguity in these matters if the findings of the auditor are to be accepted with confidence. The fact that so many of the auditors did not *report* any re-analysis of evaluation data leads us to question the basis of their judgments of the accuracy of evaluation data and the quality of the data analysis. Twenty-eight auditors commented upon the accuracy of evaluation data and 33 judged the evaluator's analysis of raw data but the impact of these judgments is lost in our confusion as to how they were made.

Auditor's Judgments of the Quality of Evaluation Procedures

One of the primary functions of the independent educational accomplishment audit is to judge the quality of the evaluation. A major section of the Audit Data Sheet dealt, therefore, with those aspects of the evaluation about which the auditor commented and made qualitative judgments. The CSE raters assigned a score ranging from -3 to +3 to elements of the evaluation. At one extreme, -3, the auditor was judged to have said that the evaluation was poor with respect to that particular item. At the other extreme, +3, the auditor was thought to have made a most favorable judgment about the evaluation. In the middle, 0, the auditor was judged simply to have said that an item had been attended to by the evaluation but made no attempt to assess the quality of the evaluation effort. Table 2.10 lists some of the components of the evaluation that were typically commented upon by the auditors.

The item "Accuracy of data gathered by the evaluator" received a mean rating of 1.75, that is, above average in quality. Thus, while the auditors did not report any re-analyses of the evaluator's data neither did they indicate that they questioned the accuracy or authenticity of the data. The item "General evaluation design" received an average rating of 1.22,

which again represents a positive judgment on the part of the auditors. The auditors' judgments ranged between 0 and +1 for most of the other items (assessment techniques, testing instruments, analysis of raw data, appropriateness of data collection techniques, and completeness of report). These findings lead to two possible interpretations. One might conclude that the auditors were making "cautiously positive" judgments about the evaluators' work in the area of data analysis. On the other hand, the auditors may have been content to report that these activities were discussed by the evaluator without making any judgments about the quality of the evaluator's work. Which ever the case, the auditors did not make extreme judgments in either direction regarding the evaluators' skills in data analysis. In general, the auditors viewed the evaluators' work in neutral or only slightly positive terms.

TABLE 2.10

Mean Ratings of Auditors' Judgment of the Quality of Selected Components of the Evaluation (N = 39)

Evaluation Component	Mean Rating*	N[a]	%>1[b]	%>0[c]
General Evaluation Design	1.22	36	77.88	86.11
Assessment Techniques	.94	36	63.89	86.11
Testing Instruments	.77	35	65.71	82.86
Interpretation of Results of Analysis	.32	34	50.00	58.82
Analysis of Raw Data	.30	33	48.48	60.61
Appropriateness of Data-Collection Techniques	.97	32	68.75	81.25
Format of Evaluation Report	.44	32	50.00	59.38
Accuracy of Gathered Data	1.75	28	75.00	85.71
Completeness of Report	0.00	28	42.46	46.43

*On a scale ranging -3.0 = auditor judged evaluation poor on the component to +3.0 = auditor judged evaluation to have been done quite well on the component.

[a]Number of audit reports in which judgments were made.

[b]Percentage of responses greater than +1.0 on a scale ranging from -3.0 to +3.0.

[c]Percentage of responses greater than 0.0 on a scale ranging from -3.0 to +3.0.

Scope of the Audit Report

The analysis of the "scope" of the audit report dealt with two issues: the auditor's agreement with the evaluator's findings and the areas of the project in which modifications were recommended by the auditor. Data on the agreement of the auditor with the evaluator's findings were derived from the CSE raters' consideration of a number of areas of project outcome such as student learning, student attitudinal change, and commu-

nity/parent involvement. In each instance, the rater indicated whether the auditor (a) agreed with the evaluator's findings, (b) did not agree with the evaluator's findings, or (c) did not comment on evaluation findings with respect to this element of the project. The auditors tended to support the evaluators' conclusions regarding student cognitive achievement, staff development, and community/parent involvement. Indeed, there were few cases where the auditor specifically disagreed with or disputed the evaluator's findings. In those cases where "agreement" was not reported, the auditor generally did not comment on the evaluator's findings.

Data on auditors' recommendations for project modification were obtained by having the CSE rater consider some fifteen areas of activity, such as evaluation design, data analysis techniques, and provision for unanticipated outcomes. In each case the rater indicated whether or not the audit report contained a recommendation for modification. The project areas where auditors most frequently recommended modifications were the design of the evaluation (34 auditors) and the data collection techniques (28 auditors). Clearly these areas are entirely appropriate for consideration by the auditor. One would wish, in fact, that the auditors' recommendations for modification in these areas had been even stronger. The recommendations and conclusions tended to be quite vague and, in general, were not based upon specific situations or data. It is worthwhile to note the areas of the project in which auditors did not make recommendations: staff development, community involvement, inservice training, and design of the instructional program. Auditors have been warned to avoid becoming involved in program management; our data suggest that they have followed this warning.

Quality of Audit Report

Twenty of the 39 audits were described by our raters as either "very useful" or "fairly useful."[2] Eighteen were described as "of limited use," while only one was thought to be "of no use." The auditors' attitudes toward the evaluations were described as "positive, but fair" in 21 cases, as "critical, but fair" in 14 cases, and as "overly positive" in four cases. In no case was an audit described as hypercritical or as a fault-finding exercise.

Data Derived from Project Director Questionnaire

A further source of descriptive data on the Title VII evaluations was information obtained from the Project Director Questionnaire (Appendix 1). The main purpose of this questionnaire was to assess the impact of

[2]Our decision to include "usefulness" as in indicator of quality derives from our position that evaluation and/or audits that are not potentially useful have failed to meet their primary objective and can therefore be considered of poor quality.

the evaluation on one of the key local decision makers and the person to whom *formative* evaluation information was directed—the project director. A second purpose was to obtain supplemental data for areas where gaps existed in the evaluation and audit reports. As noted in Chapter I, a 100% response rate was achieved by using telephone interviews as an optional response mode or follow-up for the project director.

Evaluator Characteristics

According to project directors the typical project had an evaluation team consisting of a principal evaluator supervising several individuals. This evaluation team was external to the project in 20 of the projects (48%) and internal in 16 (38%). The remaining six projects had both internal and external components. Unfortunately the accuracy of this information is suspect since project directors tended to refer to the auditor as an external evaluator, a classification we had not intended. This became apparent in the course of some of the telephone conversations with project directors and probably accounts for some disparities between response categories that should overlap. (For example, 16 projects—38%—claimed internal evaluation but only 11—26%—identified the evaluator as being a member of the project staff.)

External evaluators were frequently from a university (29%) or a private agency (29%), and in the latter case they were often connected with a university. In 18 projects, the evaluator held a master's degree; 21 evaluators held doctoral degrees. The academic fields most widely represented were education (48%) and psychology (26%).[3]

In 18 projects, evaluators were located via recommendations while 10 projects relied on contacts from previous professional affiliation. This usually meant that the evaluator had performed a similar service for the school district at a prior time. When there was more than one candidate for the position of the project evaluator, final selection responsibility was frequently (13 projects or 31%) delegated to a selection committee composed of the project director and district personnel. Finally, 32 project directors (76%) indicated that evaluation activities commenced before the school year was under way.

Formative Evaluation Activities

Since the project director is the person directly responsible for the ongoing activities of the instructional program and any possible modifica-

[3]Note that the item (PD16) on the Project Director Questionnaire from which this information was obtained makes use of the phrasing "what do you believe was . . ." This wording reflects our recognition that project directors are being asked to provide information about activities that took place two years earlier and in an area (evaluator's degree) in which the project director might not have had definitive information or interest.

tions, we expected formative studies to have an impact on him. This was indeed the case. Formative studies were reported for 39 of the 42 projects (93%) and were viewed as useful by all 39 of these project directors. This stands in contrast to our report earlier in this chapter that formative evaluation studies were described in only slightly more than half of the evaluation reports. This hidden, non-explicit nature of formative evaluation is something that we all suspect but which has been insufficiently documented in the literature. That evaluators had been actively involved in formative activities was evident from responses to two particular questions. Thirty-two project directors (73%) reported that formative evaluation studies were useful in "guiding project staff in making modifications in the program." The questionnaires clearly explained what was meant by "formative evaluation" and project directors' use of the term confirmed their understanding of its meaning. When asked to rate the extent to which evaluation was of use in various areas, "identifying possible problem areas" and "general recommendations for program changes" were among the areas receiving the highest average ratings from project directors. We thus feel confident in reporting that in many cases formative evaluation was conducted by evaluators and found useful by project directors but was not evident in final evaluation reports.

When we note the impact of the evaluation in the areas of "guiding project staff in making modifications in the program" and "general recommendations for program changes" the issue of the line separating evaluative assistance and management consulting arises. The responses above leave room for suspicion that evaluators were being co-opted as management consultants. Conversely, the project directors were perhaps co-opted into evaluation roles, making it a two-way affair. This is suggested by the fact that 18 of the project directors (43%) reported that they or their representative assisted in *writing* the final evaluation reports.

This close association between evaluator and project director (which is discussed further in Chapter V) and the provision of valuable formative evaluation information would require the frequent presence of the evaluator; project directors indicated that in 31 of the 42 projects (74%) the evaluator made "more than 5" site visits during the school year. (This was the highest response category provided on the questionnaire.)

The evaluator's generally close ties with the project are further confirmed by his participation in establishing criteria, an activity for which instructional program developers and curriculum specialists should surely take responsibility. The evaluator is not to judge the program according to *his own* criteria but to provide information upon which the decision maker can judge the program. Nevertheless, in this study project directors indicated that over half of the evaluators were involved in developing criteria for the objectives of the instructional program. When the evaluator

was not involved, criteria were generally accepted directly from the project proposal[4] (See Table 2.11).

TABLE 2.11

Procedures Indicated by Project Director as Being Most Often Used in Developing Criteria for Program Objectives

	Procedure	N
1.	Criteria Accepted Directly from the Objectives as Stated in the Project Proposal	15
2.	The Evaluator Selected Criteria	1
3.	Criteria Developed Jointly by the Evaluator(s) and Project Staff	21
4.	Criteria Based on Pilot Studies	0
5.	Other*	4
6.	Not Known	1

*The responses in this category indicated that a combination of procedures was used. In addition to the procedures described for response categories, one project reported the use of consultants.

Concerning the various activities performed by internal or external personnel, the project directors indicated that while the external personnel observed programs in operation, examined records, and took sole responsibility for data analysis, rarely were they involved in the critical activity of test administration. As we noted earlier, the procedures for gathering data are almost never reported in the final evaluations, a fact which inevitably detracts from credibility as we are left with the suspicion that tests may have been administered by teachers without training in the necessary standard procedures.

Evaluation Reporting

Typically, all final evaluation reports were sent to the project director who then distributed them to teachers, the school board, federal sponsors, state agencies, and others (See Table 2.12). Parents sometimes received reports; news media rarely. For 23 of the projects (55%) there were specially written reports for the community; these were generally condensed, in non-technical language, and aimed to increase the community's confidence in the project.

[4]This information would appear to contradict questions 62–66 on the Evaluation Data Sheet. There, 32 of the 42 evaluators indicated that the project's prespecified goals were adopted (with no modifications or review). These discrepencies could be a function of the small attention given to this area in the evaluation reports. It is somewhat surprising that such a vital issue—the standards against which the project is to be evaluated—would be so neglected by the evaluator in his reporting.

TABLE 2.12

Project Director's Indication of Final Evaluation Report Distribution

Groups	received report	did not receive report	not known
Project Director	41	0	1
Teachers	29	4	9
Parents	15	12	15
Students	1	27	14
School Board	33	5	4
Federal Sponsors	41	0	1
News Media	7	18	17
Others*	29	0	13

*In this category "state agencies" were generally mentioned.

Project Directors' Reactions to the Evaluation

In general, project directors reacted quite favorably to the evaluation. Table 2.13 displays the project directors' responses to the item, "Please rate the extent to which the evaluation was of assistance to you in the indicated areas." Inspection of this table indicates that project directors found evaluations most useful in identifying possible problem areas. Preparing reports, developing criteria, and general recommendations for project changes were other areas in which the evaluation was often viewed as very useful. The areas in which evaluations were most frequently described as "not useful" were changing personnel, changing program management, and changing community relations activities.

When asked about their preference for an evaluator, project directors in our sample of Title VII projects, where evaluations have been mandatory, indicated that they are now firmly convinced of the value of evaluators, particularly *internal* evaluators. Twenty-five project directors (60%) indicated they would prefer both an internal and external evaluator, only one indicated a preference for external only, and *no* project director indicated that he would prefer not to have an evaluator on the project. Evaluators have apparently convinced project directors of their worth.

This finding of the usefulness of formative evaluation at the local level—that is, to project directors—contrasts strongly with a report by Wholey and White (1973) concerning Title I evaluations. They reported:

> Most of these (local) project evaluations yield little information useful to local project improvement, and play almost no part in influencing Title I operations at the local level (p. 6).

TABLE 2.13

Project Director's Ratings of the Evaluation's Utility for a Variety of Activities

Activity	Average usefulness on four-point scale	Number of responses in each category: not useful 1	moderately useful 2	useful 3	very useful 4	"activity not performed"
Identifying possible problem areas	3.17	2	8	11	19	2
Preparing reports	3.15	1	8	14	16	3
General recommendations for project changes	3.12	2	8	13	17	2
Helping to develop criteria	3.02	2	10	13	15	2
Changing in-service activities	2.87	8	3	15	14	2
Changing teaching procedures	2.65	3	14	14	9	2
Changing the pace of instruction	2.50	6	14	14	6	2
Changing instructional materials	2.50	9	9	15	7	2
Changing program management	2.32	13	7	11	7	4
Changing community relations activities	2.18	12	9	15	2	4
Changing personnel	1.56	20	10	3	1	8

The differences that led to such contrasting reactions to the utility of the evaluation procedure indicate a direction to further research.

When project directors were asked to indicate agreement or disagreement with statements that the evaluation had influenced decisions to modify the project, the majority felt that the evaluation had been influential, both during the school year (formative value) and also, more often, in decisions made regarding the following school year. These responses were substantiated by questions asking how influential the evaluations had been. Again, the majority of the project directors indicated that evaluation had been influential in decision making particularly on those decisions concerning modifications in the following year.

The final items on the Project Director Questionnaire asked about the impact of auditors and federal monitors on local decision making. The extent of influence of federal monitors varied considerably among projects, ranging from "not at all" to "very influential" with frequencies rather evenly distributed. Auditors appear to have been a slightly more consistent source of influence.[5] The correspondence between recommendations from

[5]See Table 4.3, Chapter IV.

the federal monitors and those from evaluation and audit reports were generally indicated to be "average" or "strong" but interview data indicated some vagueness in recalling this information. Federal recommendations in particular did not seem to have been salient or well-remembered.

Summary

This chapter has presented descriptive data derived from extensive, systematic examinations of the Evaluation and Audit Data Sheets and the Project Director Questionnaires. In general it was found that the evaluation reports reveal little information on the evaluators themselves, their training, or institutional affiliations. While almost all evaluations included formative activities, it was noted that evaluators tended to be rather vague about the specific procedures that were employed in the formative evaluations. Turning to summative procedures, most projects (64%) employed evaluation designs that would be designated "pre-experimental" in the Campbell and Stanley (1966) classification system.

In their summative evaluation, the evaluators generally confined themselves to stated project objectives and only rarely looked at unanticipated outcomes. The most common data-gathering device for outcome evaluations was the standardized test. The evaluators most frequently employed simple descriptive statistics in their data analysis.

Evaluators tended not to report lack of success with respect to the achievement of project goals; there were no documented failures. In a related finding it was noted that few evaluations contained a complete interpretation of evaluation findings. It was rare for the evaluator to recommend major revisions on the basis of evaluation findings. Finally, the CSE raters found the evaluators to have been fair, not overly critical. Just under half of the evaluations were judged to have been either "very useful" or "fairly useful" for the purpose of analyzing the effectiveness of the project being evaluated.

The data on the independent educational accomplishment audits suggest that the auditors performed their services in accordance with the general intent of the USOE guidelines. Those activities in which auditors are expected to engage did in fact take place on most occasions. Auditors were able to make judgments about evaluation procedures and were able to make recommendations for modification on those areas of the project related to the evaluation.

The independent audit is a new concept and, as such, has not been implemented with complete perfection. Auditors should be encouraged to detail clearly and precisely the activities in which they engaged and which serve as the basis for their judgments about the evaluation in particular and the project in general. There should be no mystery surrounding the procedures that were employed by the auditor. Accountability requires full disclosure. The auditors tended to avoid making strong or harsh judg-

ments about the evaluation efforts that they reviewed. Their timidity is quite understandable. The audits in question were completed some three years ago when the concept was even newer than it is today. Furthermore, there is an inclination among the auditors to practice "professional courtesy" toward the evaluator (after all, the roles may be reversed next year!).

Finally, this investigation suggests that relatively few auditors have attempted to re-analyze an evaluator's data for purposes of verification. There are a variety of possible explanations for this fact, including the limitations of time and funds and some question as to the values to be derived from such activities. It seems that those who are responsible for commissioning audits should give some consideration to the desirability of requiring this particular procedure.

Responses from the project directors provided information about the evaluators that was not included in the evaluation reports. The evaluations were frequently conducted by university-based personnel holding doctoral degrees; all but 7% held at least a master's degree.

An expected finding was the fact that all the project directors who received formative evaluation information found it to be useful. In fact, the directors maintained a close working relationship with the evaluators in three-fourths of the projects.

Project directors found the evaluation to be at least moderately useful in seven of the 11 areas indicated on the Project Director Questionnaire. The evaluators' recommendations were especially influential in making modifications for the following year.

The project directors did not seem to be as influenced by other persons assigned to the project—the independent auditor and the federal monitor. The federal requirements seem to have been assimilated in at least one context: when asked their preference for an internal evaluator, an external evaluator, both types, or no evaluation, none of the project directors selected no evaluator—an independent decision not possible under present funding regulations.

Chapter III

EVALUATION RELATIONSHIPS

The characteristics of the evaluation and audit reports and information obtained from the Project Director Questionnaires have been described in Chapter II. The variables derived from a reading of these documents will now be organized into several underlying dimensions or "intuitive factors" reflecting aspects of an evaluation and/or audit that might influence federal or local decision making. For each factor indices will then be constructed. These indices will be used in subsequent chapters to study the impact and quality of evaluation for decision making.

Selection of Factors

Our initial data collection resulted in a large number of variables, many of which were correlated. Factor analysis would have been an inappropriate (and unproductive) technique for categorizing the variables since there were only 42 cases in the study and a much greater number of variables. Further inspection of the data revealed that the variables seemed to reflect seven underlying constructs or intuitive factors:

- project context
- project quality
- evaluation context
- evaluation report comprehensiveness
- evaluation report sophistication
- evaluation report quality
- evaluation report physical attributes

We intuited those characteristics that might be included within each of the factors. The *project context* factor, in our judgment, might measure such characteristics as grade levels involved and size and SES of the parent district. The *project quality* factor might reflect judgments concerning the value of the experimental program. The *evaluation context* factor might include the composition of the evaluation team, size of the evaluation effort, and the closeness of the evaluation team to the program. The *evaluation report comprehensiveness* factor might assess the scope of evaluation activities as well as the completeness of coverage of each activity in the evaluation report. *The evaluation report sophistication* factor might be reflected by the evaluator's training, the kinds of project areas investigated, and the kinds of analyses undertaken. The *evaluation report quality* factor might assess the excellence of the evaluator's endeavor as judged

by the auditors and the CSE raters. The *evaluation report physical attributes* factor will be described later.

Recoding of Variables and Analysis of Factors

After identification of these factors, the total set of variables was reviewed. For each factor a group of variables was selected which appeared, by examination of their content, to represent the factor. In some cases, as will be described, new variables were added. The convention for labelling variables was as follows: item 1 on the Evaluation Data Sheet was Variable E1, item 2 was E2, and so on to E135 for item 135. Similarly variables representing the 124 items from the Audit Data Sheet were A1 to A124, and variables from the Project Director Questionnaire were PD1 to PD24b. The 11 new variables added were labelled NV1 through NV11. As previously noted, the Evaluation Data Sheet, the Audit Data Sheet, and the Project Director Questionnaire are reproduced in full in Appendix 1.

The variables defining each factor were of three types. First, most variables were taken directly from the already existing bank of variables reported in the Evaluation and Audit Data Sheets and the Project Director Questionnaires. However, in order to avoid the effect of some heavily skewed distributions (and in anticipation of future data analysis, i.e., intercorrelation among variables within a factor) each of these variables was dichotomized close to the median. (All of the variables of this first type have labels beginning with an E, A, or PD.)

Second, some new variables were derived from existing variables. For example, NV1 was created to measure the extent of demographic information on the target district and community. This variable was created by combining responses to several items related to this area. Each item for which demographic information was provided was coded "1;" if no information was provided then the item was coded "0." The sum of the scores on these items was therefore an indication of the total amount of demographic information provided in the evaluation report. Thus NV1 = recoded (E103 + E104 + E105 + E106 + E107); the possible range was 0 to 5. (Variables of this second type are labelled NV1 through NV5.)

Finally, the third type of variable measured information that was not coded initially but was readily accessible either from reports on file or from government publications. For example, one piece of data in the third category is the presence or absence of the name of the evaluator on the evaluation report. We had assumed that every report would contain this information and CSE collected it after its absence became apparent. (Variables NV6 through NV11 fall into this third category.)

Figure 3.1 lists all new variables, shows their composition, and indicates, if relevant, where they were dichotomized.

FIGURE 3.1

New Variables Defined*

NV1 = **Sum dichotomized (E103 to E107)—Extent of demographic information in evaluation report. Possible range: 0-2.**

NV2 = **Sum dichotomized (E62 to E65)—Variety of procedures used to select criteria. Possible range: 0-4.**

NV3 = **Sum dichotomized (E69 to E79)—Variety of project components considered by evaluator. Possible range: 0-11.**

NV4 = **Sum dichotomized (E80 to E88)—Variety of data sources used. Possible range: 0-9.**

NV5 = **Sum dichotomized (PD20a to PD20k)—Usefulness of evaluation to project director. Possible range: 0-11.**

NV6 = **Project located in a standard metropolitan statistical area (0 = no; 1 = yes)**

NV7 = **Average enrollment per grade level for district (dichotomized at mean = 8,398: $<$8,398 = 0, $\geq$8,398 = 1). Possible range: 0-1.**

NV8 = **Number of main body pages in evaluation report (dichotomized at mean = 60: $<$60 = 0, $\geq$60 = 1). Possible range: 0-1.**

NV9 = **Evaluator identified by name on the cover or title sheet of evaluation report (0 = no; 1 = yes)**

NV10 = **Weight of evaluation report (dichotomized at mean = 15.4 oz.: $<$15.4 = 0, $\geq$15.4 = 1). Possible range: 0-1.**

NV11 = **Year of project's Title VII funding (1 = first; 2 = second). Possible range: 1-2.**

***The code values assigned to each response category are reported in this figure for the new variables. Code values for the original variables can be found by referring to the data collection forms in Appendix 1. In addition, where original variables were dichotomized, the new coding values are reported in Appendix 2.**

With new dichotomous variables composed and old variables dichotomized, correlation tables were drawn up for the variables within each of the seven factor areas. These intercorrelations were then analyzed and for each factor a best variable and an index were selected to represent the respective factors in subsequent regression analyses.

Selection of the best variables was based on considerations of the statistical properties and face validity of the variables identified as representative of the factors. For each factor, that variable which exhibited the largest number of statistically significant correlations with other defining variables and which met our *a priori* notions concerning the theoretical nature of the factor was chosen as a most representative variable (MRV).

In addition to the MRV an index was constructed; for a given factor, those defining variables with the greatest face validity and the strongest pattern of significant correlations were included in order to maximize the number of factor characteristics represented.

In the following section, the intercorrelation tables are presented and the choice of variables and indices to represent each factor is indicated.

Factor 1: Project Context

Table 3.1 displays the data from which Index 1, project context, was derived.

TABLE 3.1

Factor 1: Project Context

A. **Intercorrelations of Variables (N = 42)**

Variable	E107	NV6	NV7	NV11	FY70/71	FY71/72	FY72/73
E107	–	.29*	.02	-.15	.07	.13	.16
NV6		–	.27*	.10	.37**	.36**	.33**
NV7			–	-.20	.22	.23	.24
NV11				–	.28*	.10	.06
FY70/71					–	.80**	.80**
FY71/72						–	.99**
FY72/73							–

*$p < .05$
**$p < .01$

B. **Description of Variables**
E107 = Number of schools participating in the Title VII project
NV6 = Standard Metropolitan Statistical Area
NV7 = Average enrollment per grade level for district
NV11 = Year of project's Title VII funding
FY 70/71 = Funding level for fiscal year 1970-1971
FY 71/72 = Funding level for fiscal year 1971-1972
FY 72/73 = Funding level for fiscal year 1972-1973

C. **Most Representative Variable**
MRV1 = NV6

D. **Index**
Index 1 = NV6 + NV7 + NV11
(Range of Index 1 is 0 to 3)

*Recall that the code values for new variables can be found in Figure 3.1, the code values for original variables in Appendix 1, and the code values for dichotomized original variables in Appendix 2.

Defining variables. Seven variables were identified as contributing to the project context factor. Because of the scarcity of demographic information available from the evaluation and audit reports, three of these were new variables. New variable 6 (NV6) indicates whether or not the target project was located in a community that was part of a Standard Metropolitan Statistical Area (SMSA). Information was based on the 1970 census report. This variable was intended to serve as a measure of the target district's "urbanness" or proximity to large cultural centers (where educational stimuli as well as technical consultants may be in greater abundance). As an indicator of school district size, NV7 measured average enrollment per grade level. Once again these data were taken from the 1970 census report. Finally, NV11 indicated whether the project had been receiving Title VII funds for one year or two years.

Intercorrelations. The generally high number of correlations between NV6 and the other variables seems to indicate that projects located within a SMSA are larger in terms of number of participating schools and district size and, in addition, tend to be funded at a higher level. This "urbanness" variable was selected as the most representative variable (MRV) of project context.

The almost perfect correlation between funding level in FY71 and FY72 indicates that there was little in the way of differential funding among these Title VII projects in 1972. The lower correlation between funding for FY70 and FY71 is probably due to different "start up" costs among projects or to the possibility that some projects operated on planning grants in FY70, rather than due to refunding decisions based on project merit. The high correlation between funding levels served as a warning that one of our indicators of decision making (i.e., one of our dependent variables) was unlikely to show much meaningful variance. There is the additional problem that funding decisions for FY72 were probably made before the FY71 final evaluation reports were received by federal personnel, but might have relied on FY70 evaluation reports, where they were available.[1]

Index 1 (NV6 + NV7 + NV11). New variables 7 and 11 (NV7 and NV11) were dichotomized and added to NV6 to form an index for project context. Index 1 therefore reflects the location, urban or otherwise, of the target population (NV6), the district size (NV7), and the project's age as a Title VII effort (NV11). This index has a range of 0 to 3 points.

[1]It is perhaps helpful to note here that the federal fiscal year runs from July 1 to June 30 and is named by the year in which it ends. For example FY71 denotes the period from July 1, 1970 to June 30, 1971.

Refunding decisions for Title VII projects were usually announced in April for the following fiscal year (beginning in July). Final evaluation reports were usually submitted no earlier than June, and hence were received by federal officers after the refunding decisions for the upcoming fiscal years had been announced. In practical terms it would seem that the final evaluation report for FY71 (or school year 1970–71) could only possibly influence refunding decisions for FY73 (or school year 1972–73).

A high value indicates an urban location, a large size of district (reflected by large enrollment per grade level), and/or a project in its second year of federally funded operation.

Factor 2: Project Quality

Table 3.2 displays the data from which Index 2, project quality, was derived.

TABLE 3.2

Factor 2: Project Quality

A. **Intercorrelations of Variables (N = 42)**

Variable	A123	A120	E120	E126	E129	E133	E134
A123	–	.48**	.01	.06	.21	.25*	.61**
A120		–	.14	.37**	.31*	.40**	.50**
E120			–	.30**	.17	.27*	.11
E126				–	.01	.34*	.02
E129					–	.47**	.47**
E133						–	.50**
E134							–

*p < .05
**p < .01

B. **Description of Variables**
A123 = Raters' judgment of project based on audit report
A120 = Auditor's judgment of project
E120 = Success of parent-community relations (as reported by evaluator)
E126 = Success of affective outcomes (as reported by evaluator)
E129 = Success of staff performance and attitudes (as reported by evaluator)
E133 = Evaluator's judgment of project
E134 = Raters' judgment of project quality based on the evaluation report

C. **Most Representative Variable**
MRV2 = E133

D. **Index**
Index 2 = E133 + A120 + 1/2(E134 + A123 + E126 + E129)
(Range of Index 2 is 0 to 4)

Defining Variables. All of the variables identified with this factor come from the Evaluation and Audit Data Sheets. The variables are of two types. Variables A120, A123, E133, and E134 measure general project quality as indicated in audit and evaluation reports (and as judged by the CSE raters). Although these variables derive from a subjective rating,

the evaluator's or auditor's overall attitude toward the project was usually fairly obvious from a reading of his report. The remaining variables measure the success of specific aspects of the target project (for example, affective outcomes) as reported by the evaluator.

Intercorrelations. Variable E133, the evaluator's overall judgment of the project as perceived by the CSE raters, received a significant correlation with every other variable. This seems to suggest that not only does the evaluator share a similar opinion of the project with CSE raters and the auditor, but also his attitude reflects project successes in a variety of areas. Variable E133 was therefore selected as the MRV for Factor 2.

The highest correlation was between two judgments that raters had made of each project, one at the end of the reading of the final evaluation report (E134) and one after reading the audit report (A123). On the whole, general opinions as to quality are more highly correlated than judgments of quality in specific areas.

Index 2 [$E133 + A120 + 1/2(E134 + A123 + E126 + E129)$]. Variables E133 and A120, the evaluator's and the auditor's overall judgments of project quality, were given greatest weight. These variables were most highly intercorrelated with other variables and in addition were hypothesized to have the greatest impact on decision making, particularly at the federal level. In addition, four other variables were included. Two of these variables were subjective (E134, A123) and two were objective (E126, E129). They were each given a weight of one-half in the index. A high value for Index 2 thus represents a project perceived by evaluators and auditors (and through them CSE raters) to be of high quality and successful.

Factor 3: Evaluation Context

Table 3.3 displays the data from which Index 3, evaluation context, was derived.

Defining variables. Variables identified with Factor 3 come from the Evaluation Data Sheet and the Project Director Questionnaire. Two items, the external-internal position of the evaluator with respect to the project and the character of the evaluation team (single individual vs. organization) were in both the Project Director Questionnaire (PD1, PD3) and the Evaluation Data Sheet (E1, E2).[2] Other variables measured the extent of formative evaluation activities (PD10) and the number of site visits

[2]Several items in the Evaluation Data Sheet were repeated in the Project Director Questionnaire. In some cases we only desired to validate the data. In other cases, however, information provided by the evaluators was incomplete or simply not available. The latter situation held for variables E1 and E2. Because the relative position of the evaluator has long been debated in the literature, we sought this information from the project director. In Chapter II, data from both sources were reported. In the analyses both variables were included in the factors in order to validate each other, but only one source was selected for inclusion in an index. Usually that source was the Project Director Questionnaire responses as these provided more complete data.

made by the evaluation team (PD17). These latter two variables were hypothesized to indicate the effort that went into the evaluation.

TABLE 3.3

Factor 3: Evaluation Context

A. **Intercorrelations of Defining Variables (N = 42)**

Variable	E1	E2	PD1	PD3	PD10	PD17
E1	–	-.01	.57**	.17	-.08	-.11
E2		–	-.37**	.12	-.05	.07
PD1			–	.21	-.01	-.19
PD3				–	.01	-.36**
PD10					–	.13
PD17						–

**$p < .01$

B. **Description of Variables**

E1 = Evaluation contracted to a single individual or to an organization
E2 = Internal or external evaluator as reported by evaluator
PD1 = Evaluation contracted to a single individual or to an organization as reported by project director
PD3 = Internal or external evaluator as reported by project director
PD10 = Extent of formative evaluation reported by project director
PD17 = Number of site visits by evaluator as reported by project director

C. **Most Representative Variable**

MRV3 = PD3

D. **Index**

Index 3 = PD3 + PD1 + PD17
(Range of Index 3 is 0 to 3)

Intercorrelations. It is surprising to note that variables E2 and PD3, each asking whether the evaluator was internal or external, were not significantly correlated. (E1 and PD1, relating to the nature of the evaluation, were correlated, although not as highly as might have been expected.) There are several possible ways in which this lack of agreement could have arisen: (1) The project directors did not discriminate as strictly between the two categories as did the CSE raters. (For example, people sent to the project from the district office might be seen as external by project directors but would have been rated as internal by CSE raters.) (2) The evaluator's view of his position relative to the project differed from that of the project director. Evaluators frequently perceived themselves as "external" to the project, for example, while the project director

saw the evaluation as being paid for by project funds, utilizing project office space and services, and therefore "internal." As another example, project directors viewed members of the district research staff as "external;" CSE raters considered such individuals as "internal." (3) Project directors tended to refer to the auditor as an external evaluator. We realized the confusion too late in the investigation to avoid this data collection problem.

Index 3 (PD3 + PD1 + PD17). Analyses of the intercorrelations (Table 3.3) offered arguments for the selection of either PD1 or PD3 as the most representative variable. We chose PD3 as the MRV for Factor 3 based primarily on our view of its greater content validity. In addition to PD3, PD1 and PD17 were included in Index 3. This index reflects the evaluator's position with respect to the project (PD3), the composition of the evaluation team (PD1), and the closeness of the evaluator to the project in terms of frequency of site visits (PD17). A high value for Index 3 indicates an external evaluation team and many site visits.

Factor 4: Evaluation Report Comprehensiveness

Table 3.4 displays the data from which Index 4, evaluation report comprehensiveness, was derived.

Defining variables. The variables in this factor measured the presence or absence in the evaluation report of various kinds of information. For example, Variable E103 measured the presence of information on the target community; E30 indicated whether or not progress evaluation activities were reported. These items were coded such that a "1" represented the inclusion of information in the evaluation report and "0" the lack of this information. In addition, as noted, several new variables were created by summing the number of "no information" vs. "some information" across several variables. New variable 1 (NV1), for example, measured the extent of demographic information on the target district and community by summing items E103 to E107, which were all questions about the school district and the community in which it was located.

Intercorrelations. There were a large number of significant correlations in this factor, indicating a tendency for evaluation reports that were complete in one aspect to be thorough in other aspects. The variable that we considered most representative was NV3 (variety of project components considered by evaluator) which had a significant correlation with 10 other variables. Somewhat surprising were the many significant correlations with E16 (whether or not an evaluator's professional affiliation was recorded somewhere in the evaluation report). This observation suggests, perhaps, that anonymous evaluations were less comprehensive and less thorough than evaluations that clearly indicated the responsible author.

TABLE 3.4

Factor 4: Evaluation Report Comprehensiveness

A. **Intercorrelations of Variables (N = 42)**

Variable	E16	E21	E30	E96	E98	E99	E101	E102	E103	E107	NV1	NV2	NV3
E16	–	.50**	.29*	.22	.10	.10	.48**	.42**	.33*	.32*	.41**	.05	.44**
E21		–	.28*	.21	.16	.35*	.15	.34*	.17	.08	.16	.25*	.50**
E30			–	.18	-.09	.33*	-.06	.29*	.14	.19	.20	.00	.27*
E96				–	-.03	-.05	.00	.34*	.03	.18	.09	.14	.13
E98					–	.32*	.19	-.12	.59**	-.01	.38**	.02	.08
E99						–	.17	.35*	.01	.22	.14	.00	.29*
E101							–	.18	.13	.23	.22	.32**	.45**
E102								–	-.05	.41**	.22	.18	.33**
E103									–	.25*	.81**	.19	.26*
E107										–	.78**	.22	.30*
NV1											–	.26**	.36**
NV2												–	.33*
NV3													–

*p < .05
**p < .01

B. **Description of Variables**

E16 = Information on the evaluator's professional affiliation
E21 = An indication of implementation evaluation activities
E30 = Indication of progress evaluation activities
E96 = Information on staff performance
E98 = Well-interpreted analyses of results
E99 = A full description of assessment instruments
E101 = Comments on staff performance
E102 = Information on parent and community involvement
E103 = Information on target community
E107 = Information on number of experimental schools
NV1 = Extent of demographic information
NV2 = Variety of procedures used to select criteria
NV3 = Variety of project components considered by evaluator

C. **Most Representative Variable**

MRV4 = NV3

D. **Index**

Index 4 = NV1 + NV3 + E99 + E102 + E103
(Range of Index 4 is 0 to 5)

Index 4: (NV1 + NV3 + E99 + E102 +E103). Index 4 is the sum of those variables that appeared to define Factor 4 from a content validity point of view and which were correlated with a majority of the variables. When the set of most highly correlated variables was examined, it was decided not to include E16 in the index, in part because it did not seem compatible with the rest of the members of the index, and in part because it seemed too highly related to the variables in Factor 7. To that end, after variable selection, Index 4 reflects the extent of demographic information (NV1), the variety of project components considered by the evaluator (NV3), the extent of description of assessment instruments (E99), and the presence of information on the target community and parent involvement (E102 and E103). A high value for Index 4 indicates that the final evaluation report attended to a wide variety of project components.

Factor 5: Evaluation Report Sophistication

Table 3.5 reports the data from which Index 5, evaluation report sophistication, was derived.

Defining variables. These variables reflected our conception of a sophisticated evaluation as including both implementation and progress evaluation (E21 and E30), as considering unanticipated as well as anticipated outcomes (E68), as utilizing a variety of data sources (NV4), as containing well-interpreted analyses of results (E98), and as using a variety of procedures to select criteria (NV2).

Intercorrelations. Although the variables identified with this factor were not highly intercorrelated, Variable E30 did correlate significantly with several variables, suggesting that they share some variance. Variable E30, which represents the presence of progress evaluation information, was selected as the most representative variable.

Information regarding the selection of evaluation criteria (NV2) showed only one significant correlation (with E21, implementation evaluation). It would appear that the frequency of substantiating, or even stating criteria, is not related to either the comprehensiveness or the sophistication of the evaluations.

Index 5 (E21 + E30 + E91 + NV4). Index 5 is composed of some rather diverse variables. The presence of implementation evaluation (E21), progress evaluation (E30), the use of hypothesis-related statistics (E91), and the variety of data sources used by the evaluator (NV4) are all equally weighted to form Index 5, evaluation report sophistication. A high value on this index indicates an evaluation report that is marked by its "sophistication" in terms of the kinds of evaluations performed, in the variety of measures used, and in technical appropriateness of its data analyses.

TABLE 3.5

Factor 5: Evaluation Report Sophistication

A. **Intercorrelations of Defining Variables (N = 42)**

Variable	E21	E30	NV2	E68	E69	NV4	E91	E98	E100
E21	–	.28*	.25*	.15	.05	.40**	.20	.16	.14
E30		–	.00	.22	.23	.35*	.28*	-.09	.24
NV2			–	.19	-.16	.19	.06	.02	-.04
E68				–	.01	.07	.24	.29*	.21
E69					–	.24	.03	.12	-.04
NV4						–	.05	-.22	-.01
E91							–	.08	.84**
E98								–	.24
E100									–

*$p < .01$
**$p < .001$

B. **Description of Variables**

E21 = Indication of implementation evaluation activities
E30 = Indication of progress evaluation activities
NV2 = Variety of procedures used to select criteria
E68 = The consideration of unanticipated cognitive outcomes
E69 = The consideration of affective outcomes
NV4 = Variety of data sources used
E91 = Use of the hypothesis-related statistics (e.g., "t" tests, analyses of variance)
E98 = Well-interpreted analyses of results
E100 = Extensive technical language used in the evaluation report

C. **Most Representative Variable**

MRV5 = E30

D. **Index**

Index 5 = E21 + E30 + E91 + NV4
(Range of Index 5 is 0 to 4)

Factor 6: Evaluation Report Quality

Table 3.6 reports the data from which Index 6, evaluation report quality, was derived.

Defining variables. All variables for this factor come from the Audit Data Sheet except E135 which represented the CSE rater's opinion of the quality of the evaluation report. The remaining variables measured

the auditor's opinion[3] of various aspects of the evaluation report (assessment techniques, accuracy of data, format, etc.). Variable A124 is an assessment of the auditor's opinion of the total evaluation report quality. It should be stressed that all "auditor's opinions" represent interpretations made by multiple CSE raters on reading the audit reports.

TABLE 3.6

Factor 6: Evaluation Report Quality

A. **Intercorrelations of Defining Variable (N = 39)**

Variable	E135	A53	A57	A67	A68	A70	A71	A80	A124
E135	–	.09	.52**	.15	.53**	.27*	.21	-.01	.37**
A53		–	.40**	.37**	.29*	.44**	.52**	.29*	.39**
A57			–	.42**	.48**	.63**	.17	.12	.29*
A67				–	.63**	.46**	.31*	.41**	.21
A68					–	.42**	.36**	.15	.40**
A70						–	.53**	-.01	.60**
A71							–	.02	.64**
A80								–	-.14
A124									–

*p < .05
**p < .01

B. **Description of Variables**

E135 = Raters' opinion of evaluation report quality
A53 = Auditor's opinion of general evaluation design
A57 = Auditor's opinion of evaluator's assessment technique
A67 = Auditor's opinion of accuracy of evaluation data
A68 = Auditor's opinion of evaluator's analysis of raw data
A70 = Auditor's opinion of evaluator's data interpretation
A71 = Auditor's opinion of format of evaluation report
A80 = Auditor's opinion of formative evaluation activities
A124 = Auditor's opinion of evaluation report quality

C. **Most Representative Variable**

MRV6 = A70

D. **Index**

Index 6 = (A53 + A57 + A67 + A68 + A70 + A71 + A80)
(Range of Index 6 is 0 to 7)

Intercorrelations. As is to be expected when ratings are intercorrelated, many of the correlations are significant. Guilford (1959) points out that

[3]Originally these variables were rated on a 7-point scale, -3 to +3; however, for the purposes of these analyses they were dichotomized at zero.

there is a rater-rating form interaction in which a rater tends to use a rating form in a similar fashion. Thus, a rater's judgment of different items tend to have a latent similarity. The variables with the most correlations were those representing the auditor's opinions of general evaluation design (A53), of the evaluator's analyses of data (A68), and of the interpretation of these analyses (A70). The absence of correlations with A80 (auditor's opinion of formative evaluation activities) is perhaps due to the fact that auditors did not specifically recognize or comment on this aspect of evaluations.

For this factor it was very difficult to select a most representative variable; there were a number of possible variables with nearly equal qualifications for selection. However, by using the bases for selection previously explicated, a selection was made. The variable selected as the MRV of the perceived quality of the evaluation report was Variable A70, the auditor's opinion of the evaluator's data interpretations.

Index 6 (A53 + A57 + A67 + A68 + A70 + A71 + A80). As previously noted, there were a number of variables highly correlated with many of the other variables in this factor. Variables A53, A57, A67, A68, A70, and A71 were all of this type and were included in Index 6. Despite its failure to correlate highly with other variables in this factor, A80 was included in the index because we considered it to be an important area. A high value on Index 6 indicates that, based upon the audit report, the evaluation was considered to be of high quality.

Factor 7: Evaluation Report Physical Attributes

Table 3.7 presents the data from which Index 7, evaluation report physical attributes, was derived.

Defining variables. One of the oldest tenets of western thinking is that mass is positively related to "goodness." We therefore sought to include a physical attribute factor, which included "mass," in this study. To that end we collected the following information: (1) New variable 8—number of main body pages in the evaluation report (ignoring appendices); (2) New variable 9—the presence of the evaluator's signature or (at the minimum) name in the evaluation report either on a title page or on an identification sheet; (3) New variable 10—the weight of the evaluation report. (Since all reports had been reproduced by the same company, paper weight was held constant. Thus, this variable may be thought of as "number of pages in the original report." Together these variables defined Factor 7.)

Intercorrelations. Intercorrelating these variables resulted in two out of three significant correlations. The weight of the evaluation report, NV10, was selected as most representative of the physical attribute factor.

Index 7 (NV8 + NV9 + NV10). The index for the physical attributes factor was simply the sum of all contributing variables. This weighted

heavily sheer mass or quantity. The higher the value of this index, the more there was of the evaluation report; and, having submitted a large report, apparently, the evaluator was more likely to put his name to it.

TABLE 3.7

Factor 7: Evaluation Report Physical Attributes

A. **Intercorrelations of Defining Variables (N = 42)**

Variable	NV8	NV9	NV10
NV8	–	.20	.89**
NV9		–	.36**
NV10			–

**$p < .01$

B. **Description of Defining Variables**
NV8 = Number of main body pages in the evaluation report
NV9 = Evaluator identified by name on the cover or title sheet of evaluation report
NV10 = Weight (oz.) of evaluation report

C. **Most Representative Variable**
MRV7 = NV10

D. **Index**
Index 7 = NV8 + NV9 + NV10
(Range of Index 7 is 0 to 3)

Intercorrelations among Factor Areas

Seven factor areas have now been defined from groups of variables derived from the Evaluation Data Sheets, the Audit Data Sheets, the Project Director Questionnaires, and several data items designated as new variables. Table 3.8 presents intercorrelations between the indices chosen to represent these factors. The highest correlation was between evaluation report physical attributes (Factor 7) and evaluation report comprehensiveness (Factor 4). This result is not surprising since a comprehensive evaluation is likely to be large in mass and in number of pages. The comprehensiveness factor correlated significantly with all the other factors except evaluation context (Factor 3). Evaluation context correlated significantly with no other factor.

The index for Factor 1 was composed of three variables reflecting project context. The significant correlations with comprehensiveness (Factor 4) and physical attributes (Factor 7) could be interpreted as indicating

that projects in large urban school districts have large, comprehensive evaluation reports. In part, one might argue that since the projects in these districts are larger, the comprehensiveness and mass of the report (Factor 7) will naturally be larger. One might further argue that more data and more comprehensive reports do not necessarily mean that the reports will be more favorable to the projects—in this instance large urban school projects. However, a significant correlation of .31 with project quality indicated a slight tendency for evaluators and auditors to perceive projects in large urban districts to be of higher quality than those in non-urban areas. This pattern is not changed when intercorrelations among the most representative variables are examined in Table 3.9. One might hypothesize that large districts have more highly qualified personnel available to conduct projects, have consulting resources geographically at hand, and have district office resources to call upon. This cannot be specifically sustantiated by these data but seems like a reasonable hypothesis for further investigation.

Projects with quality programs tended to have quality evaluations associated with them (.45); or, is it possibly the other way around, and that projects with quality evaluation reports tended to be considered as having quality programs? The question of whether projects of quality are wise enough to select quality evaluators or whether quality evaluators are able to shape and maintain a project's quality through their evaluations cannot be answered in this monograph. What is clear, however, is that a strong relationship exists and either (or both) are likely.

Quality evaluation reports, aside from the fact that they reported on projects of quality, exhibited comprehensiveness (.42) and were large in size (.33). Interestingly, they were not necessarily sophisticated (Factor 5).

TABLE 3.8

Intercorrelations among Indices Representing the Seven Factor Areas (N = 42)

Factor	1	2	3	4	5	6	7
1–Project Context	–	.31*	.15	.31*	.15	-.02	.25*
2–Project Quality		–	.11	.56**	.25*	.45**	.51**
3–Evaluation Context			–	.15	.10	-.02	-.01
4–Evaluation Report Comprehensiveness				–	.44**	.42**	.63**
5–Evaluation Report Sophistication					–	.04	.28*
6–Evaluation Report Quality						–	.33*
7–Evaluation Report Physical Attribute							–

*$p < .05$
**$p < .01$

A similar pattern of relationships between factors noted in Table 3.8 is present in Table 3.9. There are a greater number of significant correlations and they are, by and large, higher for the indices than between the MRV's. The relationship between project context and comprehensiveness and project quality are reaffirmed.

TABLE 3.9

Most Representative Variables from the Seven Factor Areas

A. **Intercorrelations of Most Representative Variables (N = 42)**

Factor	MRV	NV6	E133	PD3	NV3	E30	A70	NV10
1–Project Context	NV6	–	.34*	.07	.37**	-.07	-.03	.24
2–Project Quality	E133		–	.26*	.50**	.05	.30*	.58**
3–Evaluation Context	PD3			–	.25*	-.06	-.05	-.08
4–Evaluation Report Comprehensiveness	NV3				–	.27*	.18	.40**
5–Evaluation Report Sophistication	E30					–	.11	.22
6–Evaluation Report Quality	A70						–	.07
7–Evaluation Report Physical Attribute	NV10							–

*$p < .01$
**$p < .05$

B. **Description of Most Representative Variables**

Factor 1 = NV6 = Standard Metropolitan Statistical Area
Factor 2 = E133 = Evaluator's judgment of project
Factor 3 = PD3 = Internal or external evaluator as reported by project director
Factor 4 = NV3 = Variety of project components considered by evaluator
Factor 5 = E30 = Indication of progress evaluation activities
Factor 6 = A70 = Auditor's opinion of evaluator's data interpretation
Factor 7 = NV10 = Weight (oz.) of evaluation report

Again, project quality correlated well with the other factors. This factor, represented by Variable E133 (the evaluator's judgment of the project quality as estimated by CSE raters) was significantly correlated with all other MRV's except E30, progress evaluation reported, which was most representative of evaluation report sophistication (Factor 5).

The interpretation of the data related to Factor 3 offers some interesting thoughts. At first, one might be startled by the significant negative correlations to the MRV's of Factor 2 (project quality) and Factor 4 (evaluation report comprehensiveness) until the nature of the variable and its coding are considered. The most representative variable for Factor 3 (PD3) indicates whether the evaluation was performed by an internal or an external evaluator (using the project director as the data source). An internal evalua-

tor was coded as "0" and an external evaluator was coded as "1." Thus the negative correlations with this variable in Table 3.9 indicate that internal evaluators tended to be associated with better projects (Factor 2) and more comprehensive evaluations (Factor 4).

The evaluation report comprehensiveness (Factor 4) as represented by the variable NV3 (variety of project components considered by evaluator) was significantly related to all but one of the other variables. Interestingly, and inexplicably, evaluation report comprehensiveness and evaluation report quality were not significantly related.

Summary

In this chapter we have organized the variables described in Chapter II in terms of seven underlying constructs or intuitive factors. The factors were then investigated with respect to intercorrelations among the variables and an index was selected to represent each factor.

In the following chapter, these most representative variables and indices will be used as the independent variables to predict indicators of federal and local decision making. A most representative variable and an index for federal and local decision making will be proposed and these indicators of decision making will be regressed on the seven indices and the MRV's.

Chapter IV

THE IMPACT OF EVALUATION ON DECISION MAKING

In recent years the federal government has shown great concern for evaluation. Most federally funded programs currently mandate evaluations and offer increasingly detailed guidelines. Title VII projects are required to provide their federal sponsors with an evaluation and an independent educational accomplishment audit. For the purposes of this monograph, we had initially hypothesized that the information contained in these evaluation and audit reports might be used to guide federal decision makers in allocating continuation funds and that one might thus expect to find some evidence of the impact of evaluation information by reviewing federal funding decisions.

As noted earlier, however, we realized that relationships between funding decisions and final evaluation reports could have been attenuated in several ways:

- Funding decisions for the 1971–72 school year (FY72) had to be made before the 1970–71 final evaluation reports were received. However, interim evaluation reports would have been received and their content should have previewed, to some extent, the content of the final report.
- Projects doing poorly might receive more funds to encourage or make possible remedial action.
- Large increases in funds might sometimes reflect a change from a planning grant to an implementation of the project.
- Once a project has been funded at a certain level, funds are likely to be continued with little change for the duration of the project.

In addition to funding decisions we had a second indicator of federal perceptions—the ratings of projects by federal monitors. This variable also presented problems. Since the ratings were obtained from seven different federal monitors, they might not be strictly comparable. However, we purposely did not specify any criteria by which the federal monitors were to rate projects. Indeed, one of the interests in our study was to investigate and identify variables that are related to these federal judgments. The ratings simply represent the federal monitors' selection of three "outstanding," three "average," and three "below average" projects from those under their supervision.

Another important area in which evaluation might have decision-making impact is at the local project level. The responsibility for contracting and "organizing" federally mandated evaluations lies with the local project administration, that is, the local educational authority (LEA). The evalua-

tor is hired by the project, therefore, to prepare an evaluation of that project for the on-site staff as well as the federal sponsors.[1]

In this chapter we will explore the decision-making utility of evaluation information at both the federal and local levels. Paralleling our efforts in Chapter III, for each level of decision making (federal and local) an intuitive factor is developed. In the sections below we present the variables used to define these two factors, table of intercorrelations, the resulting indices, and most representative variables.

The Decision-Making Factors

Federal Decision-Making Factor

Table 4.1 reports the data from which the federal decision-making index was derived.

Defining Variables. As mentioned above, two indicators of federal decision making were available. First were the federal funds allotted each project for FY70, FY71, FY72, and second were the federal monitors' ratings of projects. Using the fiscal data, two additional variables were derived, the percent increase in funds from FY70 to FY71 (P70–71) and the percent increase from FY71 to FY72 (P71–72). Together these variables were used to define a new construct, the federal decision-making factor.

Intercorrelations. As was discussed in the derivation of Factor 3 in Chapter III, the almost perfect correlation between funding levels in FY71 and FY72 indicated that there was little in the way of differential funding among these projects in those fiscal years. The lower correlation between funding for FY70 and FY71 is perhaps due to different "start-up" costs among projects rather than refunding decisions based on project merit.

As shown by the correlation between funding levels and ratings (.36), there was a slight tendency for the higher-funded projects to receive more favorable ratings from project monitors. However, the correlations between rating and increase in funding from year to year were not significant, suggesting that federal funding decisions did not reflect the federal monitors' perceptions of project quality.

This situation raises two questions which we do not presume to answer: (1) What federal action reflects the monitor's opinions of project quality? and (2) How are refunding decisions made? However, in defense of our choice of dependent variables, it did seem reasonable to assume that fund-

[1]The position of the evaluator relative to the decision-making structure of the experimental program and its sponsors is a recurrent issue in evaluation literature (Alkin, 1972; Caro, 1971; Ferman, 1969; Agyris, 1958). In the case of a Title VII evaluator, it is possible to be hired by the district office to provide evaluative assistance at the local, district, and federal levels. The difficulties inherent in this situation are myriad. Different decision makers can have very different information needs. In particular, higher levels of decision makers (for example, federal) might require "rather sensitive" information on the effectiveness of lower level decision makers (for example, the project director). Such a situation could possibly lead to a rather tense relationship between a project director and the evaluator.

ing decisions would present tangible evidence of a sponsor's support for a project. It is a finding of considerable interest that, at least in the time period and sample represented by the study, our assumption was probably unjustified.

TABLE 4.1

Derivation of Federal Decision-Making Index

A. **Intercorrelations of Variables (N = 42)**

Variable	Rating	FY70	FY71	FY72	P70-71	P71-72
Rating	–	.20	.36**	.36**	.15	.18
FY70		–	.80**	.80**	-.32*	-.01
FY71			–	.99**	.21	-.22
FY72				–	.20	-.12
P70-71					–	.34*
P71-72						–

*$p < .05$
**$p < .01$

B. **Description of Variables**

Rating = Rating of projects by federal monitors
FY70 = Funding level for fiscal year 1970
FY71 = Funding level for fiscal year 1971
FY72 = Funding level for fiscal year 1972
P70-71= Percent increase in funds FY70 to FY71
P71-72= Percent increase in funds FY71 to FY72

C. **Most Representative Variable**

Rating

D. **Index**

Federal Decision-Making Index = 2 (Rating) + FY70 + FY71 + FY72 + P70-71 + P71-72
(Range of Federal Index is 0 to 13)

Rating was selected as the most representative variable for the federal decision-making factor. This variable directly reflects the federal monitors' judgments of quality and possesses variability (albeit, artificial, "forced" variance). Fiscal variables might have been more objective but their great homogeneity reduces their value for this study.

Federal Decision-Making Index = *2(Rating)* + *FY70* + *FY71* + *FY72* + *P70-71* + *P71-72*. All the variables associated with this factor were combined to form an index for the federal decision-making factor. To create this index several variables were weighted and/or recoded. Reflecting its status as the only nonfiscal measure, rating was given a double

weight. (In addition, this variable has the largest range, from 0 to 3.) Variables FY70, FY71, and FY72 were dichotomized at their means, and finally P70-71 and P71-72 were divided by 100 to give the fractional increase in funding from one year to the next. (The possible range for the P70-71 and P71-72 recoded variables was from 0 to 2 and the possible range for the entire index from 0 to 13.) A high value on this index indicates a high rating of project quality from the federal monitors, a high funding level, and a general increase in funds from one year to the next.

Local Decision-Making Factor

Table 4.2 reports the data from which the local decision-making index was derived.

Defining Variables. All the indicators of the usefulness of evaluation to local decision makers were derived from the Project Director Questionnaire. PD20a through PD20k queried the extent to which the evaluation was of assistance to the project director across various areas of project activity. (Although originally rated using a 5-point scale, these variables were dichotomized for this analysis as described in Appendix 2.) A general measure of usefulness to the project director, NV5, was derived by adding the sum of dichotomized PD20a-PD20k. Finally, PD22a and PD23a assessed the overall impact of formative evaluation and PD22b and PD23b the impact of summative evaluation.

Intercorrelations. In this local decision-making factor there were many significant correlations between the defining variables, indicating little discrimination among the various areas considered. When project directors reported the evaluation useful in one area, they generally reported it useful in all areas. The lack of significant correlations with variables PD20h (usefulness of evaluation to project director in changing personnel) and PD10 (extent of formative evaluation reported by project director) is not surprising in view of the lack of variance in the responses to these items. Only a few project directors reported no formative studies. Many project directors were emphatic that the evaluation activities were not a factor in personnel decisions and many directors reported no personnel changes.[2]

The high correlation (.75) between PD20a (usefulness of evaluation to project director in identifying possible problem areas) and PD20d (usefulness of evaluation to project director in recommending project changes) indicated that evaluators who were effective in locating problem areas tended to make useful recommendations. When such recommendations were made the project director tended to perceive the evaluation as influential in modifying the project the following year as indicated by a correla-

[2]The lack of use of evaluation for personnel decisions has interesting implications. Caro (1971) and Argyris (1958) discuss the threatening nature of evaluation to the project staff. Project directors certainly gave the impression that such a fear is unjustified in actual practice.

TABLE 4.2

Derivation of Local Decision-Making Index

A. **Intercorrelations of Variables (N = 42)**

	PD20a	PD20b	PD20c	PD20d	PD20e	PD20f	PD20g	PD20h	PD20i	PD20j	PD20k	NV5	PD22a	PD22b	PD23a	PD23b
PD20a	–	.63**	.53**	.75**	.48**	.35*	.46**	.20	.51**	.44**	.54**	.26*	.38**	.54**	.37**	.42**
PD20b		–	.30*	.38**	.15	.24	.22	.20	.29*	.22	.21	.20	.38**	.26*	.37**	.30*
PD20c			–	.65**	.46**	.45**	.48**	.22	.57**	.41**	.52**	.26*	.61**	.72**	.35*	.55**
PD20d				–	.59**	.46**	.46**	.20	.51**	.33*	.43**	.25*	.49**	.67**	.48**	.65**
PD20e					–	.47**	.52**	.34*	.56**	.45**	.52**	.31*	.39**	.48**	.51**	.46**
PD20f						–	.73**	.29*	.54**	.31*	.66**	.29*	.50**	.43**	.37*	.35**
PD20g							–	.21	.54**	.59**	.69**	.26*	.55**	.50**	.41**	.25*
PD20h								–	.39**	.13	.14	.01	.24	.16	.18	.08
PD20i									–	.39*	.58**	.34*	.50**	.41**	.37**	.36**
PD20j										–	.56**	.30*	.26*	.31*	.46**	.07
PD20k											–	.31*	.46**	.41**	.52**	.31*
NV5												–	-.18	.10	-.12	.13
PD22a													–	.65**	.61**	.47**
PD22b														–	.34*	.59**
PD23a															–	.43**
PD23b																–

*$p < .05$ **$p < .01$

B. **Description of Variables**

PD20a = Usefulness of evaluation to project director in identifying possible problem areas
PD20b = Usefulness of evaluation to project director in preparing reports
PD20c = Usefulness of evaluation to project director in developing criteria
PD20d = Usefulness of evaluation to project director in recommending project changes
PD20e = Usefulness of evaluation to project director in changing pace of instruction
PD20f = Usefulness of evaluation to project director in changing teaching procedures
PD20g = Usefulness of evaluation to project director in changing inservice activities
PD20h = Usefulness of evaluation to project director in changing personnel
PD20i = Usefulness of evaluation to project director in changing community relations activities
PD20j = Usefulness of evaluation to project director in changing program management
PD20k = Usefulness of evaluation to project director in changing instructional materials
NV5 = Sum of PD20a to PD20k. Usefulness of evaluation to project director
PD22a = Usefulness of evaluation for modifications during the year (formative)
PD22b = Usefulness of evaluation for modifications for the next year (summative)
PD23a = Influence of evaluation for decisions during the year (formative)
PD23b = Influence of evaluation for decisions for the next year (summative)

C. **Most Representative Variable**

PD20c
PD20i
PD22a
PD23b

D. **Index**

Local Decision-Making Index = PD20a + PD20k + 1/2(PD22a + PD22b + PD23a + PD23b)
(Range of Local Index is 0 to 4)

tion of .72 between PD20d (recommending for project changes) and PD22b (usefulness of the evaluation for modifications for the next year—summative). We asked project directors about two other potential sources of influence in questions 23c and 23d on the Project Director Questionnaire. Table 4.3 indicates the distribution of responses. There appears to have been a tendency for the project director to have found the final evaluation report more influential than either the recommendations of the auditor or of the federal monitor.

TABLE 4.3

Comparison of the Project Director's Views of Three Sources of Influence on the Project

Source	Very Influential	Influential	Somewhat Influential	Not Very Influential or Not at All	N
The *final evaluation report* for 70-71 influenced decisions that were made to modify the program the following year (71-72)	12	17	6	6	41
Recommendations of *federal monitors* influenced decisions to modify the program the following year (71-72)	8	9	9	15	41
Recommendations of *auditors* influenced decisions to modify the program the following year	8	18	11	5	42
	28	44	26	26	1

The high correlations with PD20c were surprising since little or no mention of criteria was made in the evaluation reports. Nevertheless, most project directors credited the evaluators with being of great assistance in developing criteria. We imagine that setting criteria was one of the more difficult problems faced by project management so that this was an area in which project directors were only too happy to assign responsibility to evaluators.

Four variables were selected as most representative of local decision making for the purposes of further investigations. These were PD20c (usefulness of evaluation to project director in developing criteria), PD20i (usefulness of evaluation to project director in changing community relations activities), PD22a ("formative" usefulness) and PD23b, ("summative" usefulness).

Local Decision-Making Index = *PD20a* + *PD20k* + ½(*PD22a* + *PD22b* + *PD23a* + *PD23b*). This index reflects the usefulness of evaluation information to the project director with respect to two specific areas; identifying possible problems (PD20a) and changing instructional materials (PD20k) and the general overall usefulness for making modifications either during the school year (PD22a and PD23a) or in the following school year (PD22b and PD23b). These latter two areas roughly reflect formative and summative purposes, respectively. A high value on this index indicates an evaluation that was generally perceived as useful for the project director.

Decision-Making Relationships

To study the effect of evaluative information on decision making, a series of regression analyses were conducted with either a decision-making index or a most representative variable (MRV) serving as the dependent variable and with the indices or MRV's developed in Chapter III serving as the independent variables. In view of the special character of Factor 7, two parallel regressions were conducted for each dependent variable. The first regression equation used the indices (or MRV's) representing the first six factors only, while the second equation also included Factor 7. All analyses were conducted using the multiple regression procedure in the Statistical Package for the Social Sciences (SPSS).

As has been previously noted throughout this monograph, we recognize some deficiencies associated with our variables. With respect to indicators of federal decision making, ratings obtained from seven different federal project monitors might not be strictly comparable, particularly in view of the vague criterion of overall project quality. Measures of federal funding displayed little variation, and since funds are frequently allotted before a final evaluation report is submitted, a direct relationship between the two is hard to postulate. With respect to indicators of local decision making, one might expect a stronger relationship between information contained in the evaluation report and the project director's opinions. However, the fact that these opinions were solicited one year later would tend to weaken this relationship.

In view of these caveats, we will not attempt to cross-validate prediction equations. Rather we will consider the varying pattern of relationships between the dependent and independent variables. By investigating which of these independent variables are major contributors to a regression equation, we can gain insights into how various characteristics of evaluations were related to different levels of decision making in this sample. The results of these analyses are presented in the tables below: Table 4.4 presents summary results from the regressions for federal decision making and 4.5 presents summary results from the regressions for local decision making.

TABLE 4.4 Regression Analyses: Federal Decision Making*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
	Simple Correlations with Dependent Variables							Step1				Stepwise Regression Step2				Step3				Regression Results Entering All Dependent Variables		
	Indices and MRVs for Factors 1-7								Mult.				Mult.				Mult.			Mult.		
Dependent Variable	1	2	3	4	5	6	7	Var.	R	R^2	F	Var.	R	ΔR^2	F	Var.	R	ΔR^2	F	R	R^2	F
Federal Index	.249*	.401*	.224	.336*	.491**	-.001	.215	Index for Factor 5	.491	.241	12.68**	Index for Factor 2	.568	.082	9.300**	Index for Factor 6	.594	.030	6.899**	.614	.377	2.993*
Rating	.239	.169	.058	.313*	.282*	.057	.058	MRV for Factor 4	.313	.098	4.339*	MRV for Factor 5	.374	.042	3.164	MRV for Factor 1	.412	.030	2.590	.458	.210	1.288

Notes: *$p < .05$
**$p < .01$

Dependent Variables: Federal Index = 2 (Rating) + FY70 + FY71 + FY72 + P70-71 + P71-72
Rating = rating of projects by federal monitors

Factor	MRV for Factor #
1. Project Context	1
2. Project Quality	2
3. Evaluation Context	3
4. Evaluation Report Comprehensiveness	4
5. Evaluation Report Sophistication	5
6. Evaluation Report Quality	6
7. Evaluation Report Physical Attributes	7

NV 6 : Standard Metropolitan Statistical Area
E 133 : Evaluator's judgment of project
PD 3 : Internal or external evaluator as reported by project director
NV 3 : Variety of project components considered by evaluator
E 30 : Indication of progress evaluation activities
A 70 : Auditor's opinion of evaluator's data interpretation
NV 10 : Weight (oz.) of evaluation report

*Regression of the federal index (dependent variable) on the indices (independent variables) are shown in row 1. Row 2 contains the regression of the ratings of the projects by federal monitors on the most representative variables (MRVs) of the factors.

Explanatory Notes

1. Each row represents the results of the regression analyses for a single dependent variable.
2. Column 1. For each factor a most representative variable and an index were developed. Column 1 indicates the dependent variables used in the regression analyses for each row. The first dependent variable is the index for the federal decision-making factor. The second dependent variable is the most representative variable for this factor, federal rating.
3. Columns 2-8. These columns give the simple correlations between the independent and dependent variables. The independent variables are representatives of Factors 1-7 developed in Chapter III. Note that when the dependent variable in column 1 is an "index" then columns 2-8 give the correlations between Indices 1-7 and the dependent variable. When column 1 identifies as dependent variable a most representative variable for the factor then columns 2-8 give the correlations between the MRVs for Factors 1-7 and the dependent variable.
4. Columns 9-20. Using all 7 factors a stepwise regression analysis was performed. Stepwise regression employs the following algorithm. Step 1: All the independent variables are correlated with the dependent variable, and the dependent variable is regressed on the independent variable with the highest simple correlation. Columns 9-12 report the results of this analysis. Column 9 reports the factor that is entered on the first step. As usual the appropriate representative of the factor is used (either an index or MRV). See note 2 above. Column 10 gives the "multiple" R for the equation. (This is the multiple regression's analogue of a simple correlation coefficient when a single independent variable is correlated with a dependent variable.) Column 11 gives R^2 (the square of the multiple R) which indicates the proportion of the total variance in the dependent variable that can be accounted for by the independent variables. Column 12 presents the F value for the regression equation. The null hypothesis claims that all of the b-weights in the regression equation are not significantly different from zero i.e., Given

$$y = b_o + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k.$$

$$H_o: b_1 = \ldots = b_k = 0$$

A significant F allows us to reject this null hypothesis, and we can conclude that at least one of the b-weights is different from 0. Note that significance at the .05 level is indicated by "*" and at the .01 level by "**."

Step 2: The partial correlations of the remaining independent variables with the dependent variable (removing the effect of the variable selected in Step 1) are compared. The independent variable with the highest partial correlation is then selected and together with the original independent variable is then used to predict the dependent variable. The results of this analysis are reported in columns 13-16. In row 1 (column 15), $\Delta R^2 = .082$. This represents a change in R^2 of .082 from step 1. This means that an additional 8.2% of the variance of the federal index can now be accounted for. The remaining steps follow a similar pattern. Partial correlations are computed and the highest partial correlate is selected. Columns 17-20 report the results of step 3.

5. The last 3 columns. These columns present the summary statistics for regression analyses in which the appropriate representatives of Factors or Indices 1-7 are used to predict the dependent variables. (In row 1, for example, the multiple R is .614, indicating that 37.70% of the total variance of the federal index can be accounted for by Indices 1-7. An F level of 2.993 indicates significance at the .05 level.)

TABLE 4.5 Regression Analyses: Local Decision Making

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
	Simple Correlations with Dependent Variables							Step 1				Stepwise Regression Step 2				Step 3				Regression Results Entering All Dependent Variables		
	Indices and MRVs for Factors 1-7								Mult.				Mult.				Mult.			Mult.		
Dependent Variable	1	2	3	4	5	6	7	Var.	R	R^2		Var.	R	ΔR^2	F	Var.	R	ΔR^2	F	R	R^2	F	F
Local Index	.254*	.343*	-.158	.104	-.107	-.001	.412**	Index for Factor 7	.412	.170	8.178**	Index for Factor 5	.471	.052	5.568**	Index for Factor 2	.506	.034	4.352**	.611	.373	2.889*	
PD20c	.036	.540**	-.347*	.374**	-.067	.274*	.473**	MRV for Factor 2	.539	.291	16.410**	MRV for Factor 3	.581	.046	9.922**	MRV for Factor 7	.621	.048	7.946**	.690	.476	4.420**	
PD20i	.034	.204	-.152	.149	-.088	.111	.278*	MRV for Factor 7	.278	.077	3.353	MRV for Factor 5	.316	.023	2.176	MRV for Factor 3	.346	.019	1.178	.384	.148	.841	
PD22a	.105	.483**	.175	.388**	.312*	.210	.390*	MRV for Factor 2	.483	.234	12.188**	MRV for Factor 5	.588	.112	10.306**	MRV for Factor 4	.651	.078	9.317**	.714	.509	5.037**	
PD23b	.073	.187	-.207	-.001	-.123	.342*	.290*	MRV for Factor 6	.342	.117	5.307*	MRV for Factor 7	.433	.071	4.517*	MRV for Factor 5	.481	.049	3.936*	.577	.332	2.422*	

Notes: $^*p < .05$

$^{**}p < .01$

Dependent Variables: Local Index = Local Decision Making Index = PD20a + PD20k + 1/2(PD22a + PD22b + PD23a + PD23b)—See Table 4.2

PD20c = Usefulness of evaluation to project director in developing criteria
PD20i = Usefulness of evaluation to project director in changing community relations activities
PD22a = Usefulness of evaluation for modifications during the year (formative)
PD23b = Influence of evaluation for decisions for the next year (summative)

Factor	MRV for Factor #	
1. Project Context	1	NV 6 : Standard Metropolitan Area
2. Project Quality	2	E 133 : Evaluator's judgment of project
3. Evaluation Context	3	PD 3 : Internal or external evaluator as reported by project director
4. Evaluation Report Comprehensiveness	4	NV 3 : Variety of project components considered by evaluator
5. Evaluation Report Sophistication	5	E 30 : Indication of progress evaluation activities
6. Evaluation Report Quality	6	A 70 : Auditor's opinion of evaluator's data interpretation
7. Evaluation Report Physical Attributes	7	NV 10 : Weight (oz.) of evaluation report

Explanatory Notes

1. Each row represents the results of the regression analyses for a single dependent variable.
2. Column 1. For each factor a most representative variable and an index were developed. Column 1 indicates the dependent variables used in the regression analyses for each row. The first dependent variable listed is the index for the local decision-making factor. The next four variables are the most representative variables selected to represent the factor.
3. Columns 2-8. These columns give the simple correlations between the independent and dependent variables. The independent variables are representative of Factors 1-7 developed in Chapter III. Note that when the dependent variable in column 1 is an "index" then columns 2-8 give the correlations between Indices 1-7 and the dependent variable. When column 1 identifies as dependent variable a most representative variable for the factor (as in rows 2-5) then columns 2-8 give the correlations between the MRVs for Factors 1-7 and the appropriate dependent variable (PD20c, PD20i, PD22a, PD23b).
4. Columns 9-20. Using all 7 factors a stepwise regression analysis was performed. Stepwise regression employs the following algorithm. Step 1: All the independent variables are correlated with the dependent variable, and the dependent variable is regressed on the independent variable with the highest simple correlation. Columns 9-12 report the results of this analysis. Column 9 reports the factor that is entered on the first step. As usual the appropriate representative of the factor is used (either an index or MRV). See Note 2 above. Column 10 gives the "multiple" R for the equation. (This is the multiple regression's analogue of a simple correlation coefficient when a single independent variable is correlated with a dependent variable.) Column 11 gives R^2 (the square of the multiple R) which indicates the proportion of the total variance in the dependent variable that can be accounted for by the independent variables. Column 12 presents the F value for the regression equation. The null hypothesis claims that all of the b weights in the regression equation are not significantly different from zero i.e., Given

(continued on following page)

TABLE 4.5 (Continued)

$$y = b_o + b_1 X_1 + b_2 X_2 = \ldots + b_k X_k$$
$$H_o : b_1 = \ldots = b_k = 0$$

A significant F allows us to reject this null hypothesis, and we can conclude that at least one of the b weights is different from 0. Note that significance at the .05 level is indicated by "*" and at the .01 level by "**."

Step 2: The partial correlations of the remaining independent variables with the dependent variable (removing the effect of the variable selected in Step 1) are compared. The independent variable with the highest partial correlation is then selected and together with the original independent variable is then used to predict the dependent variable. The results of this analysis are reported in columns 13-16. The remaining steps follow a similar pattern. Partial correlations are computed and the highest partial correlate is selected. Columns 17-20 report the results of step 3.

As an example, consider row 4, in which PD22a is regressed on the MRVs of Factors 1-7. The MRV of Factor 2, E133, is entered on the first step. The multiple R = 483 and 23.4% of the variance of PD40 can be accounted for by E133. The regression is significant at the .01 level.

On step 2 the MRV of Factor 5, E30, is added. With E132 and E30 in the equation the multiple R is now .588 and an additional 11.2% of the variance in PD40 can be accounted for. This new regression equation is still significant at the .01 level.

On step 3, the MRV of Factor 4, NV3, is added to E132 yielding a multiple R of .651 and 42.4% of the variance accounted for. Once again the equation is significant at the .01 level.

5. The last 3 columns. These columns present the summary statistics for regression analyses in which the appropriate representatives of Factors 1-7 are used to predict the dependent variables. [In row 4, for example, the multiple R is .714 indicating that 50.90% of the total variance of PD22a—usefulness of evaluation for modification during the year (formative)—can be accounted for by indices 1-7. An F level of 5.037 indicates significance at the .01 level.]

Federal Decision Making

Simple Correlations

Inspection of the simple correlations in Table 4.4 reveals that the factor indices (row 1) exhibit stronger relationship to the federal index than do the *MRV's* (row 2) to the project monitor ratings. This is readily explained in terms of the composition of the indices. The federal index consisted of two components (rating and fiscal variables) which were not significantly correlated. Consequently the federal index had a greater chance of sharing variance with other measures than did rating alone. In general, the indices were much broader representatives of each factor and were therefore expected to show higher intercorrelations.

The indices for Factor 5 (evaluation report sophistication) and Factor 2 (project quality) showed the highest correlations with the federal index (.49 and .40 respectively). Thus, insofar as rating and the fiscal variables reflected federal judgment of the project, the best predictors of this judgment in our sample were the presence of a sophisticated evaluation report (characterized by the inclusion of implementation and progress data, hypothesis-testing, and a variety of assessment instruments) and evaluation and audit reports that seemed favorably disposed to the project. Factor 1 (project context), which reflected size and urbanness of the program, and Factor 4 (evaluation report comprehensiveness) also had a significant positive correlation with the federal index.

Considering now the federal MRV (rating), we again found Factor 4 (evaluation report comprehensiveness) and Factor 5 (evaluation report sophistication) to be significant correlates. Apparently, in this sample at least, the scope and technical *quality of the evaluation report* were stable correlates of federal judgments of the *quality of the project.*

Taking a gestalt look at these results one could postulate that projects in urban areas were large, had high funding levels, could employ specialists as evaluators who produced sophisticated and comprehensive evaluation reports, and tended to favorably impress the federal monitors.

Regression Equations

Dependent Variable: Federal Index. Reviewing Table 4.4, we see that Factor 5 (Index 5, evaluation report sophistication) was entered on the first step of this regression analysis and accounted for a substantial proportion (24%) of the total variation in the federal index.[3] With the addition of Factor 2 (project quality) in the second step, an additional 8.2% of the variance was accounted for. Thus evaluation sophistication as measured by the inclusion of formative (implementation and progress) evaluation, the use of hypothesis testing and a variety of assessment instruments, as

[3]Recall that the variable exhibiting the highest simple correlation with the dependent variable is always entered on the first step.

well as judgments of projects' quality in the evaluation and audit reports, were estimated to be the best two predictors, in combination, of the federal index of project effectiveness.

Regression equations were developed with Factors 1 through 7 as independent variables. With all seven factors entered into the equation, the resulting regression equation was significant at the .05 level with a multiple R equal to .614 and R^2 equal to .377, meaning that 37.7% of the variance in the federal index was "accounted for."

Dependent Variable: Rating. It is clear from inspection of Table 4.4 and previous discussion that this regression analysis was not particularly successful in accounting for variation in the single dependent variable, rating of projects by federal monitors.

On the first step the MRV for Factor 4 (evaluation report comprehensiveness) was entered, and although the associated regression equation for this step was significant, only 9.8% of the variance in the federal rating could be explained. The MRV for Factor 5 was entered on the second step. However, its contribution to the prediction profile was statistically insignificant. Only an additional 4.2% of the variance was accounted for, and the associated regression equation was not significant. Using all seven MRV's in the analysis the prediction equation was not significant with a multiple R of .458, 21% of the total variance being thus "accounted for." As was the case with the federal index, little change in prediction was produced by including MRV7 into the analysis after the other six.

Local Decision Making

The dependent variables in the following analyses were all derived from the Project Director Questionnaire responses. They generally represented the project director's responses to questions concerning the evaluation's usefulness to him in various areas, such as setting criteria or making program modifications. As above, the independent variables were the indices or MRV's from the seven factors.

Simple Correlations

In considering the simple correlations in Table 4.5, we attempted to avoid over-interpreting chance correlations by attending to recurring patterns of association between the dependent and independent variables. Several of these patterns emerged. Factor 7 (evaluation report physical attributes) was significantly correlated with each dependent variable. Apparently, in considering the utility of evaluation to the project director the weight of the evaluation report was a good "predictor." In addition, Factor 2 (project quality) displayed three significant relationships: with the local decision-making index, with PD20c (usefulness of evaluation to project director in developing criteria) and with PD22a (usefulness of

evaluation for modifications during the year—formative). It would appear then that the sheer amount of evaluation information, along with overall judgments of project quality from the evaluation and audit reports, were the variables most closely linked to indications of evaluation usefulness at the local level. These results contrasted sharply with the pattern of relationships among federal decision makers. Whereas evidence of evaluation sophistication and comprehensiveness (with the addition of judgments of project quality in the case of the federal index) were associated with *federal* decision making, *local* decision makers were influenced by the physical amount of evaluation information produced and general judgments of project quality. With respect to local decision making, the rule seemed to be: the more evaluation the better and the more positive the assessment the better.

This observation should not be taken lightly. At the present stage in the development of evaluation practice, differences in the kinds of experimental designs, measurement tools, and analytic methods employed by evaluators are not as crucial to local decision makers as general judgments of project quality and the mass of evaluative information provided to them. Several explanations are possible:

1. All evaluations were performed with *equal* competence and therefore quantity became the only dimension on which evaluation reports differed.
2. The state of the art in evaluation was rather low in 1971. Consequently, project directors' major concern was simply with gathering the needed information rather than relying on evaluators for inferences and recommendations.
3. Project directors were naive with respect to evaluation and were not sure how to use or judge the resulting information. Evaluation is a recent phenomenon in the social sciences and many professionals have not yet been trained in the proper use of evaluation information. In the absence of other criteria, project directors relied on traditional measures of quality, such as report size.
4. The significant relation between the perceived usefulness of the evaluation and project quality may have measured the relationship between the evaluator and the project director. Where the evaluator had a positive attitude towards the project, the project director was more inclined to accept the advice and counsel of the evaluator, and therefore perceived the evaluator as more useful to him.

At the present time any attempt at responding to these hypotheses is quite heuristic and provisional; keeping in mind these caveats we now offer some tentative interpretations.

Based on experience acquired in the course of this study, we must reject the first hypothesis. All evaluators did not produce reports of equal quality;

neither were all evaluation efforts equally competent. (To the contrary, we had not anticipated such great diversity in evaluation efforts!) This conclusion was further substantiated by the auditors' rather varied judgments of evaluation report quality. (Of course, variance among auditors is involved here too—that is, even if all auditors were looking at the same report, there would be variance.)

Hypotheses 2 and 3 provided more plausible explanations (particularly with respect to Factor 7). Evaluation is not a well-developed art, and the Title VII reports reviewed for this study reflected many of the field's deficiencies. For example, evaluation reports frequently lacked coverage of important areas of concern to the target projects (e.g., only 10 evaluators considered staff relations in their final report and only eight attended to unexpected cognitive outcomes).

In addition, it was possible that project directors (local decision makers) had not made maximal use of the evaluation information provided to them, most likely because it was an entirely new enterprise for them. Indeed, evaluation was simply mandated by the federal government and aside from a manual providing guidelines for evaluators, no training was given to the project directors. It seems reasonable, then, that the local project administration might initially view evaluation simply as another funding stipulation from their sponsors and, consequently, be satisfied with seemingly substantive reports (that is, large reports) and favorable reports of project quality.

Finally, we suspect that the fourth hypothesis, suggesting that the evaluator's opinion of the project was a critical element in determining an evaluation's usefulness to the local decision maker, is also relevant. In a very real sense the evaluator is a federal agent, one of whose duties is to judge the local project. Thus, the evaluator can easily be perceived as both formidable and threatening. However, when the evaluator shows a positive attitude towards the project, the local decision maker's fears diminish and he is more receptive to evaluative assistance in his own decision-making activities.[4] On the other hand, when the evaluator presents a negative attitude towards the project, the local decision maker may regard him simply as a thorn in his side and consequently hesitate to share information with him or to utilize his experience.

It is quite possible to interpret this close association between the local decision maker's perceptions of evaluation utility and the evaluator's opinion of the project in either direction. That is, not only was the project director more likely to assess the evaluator's usefulness positively when the evaluator was favorably disposed towards the project, but also, as

[4]Note that we refer here to the evaluator's general attitudes towards the project. Quite frequently evaluations offered very positive global judgments and yet made extensive criticism of specific project areas.

the evaluator became involved in the project in the course of producing information useful to the project director, this involvement was then reflected in his favorable judgment of project quality.

In summary, the close relationship among the project director's judgment of evaluation usefulness (local decision-making index), the opinions of project quality (Factor 2), and the pure objective factor (Factor 7) can be explained by (a) the desire by project directors simply to obtain needed information rather than to concern themselves with quality and (b) the project director's tendency to trust the counsel of an evaluator who on the whole displayed a favorable stance towards the project (even if the small print was a bit critical).

Regression Equations

Dependent Variable: Local Decision-Making Index. On the first step of this analysis, Factor 7 (Index 7, evaluation report physical attributes) was entered accounting for 17% of the variance in the local decision-making index. On step 2, a total of 22.2% of the total variance was explained when Factor 5 (Index 5, evaluation report sophistication) was added to the prediction equation. Thus if an index is the broadest representative of a factor, the amount of evaluative information generated (as measured by the weight of and the presence of a signature on the final evaluation report) moderated by evaluation report sophistication (as measured by the presence of formative evaluation, hypothesis testing, and a variety of assessment instruments) was most intimately related to the utility of evaluation information to the local decision maker (the project director).

With all seven factors entered, the regression equation (significant at the .05 level) had a multiple R equal to .611 and 37.3% of the total variance was accounted for. However, when only the first six factors were considered in the analysis, the resulting equation was not significant. Thus in contrast to the federal decision-making analysis, Factor 7 (Index 7, evaluation report physical attributes) here made a substantial contribution (11.4% of variance) to predicting the local decision-making index over and above the effects of Factors 1 through 6.

Dependent Variable: PD20c (usefulness of evaluation to project director in developing criteria). On the first step the MRV for Factor 2 (project quality) was entered accounting for 29.1% of the total variation. Apparently PD20c, the usefulness of the evaluation to the project director in developing criteria, was highly related ($r=.54$) to MRV2, an estimate of the evaluator's judgment of project quality. This exceptionally strong relationship, relative to other results, might be attributed to the fact that when a project was well-defined, the evaluator was able to devote time to developing criteria for success, whereas in less organized projects the

evaluator's time was consumed with deciphering and clarifying objectives and formalizing a plan to guide his evaluation. On the second step, MRV3 (representing evaluation context) was added to MRV2, and although the regression was still significant, only an additional 4.6% of the variance in PD20c was accounted for.

Regression on all seven MRV's produced an equation significant at the .01 level with a multiple R equal to .690 and accounting for 47.6% of the dependent variable's variance. Contrasting this regression result with the analysis excluding Factor 7, we see that an additional 6.9% of the total variance was predicted by the inclusion of Factor 7. Thus once again Factor 7 was able to explain a sizeable portion of the variance in the dependent variable that was not accounted for by another factor, a fact reflecting the felt importance of voluminous reports to the local decision maker.

Dependent Variable: PD20i (usefulness of evaluation to the project director in changing community relations activities). A non-significant regression equation characterized this regression analysis at each step. Hence it seemed that we were unable to account adequately for PD20i (usefulness of evaluation to the project director in changing community relations activities). Regarding the small amount of variation that was "predicted," the weight of the evaluation report (Factor 7) and the sophistication of the evaluation as measured by the inclusion of progress evaluation (Factor 5) appeared to be the most important independent variables. Factor 7 (evaluation report physical attributes) was the only variable with a significant simple correlation with PD20i.

Dependent Variable: PD22a (usefulness of evaluation for modifications during the year—formative; that is the project director's opinion of the usefulness of formative evaluation). Factor 2 (MRV2, representing project quality), the independent variable displaying the highest correlation with PD22a, was entered on the first step of the analysis yielding a significant equation, with multiple R equal to .483 and 23.4% of the variance accounted for. On step 2, Factor 5 (MRV5, representing evaluation report sophistication) was added which explained an additional 11.2% of the variance.[5] On step 3, Factor 4 (MRV4, representing evaluation report comprehensiveness) was entered and together with Factors 2 and 5 another 7.8% of the variation in PD22a predicted. Thus, the usefulness of formative

[5]A significant relationship between MRV5, the inclusion of progress evaluation activities representing Factor 5 (evaluation report sophistication) and the dependent variable, PD22a (the usefulness of formative evaluation) is not surprising as progress evaluation is a fundamental part of formative activities. (CSE defines formative evaluation to be the sum of implementation and progress evaluation.) In fact one might expect an even higher correlation. However, this was not the case as all but three project directors reported the presence of some formative evaluation thereby reducing the discriminating powers of this variable. Nonetheless, this relationship should not be considered trivial as MRV2 is representative of the entire evaluation sophistication factor.

evaluation for the local decision maker (PD22a) was positively related to the evaluation's overall quality as judged by the evaluator in his report, the sophistication of the evaluation as judged by the inclusion of progress evaluation, and the comprehensiveness of the evaluation reports as indicated by the variety of project components considered in the report.

The absence of Factor 7 in the first 3 steps of this stepwise regression, despite its relatively high simple correlation with the dependent variable, can be attributed to the significant correlations between the MRV for Factor 7 and MRV2 ($r = .57$) and MRV5 ($r = .40$). The variance that these independent variables shared in common appears to have been shared with the dependent variable.

This equation presented the strongest prediction profile. At each step the regression equations were significant at the .01 level. With all seven factors entered, the multiple R was .714 and 50.9% of the variance in the dependent variable was accounted for. Comparing the regression results using Factors 1 through 6 with those using Factors 1 through 7, little predictability was added by the inclusion of Factor 7 ($R^2 = 4.7\%$). However, this situation was not surprising in light of the above discussion.

Dependent Variable: PD23b (influence of evaluation for decisions for the next year—summative). This variable represented the usefulness of summative evaluation to the project director (who used it formatively!). On step 1, Factor 6 (MRV6, evaluation report quality) was entered, yielding a significant (at the .05 level) regression equation accounting for 11.7% of the total variance. Factor 7 (MRV7, representing the evaluation report's physical attributes or weight) was then added to Factor 6 on step 2, contributing an additional 7.1% to the prediction of the variance in PD23b. Thus the utility of the outcome evaluation report was most influenced by the quality of the evaluation report (as indicated by the auditor's opinion of the evaluator's interpretation of his analyses, MRV6) and the weight of the evaluation report (MRV7).

Comparing the analyses predicting the impact of formative (PD22a) and summative (PD23b) evaluation on the local decision maker, we see that the evaluator's opinion of the project and the sophistication of the evaluation efforts were closely linked to formative evaluation's usefulness, while evaluation report quality and the sheer amount of evaluation information provided were linked to summative evaluation's usefulness. These observations might reflect the following events. During formative evaluation the evaluator's attention is directed almost exclusively to the local project in order to detect areas requiring modification. Thus the regression analysis reflects the value of the evaluator's opinion of project quality and the sophistication of the evaluation effort. On the other hand, the project director is responsible to the federal sponsors for outcome evaluation and making decisions for the next year and thus is most affected by the *auditor's* opinion of evaluation report quality and the size of the

report! In addition the project director, who may have based many of his decisions throughout the year on evaluation information and perhaps is planning further changes for the following year, may be anxious to have the quality of this information validated by an independent audit team.

With all seven factors entered into the regression analysis the equation was significant (at the .05 level), the multiple R was equal to .577, and 33.2% of the variance accounted for. When only Factors 1 through 6 were entered, the regression equation was not significant and only 20.8% of the variance was accounted for.

Caveat

At his point we must, once again, take note of the instability in our data. In Table 4.6 we present 95% confidence limits for selected values of correlation coefficients based upon N=42.

TABLE 4.6

Ninety-five Percent Confidence Limits for N = 42

r	95% Confidence Limits
.30	.00 to .55
.40	.11 to .63
.50	.23 to .70
.60	.35 to .76
.70	.50 to .83

Summary

This chapter has examined the decision-making utility of evaluation information at federal and local levels. For each of these decision-making levels an intuitive factor was developed. The variables that were used to define these factors, as well as tables of intercorrelations, indices, and most relevant variables have also been presented. Figure 4.1 provides a graphical visualization of the summary results of the stepwise regression analyses.

Key to Figure 4.1: Summary Results of Stepwise Regression Analyses*

KEY
to
graphical representation
of stepwise regression
results

FACTORS
1–project context
2–project quality
3–evaluation context
4–evaluation report comprehensiveness
5–evaluation report sophistication
6–evaluation report quality
7–evaluation report physical attributes

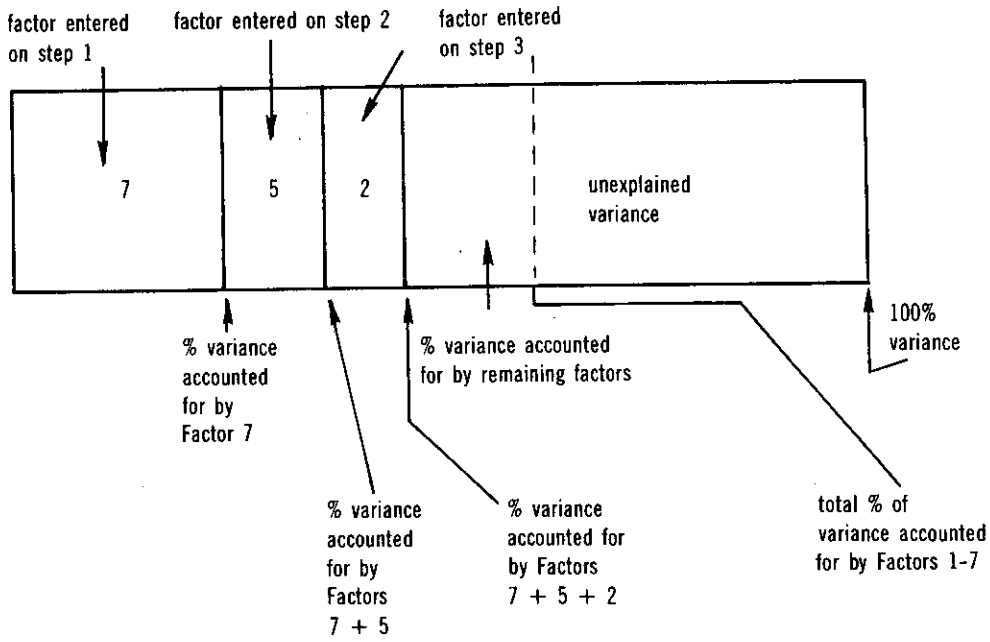
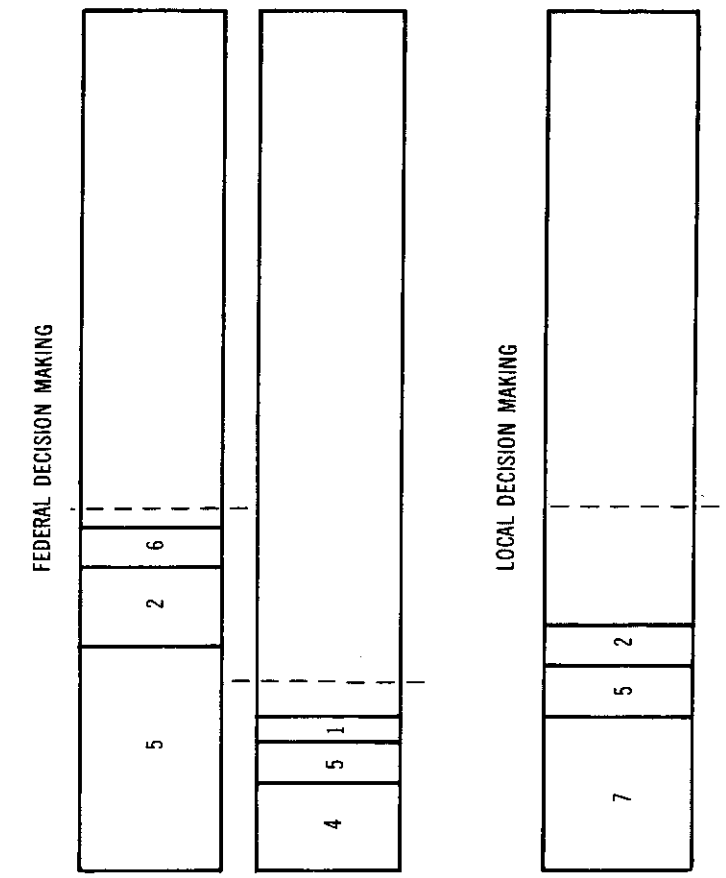
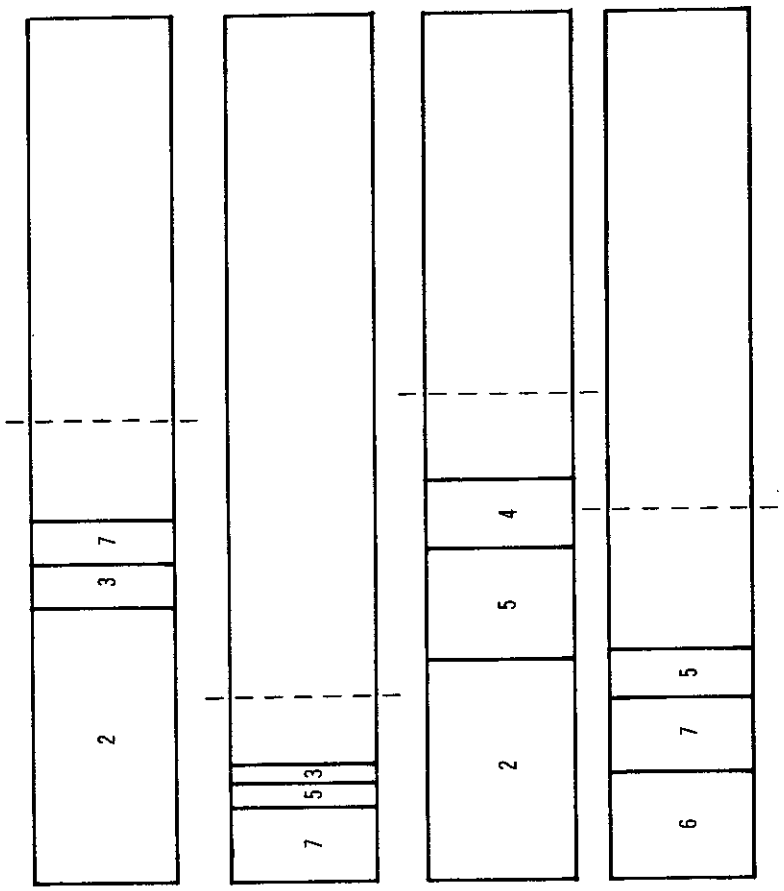


Figure 4.1 appears on the following pages.

FIGURE 4.1

Summary Results of Stepwise Regression Analyses




PD20 c (Criteria)
MRVs (1-7)
2
3
7
PD20i (Community Relations)
MRVs (1-7)
7
5
3
PD22a (Formative Influence)
MRVs (1-7)
2
5
4
PD23b (Summative Influence)
MRVs (1-7)
6
7
5

Chapter V

EVALUATION IN THEORY AND PRACTICE

In this monograph we have suggested that the purpose of evaluation is the provision of timely, useful information. A necessary condition to insure this timeliness and utility is that the decision context of the evaluation be clearly identified. This decision context can be considered to consist of: (1) the relevant decision maker(s), (2) the decisions to be made, and (3) the socio-political setting of the target project (Alkin & Kosecoff, 1973).

Components of the Decision Context

Innovative projects and, in particular, social action projects, are generally blessed with a multitude of decision makers; in Title VII projects there are the project director and his staff, the LEA supervisory and administrative personnel, the federal project monitors, and other federal officers who participate in making the ultimate refunding decisions. The evaluator faced with this multiplicity of decision makers can identify those persons to whom he is responsible by considering the second component of the decision context—the decisions to be made. One of the major determinants of these decisions is the developmental stage of the project. In a situation in which the project is being organized and implemented and its operations are being established and routinized, these functions will in large part determine the critical decisions for which evaluation data are required. Decisions during early stages should relate primarily to needed modifications in the project. This is the responsibility of the local project staff who are therefore the decision makers of concern during development of the project. On the other hand, decisions to be made about a mature, stable, replicable project relate to the project's future use and therefore are the responsibility of present and future sponsors. (In the case of Title VII, these may be district, state, or federal officials.)

The final component of the decision context requires the evaluator to be aware of the social and political realities within which the project must function. (This will hardly seem like a startling discovery to those who have had their evaluation skills tempered on the firing line of real-world evaluation.)

Theoretically, the necessity of clearly defining the decision context appears incontestable. A major question addressed in this chapter is the extent to which this necessity is borne out in practice. While we have not researched all aspects of the decision-context question, we nonetheless can draw considerable insights about the decision context. We will structure

our analysis by using as a guiding framework the general issue of the formative-summative nature of an evaluation effort.

Formative and Summative Evaluation

Formative and summative are perhaps the most widely employed terms in the evaluative literature. As previously discussed, the terms can be distinguished by reference to the different decision makers whose needs are served by each kind of evaluation. First introduced by Scriven (1967, p. 43), formative evaluation is evaluation of educational programs while they are still "fluid;" that is, while a program is considered to be in some stage of development. Formative evaluation serves the decision needs of the developer or manager of a project. Summative evaluation, on the other hand, serves the decision needs of the user. It aims to facilitate sound judgments about the project as it exists, judgments such as whether or not it should be continued, how it is best used, whether the project meets the needs of various groups, etc.

Is the kind of evaluation that he is called upon to perform of any practical concern to the evaluator? Can he not simply produce data on how the project is operating and how well it is doing and let others decide how to use the information in their decisions? We shall argue that to attempt to perform both a formative and a summative evaluation is conceptually schizophrenic, and leads in practice to role confusion and inadequacy. We shall now consider several consequences—both theoretical (as documented in the evaluation literature) and practical (as evidenced in our study)—of the formative-summative distinction. The consequences to be considered are: (1) the conception of the project, (2) the role of the evaluator, (3) evaluation techniques, and (4) the utility of evaluation information. Each of these will be examined from the point of view of the critical distinctions between formative and summative evaluation(s).

The Conception of the Project

Consider first the conception of the project that the formative evaluator must maintain. In the beginning the project exists as a plan. The formative evaluator must assist the developer as the project makes the transition from a paper-and-pencil blueprint to a well-defined process. To that end, the formative evaluator must not only assess progress toward the realization of educational goals and objectives but must also provide evaluative feedback on the extent to which the project has been implemented and is operating as planned. This assistance involves careful monitoring of the project, discovering and drawing attention to deficiencies, and recommending changes throughout the formative stages of project development. During formative evaluation the project is thus conceptualized as fluid, protean, and undefined.

In contrast, the summative evaluator conceives of a project as finished, well-defined, and replicable, with effects that can be assessed. It is the summative evaluator's task to certify the project as a completed entity to its sponsor and its potential users. This task will include comparing the target project's qualities with those of other established projects, clarifying and documenting the outcomes that can be produced, exploring the generalizability and transportability of components, and recommending how and when the project should be implemented.

The Role of the Evaluator

Second, consider the role of the evaluator vis-à-vis the project staff. In helping the project to develop and mature, the formative evaluator must serve primarily the decision needs of the developers (the project director and his staff in the case of Title VII projects). He works closely with the project staff, providing frequent recommendations for improvements. As the project develops in response to his suggestions, the formative evaluator will tend to identify with it and view its success as a personal accomplishment. It is naturally difficult for the formative evaluator to then don a summative hat and provide an unbiased critique of his own efforts.

The summative evaluator, on the other hand, assumes an external role with respect to the project and its staff. In assessing project merit and serving primarily the needs of potential users and funding agencies, the summative evaluator must maintain an objective, unbiased perspective.

The failure to distinguish and separate formative from summative evaluation, therefore, leads to a schizophrenic conceptualization of the project as both finished and unfinished, static and fluid. The evaluator who attempts to fulfill both formative and summative evaluation activities is in a position of trying to serve several masters.[1]

Evidence of Role Conflict in Title VII. Thus far we have presented the notion of formative and summative evaluation roles and have sought to establish the need to make a clear distinction between these evaluation types. We now ask if evidence of these problems can be found in Title VII and, if so, what the implications are. We shall suggest that it was failure of the federal guidelines to define the decision context that caused the formative-summative role conflict in Title VII.

Although recognition of the distinction between formative and summative evaluation roles has existed for some time, this issue seems to be overlooked in the planning and mandating of Title VII evaluation (a situa-

[1]The problem of serving multiple decision makers has been long debated in the literature. The most definitive position on this issue has been taken by Matthew who warns, "No man can serve two masters for either he will hate the one and love the other, or else he will stand by the one and despise the other." *Matthew VI, 24.*

tion which, we suspect, simply reflects current practice). In providing guidelines for the evaluator, the Title VII Manual (HEW, 1971) appears to mandate both formative and summative evaluation. With respect to formative activities, the Manual states that the evaluation should "permit timely revision of plans so operational deficiencies may be corrected" (p. 50). Clearly, this directive leads to formative evaluation and presupposes a project in its development stages. However, a few sentences later the Manual directs the evaluator to summative activities. "Moreover, the information provided through evaluation must enable staff in the funding agencies to identify which projects warrant continuation as models that can be replicated" (p. 50). In this passage a summative evaluation is clearly founded on the assumption that the project is sufficiently well formed to serve as a replicable model.

The conflicting requests for formative and summative evaluation created serious problems for Title VII evaluators; the incompatability of these tasks was evident both in their approaches to evaluation and the content of their final reports. Some evaluators accepted the Manual's second directive as binding and consequently attempted summative evaluation. In the following quotation from an evaluation report, the formative-summative conflict has been forthrightly documented by an evaluator who has decided, apparently without any qualms, that he is a summative evaluator only:

> The interim project director felt that the pre-test data might be used to substantiate the impression of the teachers that some of the children in the bilingual classroom were inappropriately placed there in that they showed little initial understanding of the language. Such corroboration could be used to switch the children to the traditional English classroom. The interim project director insisted that, under the circumstances, it was legitimate to use the data in this manner since to do so would conform to the need for the evaluator to provide immediate feedback. The evaluator explained that pre-test data, originally gathered for the purpose of *external* program evaluation, would have its usefulness vitiated were it used for making internal program changes.

Other evaluators, on the other hand, accepted the Manual's first directive, and consequently attended exclusively to formative activities. This stance seems particularly reasonable in view of the developmental nature of the Title VII projects. Of the 42 projects included in this study, less than half were in their second year of operation; the remainder were entirely new.

A second quotation from an evaluation report indicates how another evaluator confronted the problem of a formative or summative role and interpreted his mandate as "formative."

> Initially, the evaluator conceived his role in a way consistent with his background in Social Psychology. He saw his role as one providing an *external* assessment, by a disinterested party, of the kinds of fundamental changes expected to take place in children's language capabilities as a result of their experience in a bilingual classroom. As such, the initial

> intent was for the evaluator to maintain a nonmanagerial perspective to guarantee that his activities would not interfere with the very process he intended to measure. Under the requirements outlined under Title VII guidelines, the role of the evaluator is specified to operate in a manner different from the one he initially conceived. This plan called for the evaluator to become part of the program management team, concentrating his efforts on the internal processes of program functioning, rather than solely on the outcomes of the program as judged from pre- and posttest measurement provided at the beginning and end of the school year.

Evaluator's Relationship to the Project. The evaluator's affiliation with the target project has been a frequent subject of inquiry in the evaluation literature. Of particular concern is the evaluator's position as an insider or an outsider. Internal evaluation occurs when the evaluator is a member of the target agency. For example, regular personnel (i.e., teachers) or special evaluation units controlled by the agency receiving the funding (i.e., a school district's research and evaluation group) assigned to evaluation duties for the target project are considered internal. External evaluators, on the other hand, are not part of the target agency but rather are commissioned to evaluate the target project.

The literature leads one to expect significant differences between external and internal evaluators. It is generally suggested that the internal evaluator, while more sensitive to the project and its staff, is subject to pressure from the project director and his immediate superiors in the target organization. The internal evaluator's job, his security, and his critical eye can be influenced by this dependence. The external evaluator, though less sensitive to the unique character of the project, is free from this kind of pressure and therefore might be in a better position to give a critical appraisal of the project's worth (Caro, 1971). Moreover, an external evaluation is felt to be more likely to bring with it prestige that will lead to findings and recommendations receiving full attention.

This situation, however, was not observed in our analysis of Title VII projects. Although the majority of project directors preferred an internal evaluation team (only one project director indicated preference for a completely external evaluation), most project directors explained that this reflected their concern with maximizing the amount of time an evaluator was on-site. Apparently trouble-shooting and quick feedback were qualities highly valued by the project directors contacted in this study. The external-internal distinction drawn so carefully in the literature regarding the prestige and objectivity of the evaluator apparently was not a practical concern in the sample of Title VII projects studied.

Once again this result might reflect the emphasis on formative evaluation. In formative evaluation the focus is on gathering frequent information for project modifications—modifications to be made as quickly as possible while the project is still developing. The most suitable candidate for an

evaluator, therefore, would be someone very familiar with the project, its staff, and its goals—someone who would be present at all times and therefore would be in a position to detect potential problems before they mature—in short, an internal evaluator or an external evaluator who is both conceptually and physically close to the project. In contrast, it might be expected that in a summative situation designed to validate a project for the user, an external evaluator's objective assessment would be most appropriate.

In contrast to the severe distinctions drawn in the literature, the nature of the evaluator's efforts and his relationship with the project staff did not seem to be predicated on his internal or external status. Our data suggest some explanation for the similar roles adopted by internal and external evaluators. In many Title VII projects the evaluator might have suspected that he owed his present job largely to the project director and could well be similarly indebted in the future. The most typical procedure for selecting evaluators (31% of cases) relied on a committee that included the project director. In almost one-fifth of the projects, the project director alone chose the evaluator. In the case of an internal evaluator (38% of projects) the evaluator and project director might realistically have anticipated a close working relationship for some years to come and consequently have been at pains to assist each other. Furthermore, 43% of the project evaluators were located by "recommendations" and 24% had been known from previous professional association. These facts imply that evaluators are well advised to be likeable if they want to work.

Although evaluation is federally mandated and evaluation guidelines are provided by federal sources, the evaluator's real contact with Title VII is through the project director and the local project staff. The evaluator is hired by the project staff and paid with "their" funds. Further, he must work with the project director and his staff to organize his evaluation and to collect the needed information. The evaluator's findings and recommendations are directed to the local project.

These observations, considered in conjunction with the project's early stage of development, might also account for another contrast that we noted between our findings and the evaluation literature, namely, the very positive evaluator-project director relationships.

Evaluator's Relationship to the Project Director. A reading of the evaluation literature leaves one with an impression that evaluators are less than beloved:

> . . . . what the staff and participants of the evaluated program are faced with is the presence of a team of outsiders (the evaluators) poking around the program for awhile. . . (Bend, 1970, p. 117).

> Trapped administrators have so committed themselves in advance to the efficacy of the reform that they cannot afford honest evaluation (Campbell, 1969, p. 428).

Policy has been formed without considering the kinds of evaluation data needed to sustain the worth of a program. Objective evidence of the effects of programs has not been demanded as a basis for modifying programs. Satisfied with informal evaluation, administrators have often included evaluative research only when it was required by a funding agent (Caro, 1971, p. 92).

Evaluation is perceived by the program staff as organized fault finding usually without adequate understanding or explanation of why the fault exists (Ferman, 1969, p. 147).

Men in power are not merely technicians, concerned solely about the use of effective means to their ends; they are also politicians, committed to morally tinged precepts and symbols, and striving like all other men to maintain a decent self-image. Truths which are inconsistent with their own self-images are demoralizing and thus, in this very real sense, by no means 'useful' to them (Gouldner, 1961, p. 651).

The traditional methods of evaluation have failed educators in their attempts to assess the impact of innovations in operating systems (Guba, 1972, p. 250).

No one wants to be evaluated by anybody at any time (House, 1973, p. 405).

As long as evaluation is viewed as an impersonal, useless edict from a state capital hundreds of miles away, it may take more than public relations' incentives and ideals to sway pragmatic administrators (Kerins, 1973, p. 79).

An analysis of these statistics led to several basic conclusions. First, most evaluative research uses the simplest possible experimental design, substituting simplicity for efficiency. Second, the findings of evaluative research are unrelated to the ways in which change is measured. Third, the technical proficiency of this research is at a low level . . . (Mann, 1972, p. 276).[2]

Two conclusions are suggested by the analysis: the quality of evaluative research is remarkably poor, and there is little difference in the results of evaluative studies conducted in different content areas. Specifically, there is no indication that the findings of evaluative research are influenced by the method tested, the content area in which the test is conducted, the change criteria used, or the methodological quality of the study of which the evaluation is made. . . (Mann, 1972, p. 278).

The general conclusion of this comparative review is, therefore, that evaluative research represents a specific blind alley. It has failed to validate itself in practice and the sooner its failure is accepted and recognized, the easier will be a transition to another approach to the same problem (Mann, 1972, p. 278).

[2]Mann's study was concerned with projects in the affective domain, such as psychotherapy, counseling, and human relations training.

> To the program operator, who knows that his program is doing well, evaluation is at best unnecessary and at worst, if it shows few positive effects, a calumny and a threat to the future of the program, his job, and needed help to clients (Weiss, 1972, p. 331).

Nevertheless, despite the strong theoretical and historical arguments both predicting and documenting poor evaluator-project director relationships, the evidence we have assembled in this study is to the contrary. When asked about preferences for evaluators, not one project director indicated that he would prefer not to have any evaluators. In addition, 38 of the 42 project directors indicated that evaluation findings were of assistance in generating recommendations for project changes (PD20d).

Further evidence of the close relationship between the project director and evaluator is the fact that a substantial number of project directors (N=15, 36%) reported assisting in the *writing* of the final evaluation report. That the project director was the decision maker of most immediate concern to the evaluator is further suggested by the fact that in 69% of the cases, the final evaluation report was sent to the project director who then assumed the responsibility for dissemination to federal sponsors, etc.

Thus, we develop a picture of the evaluator in frequent contact with the project director. (When asked how many site visits were made by the evaluator many project directors said something like, "Oh, he was always here," or "too many to count.") Further, an image develops of the evaluator working with the project director to improve the project. There should naturally develop some involvement on the part of the evaluator—a sense of being part of a cohesive group with a common goal. This associative impulse could well be strengthened in the case of an external evaluator by his realization that his future employment might depend heavily on recommendations and on his reputation among project directors.

House (1972) has written: "Who sponsors and pays for the evaluation makes a critical difference in the eventual findings" (p. 409). What we are arguing here is somewhat different. The federal agency may pay for an evaluator but if the project director hires him and exercises a certain amount of control over his professional career, then it is the project director to whom the evaluator will pay most attention. This tendency will be strengthened by the proximity and frequent contact between project director and evaluator; the federal sponsor is a remote entity. We are hypothesizing, then, that it is the social-psychological context (salience, proximity, group pressures, fate control) and the developmental stage of the project that together determine which decision maker will receive the most attention and, consequently, the kind of evaluation (formative or summative) that is performed.

Evaluation Techniques

Now we will examine the way in which the formative-summative role conflicts in Title VII were manifest in the evaluative techniques employed and the subsequent evaluation reports.

Evaluation Design. One critical consequence of the distinction between formative and summative evaluation is the differentiation of the methodologies. Is the appropriate methodology for formative evaluation the same as that for summative evaluation? We suggest that formative evaluation, aimed at maximizing the success of a specific project, does not require the same methodology as summative evaluation. There does seem to be some consensus supporting this suggestion. One discussion characterizing current thought in this area can be found in Popham (1972). In considering the various analytic techniques available, Popham suggests that formative evaluation can be more informal utilizing pre-experimental (i.e., the one-shot case study) designs aimed at maximizing the successful development of a specific project. The emphasis here is in gathering frequent information about program processes and effects without the need for comparison groups and strict controls. Summative evaluation, on the other hand, is comparative in nature and therefore necessarily requires either a carefully executed quasi-experimental design (Campbell & Stanley, 1966) or pure experimentation (Wiley & Bock, 1967).

In formative evaluation the focus is short-term; effects are observed immediately following treatment. For example, a pretest-teach-posttest situation, that an evaluator might assess, frequently occurs in the space of a few hours. This design

0_1	X	0_2
(pretest)	(teach objectives)	(posttest)

is considered pre-experimental and, if observation #1 (0_1, the pretest) and #2 (0_2, the posttest) were widely separated in time it might well be that any difference between 0_1 and 0_2 could be attributed to some other source than the lesson (history or maturation in Campbell's and Stanley's 1966 terms). However, if little time has elapsed between 0_1 and 0_2 and if it is only during the lesson itself that the student's activity has been related to the lesson's objectives, then the inference that X, the lesson, caused the change is a very strong one. Although this design does not answer the question as to whether X could be replaced by some more effective treatment, it does assess the impact of X under the conditions described. This design, 0_1 X 0_2, is quite inappropriate, however, in assessing the impact of X where X is a full year's project. The competing hypotheses that children learned from some other source than the project or that simple maturation would have produced significant changes between 0_1

and 0_2 are usually too strong even to permit the inference that the project caused any observed effects, let alone to establish that the special project was any more effective than a (less expensive) year of normal schooling would have been. The summative evaluator must provide answers to these latter two questions: Did the project cause the prepost gain? Was the gain significantly better than the gain obtained by a regular, less expensive school program? In order to answer these questions the evaluation must be comparative and must possess a well-defined experimental or quasi-experimental design.

We noted in Chapter II that generally few evaluators employed experimental designs. Specifically, less than one-third of the evaluators successfully implemented a quasi-experimental design while no evaluators were able to attempt a true experimental design. In terms of Title VII, involvement in formative activities might account for the lack of experimental designs: evaluators realized that these were not in fact relevant to their immediate concern, which was to assist the project director in creating an effective, operable project. This meant the evaluations frequently provided little basis for summative kinds of decisions.

The few experimental designs observed in the evaluation reports probably represent attempts at summative evaluation efforts. The difficulties reported in implementation of these designs possibly reflect the general "unreadiness" of the target projects for such an evaluation. Typically these projects were quite new, and as such were heavily involved in developmental activities in which the definition of the project was subject to continual change.

In general, the overall organization of an evaluation design in Title VII was quite vague and the structuring of specific analyses even less visible. This situation, however, is not new. The lack of experimental designs and the unsuccessful implementation of those designs that are attempted is well documented in literature.

To that end, Guba (1972) observed in an examination of Title III program proposals that:

> Twenty-one (Title III) proposals were examined, but only one was found that could be considered to have an adequate design from a traditional methodological point of view. Most had no design at all; while those that did, offered designs known to suffer from serious deficiencies. Yet the majority of these 21 proposals purported that the services of an evaluation specialist had been employed. . . It is certainly a serious symptom of disorder when the experts in the field of evaluation seem to be unable to design evaluations that meet even their own criteria of technical soundness (p. 253).

We wonder if the evaluators, both in Guba's and the present study, were really *unable* to meet their "own criteria of technical soundness." While this is probably true in some cases we suspect that it was very often the case that the evaluator sensed, if not verbalized, his essentially

formative role, and purposely avoided experimentation. *He may have been wise to do so.*

The most likely outcome after only a year of project operation is "no significant difference" between two randomly assigned groups. The prospect of such an outcome does not overjoy an evaluator, particularly where significant differences are considered the hallmark of successful projects. But as Scriven (1967) has pointed out, a finding of no significant difference does not mean nothing has been learned from the project. In the early years of a project, perhaps the more vital information is an understanding of its operational and educational deficiencies so that it can be modified. If Title VII projects that became stabilized in the second or third year still failed to move bilingual children ahead in life, if they did not close the gap between the advantaged and disadvantaged, this is also vital information. In such a case, we must then look for other means of meeting our goals (such as the "multiple push" approach: Scriven, 1967, p. 66). Information derived from many projects about both effective and ineffective strategies might contribute to productive decisions about future efforts.

Evaluation Criteria. Evaluation never takes place in a vacuum but is rather involved with judgments of quality. The criteria and standards against which educational innovations are judged is perhaps the most sensitive, least agreed upon issue confronting evaluation theorists and practitioners alike. Although a great deal of effort and numerous texts and articles have dealt with the importance of establishing measurable, meaningful performance standards (Tyler, 1950; Popham, 1969), such efforts are typically overlooked in practice. At best, arbitrary standards such as a percentage gain on a test are provided. This situation is further confounded by the lack of empirical or theoretical evidence establishing educationally relevant gains.

The existence or formulation of criteria for formative evaluation is even more difficult. In the absence of structured experimental designs and comparison groups, subjective judgments of project quality are frequently the sole standard for assessing its merit.

The formative nature of Title VII evaluations performed in the 1971–72 school year may account for our finding, reported in Chapter III, that evaluators almost always considered projects successful in the important area of cognitive outcomes. Only one evaluator reported unsuccessful cognitive outcomes while 32 reported definite successes (E118). Perhaps, unlike most other attempts at compensatory education, success *was* consistently obtained but it seems at least as plausible to suggest that the heavy involvement of evaluators in *formative* evaluation naturally affected their judgment, especially since the extent of "success" was bound to be a subjective judgment in the characteristic situation of the absence of an experimental design or other well defined standards.

Hidden Nature of Evaluation Reporting. By its very nature formative evaluation is oriented towards uncovering areas in need of modification. Thus, evidence of a "successful" formative evaluation might be a long list of the resultant project changes. These changes are generally assumed to represent mistakes or errors in the plan. However, these formative-based changes need not reflect negative aspects of the developing project. Formative evaluation, for example, might reveal an unintended but very positive effect of the experimental program; the program plan might therefore be altered to further facilitate this result. Nevertheless, formative evaluation is frequently associated with fault-finding and consequently can meet with fear or outright hostility. Project directors and their staff who otherwise would welcome helpful criticism fear that when these results are reported to their funding agency, the presence of modification will be interpreted as sloppy planning and administration.

This situation has caused formative evaluation to adopt a hidden role; that is, although such evaluation does take place it is infrequently reported to funding agencies. This "hidden nature" of formative evaluation is documented in the literature. Kosecoff and Wingard (1972), reporting on an evaluation of a campus drug education program, noted that although the entire staff welcomed and made use of formative information, they were reluctant to report the results to their sponsors. Alkin (1973) points out that this "hidden nature" might be caused by the multiple decision maker or no-man-can-serve-two-masters bind. Since formative evaluation is developer-oriented, it is presumed to be better received if provided to developers (that is, the project director and his staff alone) rather than shared with the project sponsors. These sponsors, on the other hand, might make better use of an audit report that validates the soundness and the appropriateness of the formative evaluation activities.

In terms of Title VII, it was suggested in the previous section that the lack of experimental designs reflects a concentration on formative activities. However, evidence of extensive formative evaluation could not be readily found in an initial reading of evaluation reports. Little was mentioned in the reports in the way of problems encountered or modifications made during the ongoing implementation of the project—the kinds of situations to which we would expect a formative evaluator to attend. For example, the Evaluation Data Sheet indicated that 17 out of 42 evaluators did not report any procedures for verifying whether the project was implemented as planned (See Appendix 1, Evaluation Data Sheet, Question E21).

However, the picture obtained from the Project Director Questionnaire was entirely different. From this questionnaire we discovered that extensive formative evaluation was conducted. Indeed, only two project directors reported that no formative evaluation was conducted for their project (PD20) while 16 reported "extensive" formative studies.

This dearth of formative results reported by evaluators and the emphasis on formative activities reported by project directors confirms the hidden nature of formative evaluation in Title VII. The knowledge of the summative "refunding and continuation" judgments to be based on the evaluation reports and the involvement of the formative evaluator in the successful development of the project might have stimulated many evaluators to avoid reporting the full extent of their evaluation activities, and in particular to skip over project flaws and the recommended modifications.

Although formal formative activities were infrequently cited in evaluation reports, the presence of some formative evaluation can be inferred from these reports as evidenced in the following examples.

In one excerpt from an evaluation report, the late arrival of materials (a formative concern) is only briefly mentioned in passing and without discussion or recommendations.

> As evaluators we have worked very closely with the rest of the staff, and have a good knowledge of the quality of the school personnel who worked in the program. Except for one or two exceptions they are performing admirably. Even when materials were late in arriving their creativity filled the gap. As a case in point, one of the communities is losing an excellent community coordinator because they recognize her skills and have given her a teaching position in next year's expanded bilingual program.

Evidence of summative evaluation in conflict with formative, and the lack of attention subsequently given to formative concerns, is provided in a second quotation. This passage was included in the evaluation report as a brief addendum to a table displaying the results of a t-test. The evaluator said, "After testing, it was found that the teacher of the bilingual grade one at. . . had changed the prescribed curriculum."

The final excerpt documents one evaluator's sensitivity to reporting negative results.

> It will be instructional to look for a moment at some of the information available from the teachers' final report. Although this Overview is meant to be positive, the teachers did provide some very instructional criticisms in their suggestions for the future. These criticisms of the program, just as those in the body of the report, are not meant to report failure, but to improve the quality of the program. It the teachers felt the program was a failure, they would not be returning, and all but two are.

The hidden, almost subordinate nature of formative evaluation reporting is apparent in Title VII. It is our belief that once again the formative-summative role conflict is responsible for this situation. If the kind of evaluation desired had been clearly mandated in the Title VII guidelines and if evaluators had been free to attend to the information needs of the appropriate consumers of that evaluation information (i.e., the project director/program developer in the case of formative and the federal moni-

tor/external user in summative) then the confusion surrounding evaluation reporting and the feelings of defensiveness might have abated.

Utility of Evaluation Information

The lack of utilization of evaluative information has been the source of much dismay in evaluation literature. The frustration of evaluators is perhaps best expressed by House (1972, p. 412) who remarked, "Producing data is one thing! Getting it used is quite another." Weiss (1973) has noted that the minimal utilization of evaluation studies is attributed to the resistance of programs to change and to the poor quality of evaluation endeavors. Ferman (1969) stated "One of the principal problems in evaluation research is how to insure that the results of an evaluation will be utilized" (p. 154), and urged that "A major concern at the earliest point in the evaluation process should be giving some thought to the use of the information obtained" (p 155).

With respect to Title VII evaluations, the predictions of disuse of evaluative information were not fulfilled. The Project Director Questionnaires indicated that evaluation activities were *very* useful to project directors. Only one project director reported that the final (year end) evaluation report was "not at all" influential in decisions to modify the project the following year (PD23b). Sixty-four percent "agreed" or "strongly agreed" that evaluation influenced decisions that were made to modify the project during the year (PD22a). When asked if the evaluation had been of assistance to them, project directors strongly contended that it had been. This applied in many areas but particularly in "identifying possible problems," "preparing reports," and "providing general recommendations for program changes" (PD20a through k).

Once again we must emphasize that this was endorsement of *formative* evaluation, evaluation at the local level assisting in shaping the project into a maximally effective enterprise. The great utility of formative evaluation probably reflects the appropriateness of this kind of evaluation for the target sample of projects. All projects were new and in need of developmental assistance and formative evaluation was most appropriate.

At the federal level, despite the injunction to evaluators to provide information enabling federal personnel to judge which projects "warrant continuation," we found no evidence that evaluations were in fact used for this kind of *summative* decision (see Chapter IV). Refunding levels remained rather constant from year to year with all projects receiving similar percentage increases or decreases in funding.

As we noted in Chapter II, the finding that evaluation was seen, at the local level, as beneficial and useful contrasted sharply with previous observations. We can hypothesize several explanations for these differences.

One possible explanation for the willingness to accept and utilize evaluation information at the local level might be the realization on the part of the local staff that: (1) evaluators can provide valuable technical assistance; and (2) refunding decisions are seemingly unrelated to evaluation report findings and thus the reports are not viewed as a major external threat to survival. In part this lack of usefulness of evaluation data to federal decision making is caused by timing of the evaluation reports. Refunding decisions are made some months before evaluation reports are submitted. Thus, the information in the reports are too late to influence the funding for the next fiscal year, and too old to influence the funding on the year that follows. If deficiencies are noted in the report, project directors could maintain (without evidence available to contradict) that the problem had now been alleviated. Thus, as the externally threatening aspects of evaluation abate, the project directors are able to establish better working relationships with evaluators and to make better use of their talents. In his formative role the evaluator's expertise is applied to the project's successful development.

Summary

In the beginning of this chapter the concept of a decision context was developed and the theoretical necessity (usefulness) of defining an evaluation's context was firmly established. The question was then raised as to whether this necessity was borne out in practice, as evidenced in the Title VII experience.

Our answer was a resounding "yes." The evaluator's perception of the relevant decision maker(s), of the decisions to be made, and of his own relationship to the target project probably framed his evaluation strategy, thereby determining the kind of information that was collected, the data collection and analysis procedures, the way evaluative information was reported and, consequently, the usefulness of the evaluation for decision making.

In the Title VII projects studied, the decision context was far from well-defined. The identity of the relevant decision maker(s) was not clear. Evaluators were hired by the local decision makers and yet were instructed by the federal guidelines to serve the decision needs of both local and federal decision audiences. This situation led to a role conflict for the evaluator. Confusion surrounding relevant decisions to be made rendered the decision context even more ambiguous. In the federally prepared Title VII Manual (HEW, 1971), evaluators were directed to provide information for both formative and summative kinds of decisions. Because concurrent formative and summative evaluations are not compatible, this situation eventuated in role conflict for the evaluator. A formative directive presumes a project in the developmental stage while a summative directive presumes a finished, replicable product.

Faced with ambiguity and confusion concerning the evaluation purpose as defined by the Title VII guidelines, evaluators were forced to define the evaluation decision context based on their own experience and judgments of the project's most immediate information needs. The result was that the majority of evaluators opted for formative evaluation and a decision context in which the relevant decision maker was the local project director and his staff, the decisions to be made concerned needed modifications and changes for project improvement, and the project was viewed as being in a developmental stage and subject to change. This decision seems reasonable in view of the projects' status (most were in their first or second year of development) and the evaluator's responsibility to the project director. The decision to pursue formative evaluation may explain our findings concerning the lack of true experimental designs, the ambiguity surrounding evaluation criteria, the hidden nature of evaluation reporting, and the usefulness of evaluation information to the project director.

Chapter VI
CLARIFYING FORMATIVE AND SUMMATIVE EFFORTS

This chapter will draw upon the findings presented earlier in the monograph in order to clarify the nature of evaluation efforts. In the data of Chapters II–IV we examined the nature and characteristics of Title VII projects vis-à-vis their evaluators and auditors. In Chapter V we focused on a particular theoretical issue which, in our estimation, forms the basis for many of the deficiencies noted in evaluation reports. That issue is the understanding of the appropriate decision context of the evaluation which leads to recognition of the basic distinctions between formative and summative evaluation. The importance of attending to the formative-summative distinctions has been documented as have the difficulties encountered by evaluators who have been unable to place their efforts in proper decision focus.

Drawing from the evaluation theory presented in the preceding chapter as well as the empirical data presented earlier in the monograph, we can now make recommendations for further clarifying the formative-summative distinction. These recommendations will be stated in terms of:

1. distinctions between formative and summative evaluation that have been generally accepted in the literature but which are to a large extent inadequately heeded, and
2. distinctions between formative and summative evaluation on which there is a considerable lack of consensus in the literature or which are conceptually new ideas.

Some Generally Acknowledged Distinctions

Chart 6.1 illustrates four generally acknowledged bases for distinguishing between formative and summative evaluation. These are: (1) the developmental stage of the project during which the evaluation takes place; (2) the evaluator's stance vis-à-vis the project; (3) the design of the evaluation; and (4) the type of instruments used in the evaluation. The chart suggests the manner in which each characteristic might be evident in a form unique to formative or summative evaluation.

Developmental Stage of the Project

Almost without exception, projects require some cycles of development during which modifications and improvements can be introduced. This fluid period of project maturation requires formative evaluation. On the other hand, in order to assess the final impact of the project it should be replicable, well-defined, and stable. Summative evaluation must therefore be postponed until the project is mature. We would be hard put

to find a single evaluation theorist or practitioner who would challenge this principle. Yet, our experience has shown that both evaluation mandates and evaluators unabashedly violate this basic evaluation distinction.

CHART 6.1

Some Generally Acknowledged Distinctions Between Formative and Summative Evaluation

Characteristic	Formative	Summative
Developmental stage of project	fluid, undefined in development	well-defined, replicable, stable
Evaluator stance towards project	involved in project, partisan	external to project, impartial
Design of evaluation	exploratory flexible focuses on individual components iterative process not necessarily comparative but seeking influential variables	well-defined evaluation design as unobtrusive, non-reactive as possible comparative concerned with a broad range of issues, e.g., implications, politics, costs, competing options
Instrumentation	great variety of instruments locally developed as well as standardized reliance on observation and frequent informal data collection devices locally chosen some concern for validation of instruments to be used in subsequent summative evaluation	instruments selected to reflect decision concerns of sponsor/ decision maker reliable and valid instruments

Evaluator's Stance Towards Project

The formative evaluator works with the local project staff to effect improvements. As the project improves, in part due to his efforts, it is natural that the evaluator will feel increasing commitment to the project and its goals. Such partisanship is not out of place for the formative evaluator. The summative evaluator, on the other hand, must maintain an objective stance; he must not be co-opted into project management. His necessary objectivity must be understood and accepted by all concerned.

Design of Evaluation

It is the formative evaluator's responsibility to locate and measure the strengths and weaknesses of a project and, in doing so, provide the developers with constant feedback. The designs that he employs may, consequently, focus on specific project components. The designs will allow for the exploration of many variables, they will frequently be iterative, and they will always be flexible and responsive to changes in the project's plan. The provision of continual feedback is the distinguishing characteristic of formative evaluation.

In contrast to formative evaluation, summative efforts demand a stable, well-defined evaluation design capable of leading to inferences about project worth. These inferences about project impact derived from summative evaluation must be valid; this necessitates an evaluation design that adequately accounts for all competing hypotheses. One source of contamination of project effects might be the evaluation itself, which consequently should be as unobtrusive and nonreactive as possible.

A final distinction is the evaluator's perspective. The formative evaluator is concerned with the examination of process and to the extent that he considers the outcomes of the project he is restricted to the examination of progress towards determined goals. The summative evaluator's perspective, on the other hand, need not be limited to measuring the project's immediate goals. He may take a broad view of the project in its social context in the manner advocated in Scriven's (1972) goal-free evaluation model. Whether or not he takes this panoramic view or limits himself to assessing goal achievement is determined by the sponsor's and the potential user's information needs.

Instrumentation

The formative evaluator should use a great variety of instruments—both locally developed as well as standardized. In so doing he can maximize the chance of finding project effects and identify the instruments that are reliable and valid for the situation at hand. The choice of instruments to be used by the formative evaluator should be heavily influenced at the local level, but nevertheless be made with an eye to the summative evaluation to follow. This might well involve the validating of locally developed instruments, the use of which might be acceptable, subsequently, to the summative evaluator.

Because of the crucial nature of summative decisions, the instruments used in summative evaluation should have been demonstrated to be valid and reliable in the target population. Further, the choice of instruments should, in large part, be influenced by the sponsors or potential users.

A major benefit in having sponsors participate in instrument selection would be the increased possibility of making valid comparisons across

projects. For example, if all Title VII preschool projects had employed the same measures, some comparative analyses would have been possible; at the very least, some reliable baseline data that might indicate criteria levels might be established.

Some Far-from-Generally Acknowledged Distinctions

Chart 6.2 illustrates additional distinguishing characteristics of formative and summative evaluation. These are: (1) the decision maker(s) to whom the evaluator will report; (2) the timing of the evaluation; (3) the evaluator's relationship to the project; (4) the need for the educational audit; and (5) the evaluator's attitude towards research. In addition, the differing nature of each specified characteristic is summarized for formative and summative evaluation. These distinctions are less generally acknowledged and/or provide conceptually new bases for distinguishing between formative and summative evaluation.

CHART 6.2

Some Far-From-Generally Acknowledged Distinctions Between Formative and Summative Evaluation

Characteristic	Formative	Summative
Decision makers	project developer (e.g., project director and his staff)	sponsors (e.g., federal government, school board) potential users (e.g., other schools, teachers, school boards)
Timing of evaluation	early stages of project development (e.g., during first two years)	at point in time when project can reasonably be viewed as stable (e.g., during third and fourth years)
Evaluator's relationship to project	internal or external hired by project personnel	independent of project, external only assigned by project sponsors
Need for external educational accomplishment audit	needed	not needed
Evaluator's research stance	concerned with advancing reliable basic research	concerned with advancing reliable basic research

Decision Makers

Identification of the project's decision maker(s) is perhaps the most elusive aspect of organizing an evaluation effort. Different organizations characteristically have very different decision structures ranging from those in which specific decision responsibility is emphasized to those in which rather amorphous divisions of decision-making responsibility exist. The spectrum of potential decision makers is equally diffuse, ranging from those individuals who have daily contact with the target project to those more distant from the project but who have, for example, provided for its funding. Evaluative information might thus be used by project staff (teachers), the project director, other project administrators, district administrators, community groups, special interest groups, project sponsors (state, federal, or private granting agency), or Congress. One of the major problems for the evaluator is to identify the decision audience to whom his reporting will be primarily directed. The important consideration is that the evaluator be quite clear as to which decision maker or audience he has responsibility for evaluation information reporting. General purpose information provision cannot be considered evaluation.

During formative evaluation the appropriate decision makers are those responsible for instructional program development, that is, the project director and his staff. Their concern is with detecting program weaknesses and making needed modifications while the program is still fluid, as well as validating the finished program as suitable and meeting its goals.

During summative evaluation, on the other hand, the emphasis is not on program modification, but rather on a thorough description and analysis of program merits directed at such questions as: Who benefits from the program? Under what conditions? How does the program compare in goal achievement and costs with other program alternatives? The appropriate decision makers during summative evaluation are, therefore, those who are interested in using, extending, continuing or discontinuing the program—that is, the potential users and sponsors.

Timing of Evaluation

In the discussion in the previous chapter, the conflicts between the formative and summative evaluation roles in Title VII evaluators were documented. One means of partially resolving this role conflict is through a careful delineation of the timing of each evaluative function. We would start with the recommendation that formative and summative evaluations be kept as discrete and separate as possible and that there be an initial formative stage in which no summative inquiries occur. Should summative evaluation require data from this period it would be available from the formative evaluator's records.

Clearly, formative evaluation takes place during the early cycles of the project. The number of cycles of the project allotted solely for forma-

tive evaluation should be specified so that the formative evaluator and project management are aware of this time constraint, feel free to modify the project as much as necessary in these early cycles without fear of criticism, and produce a stable instructional program in time for the summative evaluation period.

After this initial formative period summative evaluation efforts should begin. This does not mean that the formative evaluator abandons his job on the commencement of summative evaluation, but rather there is a period of overlap in which the formative evaluator briefs the summative evaluator on all aspects of the program and on measurement techniques he has found effective.

In the case of a Title VII project, we feel that two years of undisturbed formative evaluation would be reasonable followed by two years of replicated summative evaluations. As we have just noted, the formative evaluator provides the summative evaluator with an immense amount of useful advice and information based upon his intimate and first-hand knowledge of the project. Of course, in some situations, the summative evaluator may not wish to be "contaminated" by this information (cf. Scriven's Goal-free Evaluation, 1972).

Relationship to Project

In the formative stage, whether the evaluator is internal (an employee of the target agency) or external (an employee of an agency not affiliated with the target agency) does not seem to be of great importance, since whatever his internal/external status is, his duties require close contact with the project and its staff. It would be impossible to properly perform the formative evaluation function without a great deal of involvement and partisanship. Even a member of an external agency would become committed to a project that grew in response to his recommendations.

In the summative stage, however, the evaluator should not only be external to the project but should also be professionally independent of the effect of recommendations or the good will of project directors. One possible means of making available this kind of independent evaluator is the prior selection of highly qualified professional evaluators selected to serve a particular geographical region. These evaluators might either be constituted as units of state or county education agencies, private agencies, or special regionally based federal units. The important distinction between this recommendation and current practice is the independence of the unit from the agency whose project is being evaluated and the choice of unit or evaluator by the project sponsor. We recognize this suggestion is not a standard operating procedure at present. However, the existence of a nonpartisan pool of investigators is not unknown in organizational structures; for example, Congress' investigative arm—the General Accounting Office.

Need for Educational Audit

The formative evaluator's responsibility is to the project developers. During the formative years, however, the sponsors of the project have a legitimate concern that the monies allocated to the project and its formative evaluation are being effectively used. An independent educational accomplishment audit can assume this accountability function. The auditor, of course, is independent of and external to the project. His report will document the technical acceptability of the formative evaluation activities and the extent to which evaluation feedback is being effectively utilized by the project director. His function is not to act as a pseudo-summative evaluator or as a second formative evaluator.

The audit function, however, would be superflous in our proposed summative stage in which the evaluator is himself legitimately and demonstrably independent of the project. This recommendation will alter somewhat the current Title VII policy; at present, an audit is an everpresent aspect of a Title VII evaluation effort, whether the evaluation is oriented to formative or summative decisions.

The Evaluator's Research Stance

The preceding recommendations have created a rather complete organizational distinction between formative and summative evaluation focusing primarily on the differences in the decision audience, timing, the evaluator's relationship to the project, and the necessity for an independent, external educational audit. We now suggest an attitude that should be present in both formative and summative evaluators—the evaluator's research stance.

Some might suggest that summative evaluation as we envisage it is exceedingly like research, with its emphasis on unobtrusiveness, good experimental designs, objectivity, etc. A difference between research and evaluation lies in the time perspective. The researcher may feel that two cycles of experimental tryout are insufficient to provide strong inferences. Although the summative evaluator may feel the same, he must nevertheless use the best possible design and present to the decision maker the *likely* differences between experimental and control groups *along with all necessary caveats.* He must provide this summary information when needed, simply because decisions are frequently unavoidable and cannot be postponed. This is a reality; we must frequently act on the basis of insufficient evidence and take a leap in the dark. The evaluator musters as much light as possible in the time available. The researcher, meanwhile, struggles to construct a lasting beacon.

However, the formative as well as the summative evaluator should be concerned with advancing basic research even though each has as his first priority meeting the needs to decision makers. Is this stance as cross-

eyed as having a single evaluator try to perform formative and summative evaluations simultaneously? Do we put the evaluator into a state of conflict between what Cronbach and Suppes (1969) have distinguished as different research styles; that is, conclusion or decision-oriented research? Can the problem be avoided by having the evaluator ignore basic research problems and concentrate solely on decision making?

There are several reasons why evaluators cannot be and should not try to be purely and exclusively decision oriented. In the first place, a real-life situation does not come neatly packaged as decision or conclusion oriented. A problem of measurement confronted by an evaluator will frequently turn out to involve some basic theoretical question. As an example, consider the choice of which variables to study. This choice is heavily dependent on basic theory. Consider "critical thinking ability" which some curriculum evaluators have wished to measure. Does such a trait exist, generalizable across subject areas? This is a theoretical question for conclusion-oriented basic research and any results from tests purporting to measure it should be interpreted in this light. But if "critical thinking" is of concern to the developer, the formative evaluator might make clear the theoretical problems and, if he uses such tests, contribute his findings to the accretion of knowledge necessary for theoretical advances. The decision-oriented researcher (the evaluator) is concerned with specific decisions, but when he sees that generalizations can be drawn from specific instances he should rejoice and report his results.

The evaluator should also be concerned with the accumulation of knowledge if for no other reason than the sake of efficiency. An evaluation should provide information not only for immediate decisions, but also, if possible, for superordinate decisions that may be made at a later date. Such decisions demand generalizations. Indeed, Wardrop (1969) has noted that generalizations are likely to be made with or without the evaluator's approval. Referring to the evaluator, Wardrop states:

> He must guard against the very real tendency of his audience (intended or otherwise) to make inferences, to base decisions about contemplated actions on the results of the evaluation study. The evaluator's most effective,—if not his only—safeguard lies in using the most scientific, the most generalizable design he is able (within the constraints he always encounters) in planning his study (Wardrop, 1969, p. 42).

Fears have been expressed that evaluators, trained in research methodology, will use evaluation opportunities to further their own academic interests. Such researchers may spend the bulk of their time exploring basic theoretical problems and selecting measures with these in mind, rather than attending to the decision needs of the project. In the case of Title VII, an antithetical problem was evident. Evaluation reports were so atheoretical as to seem like cookbooks ("administer two tests and one questionnaire, mix with process objectives, and write a report," seems all too

frequently to have been the recipe). Even basic queries in research methodology such as "are the measures reliable or valid in this context?" were frequently never considered. Few evaluators appeared to hold any theoretical stance whatsoever, either about themselves, about evaluation theory, or about the projects. The Title VII data presented in this monograph, for example, showed that very few evaluators even gave evidence of using an evaluation model or framework to guide their own efforts.

Finally, we trust that the evaluator is an intelligent person, well versed in educational research. He cannot suddenly disengage his reflective curiosity about basic theoretical questions. Everyone learns to cope with the tension between the necessity for action (decision orientation) and the desire for firmly grounded knowledge (conclusion orientation). The dilemma is intrinsic to life.

The Liaison Between Formative and Summative Evaluators

To conclude this monograph with the impression that it is our intent that formative and summative evaluation are to be completely bifurcated, each relegated to its own capsule, not to be contaminated in any way by the other, is a miscalculation of our position. The formative evaluator must constantly bear in mind the impending summative evaluation. He will eventually sit down with the summative evaluator and recommend a variety of outcomes that should be measured and the possible means by which this can be accomplished. The formative evaluator should also take pains to document existing local conditions of which the summative evaluator should be aware.

Let us attempt to clarify what we have in mind with an example. Suppose that a project has as an objective the involvement of parents in school activities but few parents in fact participate. The formative evaluator may provide data, from the formative years and earlier, showing that parent participation has always been poor but that it is now increasing. The reasons for the low level of participation may have been investigated by the formative evaluator by such means as parent interviews and by some other data-collection techniques. Perhaps he found that when both parents hold jobs there is little participation in school activities; income level, on the other hand, may be uncorrelated with parent participation. The formative evaluator should, after two or three years, have formed some hypotheses accounting for such defects in the project, defects of which he was aware but about which little could be done. Such information will be important to the summative evaluator not only in providing caveats to some negative conclusions but also in the selection of variables to be attended to in selecting control groups. To continue the example, the summative evaluator may have considered stratifying families of similar income levels in his random assignments to experimental and control

groups. The formative evaluator may suggest on the basis of his experience (supported by data verified by the independent auditor) that the stratification could be better based upon number of hours worked per week by each parent.

Stake (1972, p. 47) has stated that deciding which variables to study is an "essentially subjective" commitment in evaluation. As in any decision, however, the choice is facilitated by the existence of information about the available alternatives. In this area the summative evaluator faces very difficult decisions. He must measure thoroughly and cannot dabble tentatively in all aspects of the project. The exploratory, dabbling, hypothesis-generating phase belongs to the formative evaluator who eventualy provides the summative evaluator with his findings. The summative evaluator can then select the few aspects of the project that are to be thoroughly measured and the measures to be applied. He should be required to provide an explicit rationale for these decisions, a rationale drawing on data from the formative evaluation and also stating clearly where the choices have been subjective.

As we pointed out in Chapter II, this kind of rationale for the selection of measures was notably absent from evaluations. Test instruments were occasionally described but almost never was there any discussion of a rationale for their selection. This situation has been previously noted by Fitzpatrick (1970):

> The rationale . . . is a written version of the thought process by which the evaluator ought to have arrived at his decision concerning each measure. By committing the process to writing, he increases the likelihood that he will make a balanced decision and he makes it possible for his colleagues to understand and criticize his decision. This may seem obvious, but how many evaluation reports have provided anything approximating a rationale? (p. 75).

Summary

In this chapter we have presented nine distinctions between formative and summative evaluation. The characteristics that form the bases of the more well-known and well-accepted distinctions were listed as follows:

- developmental stage of the project
- evaluator stance vis-à-vis the project
- design of the evaluation
- type of instruments used

The principles stemming from these characteristics are accepted in evaluation literature, yet they are infrequently applied in practice. It is our contention that adherence to these principles in the planning and implementation of an evaluation would greatly aid an evaluator in avoiding a host of decision-making entanglements.

A second class of distinctions was also presented for consideration. These were new or less consensual in nature and were based on the following five characteristics:

- decision makers
- timing of the evaluation
- evaluator's relationship to the project
- need for external educational accomplishment audit
- evaluator's attitude toward research

Emphasis was placed on the importance of clarifying differences in these characteristics for formative and summative evaluation. Although it is suggested that the evaluator be aware of and adhere to the many distinctions between these two types of evaluation, it is essential that these differences do not impede extensive communication and cooperation between formative and summative evaluators.

REFERENCES

Alkin, M. C., & Kosecoff, J. B. A new eclectic model for the redirection of evaluation efforts. *CSE Report* No. 84, March 1973. (ED 075 086.)

Argyris, C. Creating effective relationships in organizations. In R. N. Adams & J. J. Preiss (Eds.), *Human organization research.* Homewood, Ill.: Dorsey, 1960.

Bend, E. The impact of the social setting upon evaluative research. In *Evaluative Research,* Pittsburgh: American Institutes for Research, 1970.

Campbell, D. T. Reforms as experiments. *American Psychologist,* 1969, *24*(4), 409–429.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1966.

Caro, F. G. Issues in the evaluation of social programs. *Review of Educational Research,* 1971, *42*(2), 87–114.

Cronbach, L. J. & Suppes, P. (Eds.) *Research for tomorrow's schools.* New York: Macmillan, 1969.

Department of Health, Education and Welfare. *Manual for project applicants and grantees.* Programs under Bilingual Education Act (Title VII, ESEA), Washington, April, 1971.

Ferman, L. A. Some perspectives on evaluating social welfare programs in evaluating the war on poverty. *The Annals of the American Academy of Political and Social Science,* Philadelphia, September, 1969.

Fitzpatrick, R. The selection of measures for evaluating programs. In *Evaluative Research,* Pittsburgh: American Institutes for Research, 1970.

Gouldner, A. W. Engineering and clinical approaches to consulting. In W. G. Bennis, K. D. Benne, & R. Chin (Eds.), *The planning of change: Readings in applied behavioral sciences.* New York: Holt, Rhinehart and Winston, 1961.

Guba, E. G. The failure of educational evaluation. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education.* Boston: Allyn and Bacon, 1972.

Guildford, J. P. *Personality.* New York: McGraw-Hill, 1959.

House, E. R. The conscience of educational evaluation. *Teachers College Record,* 1972, *73*(3), 405–414.

Kerins, G. T. Accountability and the state. In E. R. House (Ed.), *School evaluation: The politics and process.* Berkeley: McCutchan, 1974.

Kosecoff, J. B., & Wingard, J. A. Evaluating a campus drug education program. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, February, 1973.

Morin, A. J. *Handbook for educational program audit.* (Rev. ed.) Washington: U.S. Office of Health, Education and Welfare, Office of Education, Bureau of Elementary and Secondary Education, June, 1971.

Popham, W. J. *An evaluation guidebook.* Los Angeles: The Instructional Objectives Exchange, 1972.

Popham, W. J. Objectives and instruction. In W. J. Popham, E. W. Eisner, H. J. Sullivan, & L. L. Tyler (Eds.), *Instructional objectives.* AERA Monograph Series on Curriculum Evaluation, No. 3. Chicago: Rand McNally, 1969.

Scriven, M. The methodology of evaluation. In R. Tyler, R. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation.* AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

Scriven, M. Prose and cons about goal-free evaluation. *Evaluation Comment,* 1972, *3*(4), 1–4.

Stake, R. E. The countenance of educational evaluation. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education.* Boston: Allyn and Bacon, 1972.

Tyler, R. W. The functions of measurement in improving instruction. In E. F. Linquist (Ed.), *Educational measurement.* Washington: American Council on Education, 1950.

Wardrop, J. L. Generalizability of program evaluation: The danger of limits. *Educational Product Report,* 1969, *2*(5), 41–42.

Weiss, C. H. Between the cup and the lip. *Evaluation Comment,* 1973, *1*(2), 1–6.

Weiss, C. H. Utilization of information: Toward comparative study. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education.* Boston: Allyn and Bacon, 1972.

Wholey, J. S., & White, B. F. Evaluation's impact on Title I elementary and secondary education program management. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, February, 1973.

Wiley, D. E., & Bock, R. D. Quasi experimentation in educational settings: Comment. *School Review,* 1967, *75*(4), 353–366.

Appendix 1:
DATA-COLLECTION INSTRUMENTS AND LISTING OF SUMMARY STATISTICS

Project ID _____
Rater _____
Year of Project 1970–71 _____
Date _____

EVALUATION DATA SHEET

Part I: Characteristics of Evaluator(s)

E1. To whom was the evaluation contracted?
- *17* 01 Single individual
- 23 02 Organization or more than one individual
- *2* 98 No information

E2. Was there one or more individual(s) specifically designated as evaluator(s) and who had no other duties within the project?
- *33* 01 Yes
- *3* 02 No
- *6* 98 No information

E3. What was the position of this evaluator(s) with respect to the project?
- *14* 01 Internal
- 23 02 External
- *1* 03 Both internal and external evaluators
- *0* 04 Internal with minimal external consulting aid
- *0* 05 External with minimal internal aid
- *4* 98 No information

For items 4–15 circle "external" if the activity was performed solely by external evaluator(s). Circle "internal" if the activity was performed solely by internal evaluator(s). Circle "both" if the activity involved joint internal and external efforts. Circle "no info" if the activity was performed but it is unclear as to who was responsible. Circle "not applicable" if the activity was not performed *at all.*

	External	*Internal*	*Both*	*No Information*	*Not Applicable*
E4. Formative data collection in general	01(3)	02(*18*)	03(7)	98(3)	99(*11*)
E5. Formative data collection: observation	01(3)	02(*19*)	03(5)	98(4)	99(*11*)
E6. Formative data collection: testing	01(2)	02(*15*)	03(5)	98(5)	99(*15*)
E7. Formative data collection: examination of records	01(3)	02(*10*)	03(6)	98(5)	99(*18*)

	External	*Internal*	*Both*	*No Information*	*Not Applicable*
E8. Formative data	01(*7*)	02(*12*)	03(*4*)	98(*6*)	99(*13*)
E9. Formative data reporting	01(*11*)	02(*12*)	03(*5*)	98(*3*)	99(*11*)
E10. Summative data collection in general	01(*6*)	02(*17*)	03(*13*)	98(*6*)	99(*0*)
E11. Summative data collection: observation	01(*3*)	02(*17*)	03(*13*)	98(*9*)	99(*0*)
E12. Summative data collection: testing	01(*3*)	02(*19*)	03(*11*)	98(*9*)	99(*0*)
E13. Summative data collection: examination of records	01(*14*)	02(*12*)	03(*3*)	98(*8*)	99(*5*)
E14. Summative data analysis	01(*20*)	02(*14*)	03(*2*)	98(*6*)	99(*0*)
E15. Summative data reporting	01(*22*)	02(*14*)	03(*1*)	98(*5*)	99(*0*)

E16. Professional affiliation of principal evaluator:

6 01 University (specify _____)
4 02 Government Agency (i.e., regional lab) (specify _____)
16 03 Private agency (specify _____)
10 04 Project staff member (specify _____)
6 98 No information

E17. Highest degree attained by principal evaluator:

1 01 BA
1 02 MA or MS
14 03 Ph.D. or Ed.D.
0 04 Other (specify _____)
26 98 No information

E18. What is principal evaluator's major field of study?

4 01 Education
2 02 Psychology
0 03 Sociology
0 04 Political Science
0 05 Business Management
0 06 Other (specify _____)
36 98 No information
0 99 Not applicable if there is more than one principal evaluator with different fields of study

E19. Indicate which evaluation model was reported to have guided evaluation efforts:

0 01 Decision oriented model (CIPP/Provus/CSE)
0 02 Descriptive model (Stake)
7 03 Objectives-based (Tyler/Popham/EPIC)
2 04 Other (Specify _____)
33 99 Not applicable—no model reported

E20. Indicate point of entry of evaluator(s)

1 01 At or near program inception
9 02 Prior to program implementation, during program planning
15 03 During program implementation and operation
0 04 After completion of program
17 98 No information

Enroute Evaluation Activities

E21. Did the evaluators report any procedures for verifying whether the program was implemented as planned?

25 01 Yes
17 02 No

For items 22–28, indicate whether the procedure was used to monitor program implementation by circling appropriate number. If the answer to item 21 was "no," circle "not applicable" for items 22–28.

	Used	*Not Used*	*Not Applicable*
E22. Program observation	01(*24*)	02(*1*)	99(*17*)
E23. Review of written reports by staff members	01(*7*)	02(*16*)	99(*19*)
E24. Interviews (specify with whom _____ _____)	01(*12*)	02(*11*)	99(*19*)
E25. Questionnaires	01(*3*)	02(*20*)	99(*19*)
E26. Checklists	01(*7*)	02(*17*)	99(*18*)
E27. Miscellaneous school records (specify _____)	01(*8*)	02(*16*)	99(*18*)
E28. Other (specify _____)	01(*9*)	02(*14*)	99(*19*)

E29. Did monitoring procedures involve entire program or a sample?

15 01 Entire program
2 02 Sample
8 98 No information
17 99 Not applicable, if answer to item 21 is "no."

E30. Did the evaluators report any procedures for monitoring progress toward program objectives during the course of the program?

22 01 Yes
20 02 No

For items 31–36, indicate if the procedure was used to monitor program progress by circling the appropriate number. If the answer to item 30 was "no," circle "not applicable" for items 31–36.

	Used	*Not Used*	*Not Applicable*
E31. Program observation	01(*16*)	02(*6*)	99(*20*)
E32. Written or oral reports by staff members	01(*14*)	02(*8*)	99(*20*)
E33. Interviews (with whom _____ _____)	01(*3*)	02(*19*)	99(*20*)
E34. Nationally normed tests	01(*3*)	02(*19*)	99(*20*)
E35. Locally designed tests	01(*6*)	02(*16*)	99(*20*)
E36. Check lists	01(*16*)	02(*6*)	99(*20*)
E37. Documentation	01(*8*)	02(*14*)	99(*20*)
E38. Other (specify _____)	01(*8*)	02(*14*)	99(*20*)

E39. Did monitoring procedures involve the entire program or a sample?

18	01	Entire program
1	02	Sample
3	98	No information
20	99	Not applicable if answer to item 30 is "no."

For items 40–43, indicate which sampling techniques were used in the monitoring procedures by circling the appropriate number. If the answer to item 39 is *not* "02," circle "not applicable" for items 40–43.

	Used	*Not Used*	*Not Applicable*
E40. Matrix sampling	01(*0*)	02(*1*)	99(*41*)
E41. Random sampling	01(*1*)	02(*0*)	99(*41*)
E42. Stratified sampling	01(*0*)	02(*1*)	99(*41*)
E43. Other techniques (specified or unspecified)	01(*0*)	02(*1*)	99(*41*)

E44. Was any type of analysis performed on the progress data?

17	01	Yes
6	02	No
19	99	Not applicable, if no progress data gathered

For items 45–50, indicate the kind of analytic procedures used by circling the appropriate number. If the answer to item 44 was "no," circle "not applicable" for items 45–50.

	Used	*Not Used*	*Not Applicable*
E45. Descriptive statistical procedures (e.g., means, confidence intervals, proficiency level. etc.)	01(*1*)	02(*10*)	99(25)
E46. Hypothesis-related statistical procedures (e.g., t-tests, ANOVA, etc.)	01(*3*)	02(*14*)	99(25)
E47. Quantified ratings from observations and written reports	01(*13*)	02(*4*)	99(25)
E48. Narrative reports, and documentation	01(*12*)	02(*5*)	99(25)
E49. Summaries of personal observations	01(9)	02(*8*)	99(25)
E50. Other analysis (specified or unspecified)	01(*0*)	02(*17*)	99(25)

E51. If any enroute data (i.e., implementation and progress) was gathered, was it used for summative purposes (e.g., used in final report for final evaluation) or formative purposes (e.g., program modification and improvement)?

6	01	Summative purposes
0	02	Formative purposes
21	03	Both
2	98	No information
13	99	Not applicable if no enroute data was collected

Summative Evaluation Activities

E52. Did the evaluator(s) claim to use an experimental or quasi-experimental design to measure the program's effectiveness in attaining its goals?

42 01 Yes
0 02 No

For items 53–60, indicate which variables characterize the experimental or quasi-experimental design by circling the appropriate number. If the answer to item 52 was "no," check "not applicable" for items 53–60.

	Characterized	*Did Not Characterize*	*Not Applicable*
E53. Random assignment of subjects	01(*4*)	02(*38*)	99(*0*)
E54. One control group only	01(*22*)	02(*20*)	99(*0*)
E55. Two or more control groups	01(*7*)	02(*35*)	99(*0*)
E56. One shot case study	01(*6*)	02(*36*)	99(*0*)
E57. Posttest only	01(*13*)	02(*29*)	99(*0*)
E58. Pre- and posttest	01(*39*)	02(*3*)	99(*0*)
E59. Non-equivalent control groups	01(*20*)	02(*22*)	99(*0*)
E60. Other (specified or unspecified)	01(*2*)	02(*39*)	99(*1*)

E61. Extent to which design implemented:

23 01 Design implemented as planned
16 02 Design not implemented as planned; reasons noted in evaluation report
3 03 Design not implemented as planned; no reasons noted in evaluation report
0 98 No information
0 99 Not applicable if evaluator did not use an experimental or quasi-experimental design

For items 62–66, indicate how the evaluation criteria were selected by circling the appropriate numbers.

	Procedure Used	*Procedure Not Used*	*No Information*
E62. Program's prespecified goals adopted by evaluator (tests found to measure goals)	01(*32*)	02(*6*)	98(*4*)
E63. Evaluator selected criteria	01(*1*)	02(*35*)	98(*6*)
E64. Expert opinion used to select criteria (e.g., Delphi Technique)	01(*1*)	02(*35*)	98(*6*)
E65. Evaluator and staff choose criteria together	01(*8*)	02(*29*)	98(*5*)
E66. Other (specify _____)	01(*0*)	02(*35*)	98(*7*)

For items 67–79, indicate whether the described area was considered in assessing program success by circling the appropriate number.

	Was Considered	*Was Not Considered*
E67. Cognitive achievement: pre-specified goals	01(*42*)	02(*0*)
E68. Cognitive achievement: unanticipated outcomes	01(*8*)	02(*34*)
E69. Affective/Attitudinal changes; pre-specified goals	01(*26*)	02(*16*)
E70. Affective/Attitudinal changes: unanticipated outcomes	01(2)	02(*40*)
E71. Community/Parent relations: pre-specified goals	01(*38*)	02(*4*)
E72. Community/Parent relations: unanticipated outcomes	01(*4*)	02(*38*)
E73. Process Objectives	01(*31*)	02(*11*)
E74. Cost information	01(*10*)	02(*32*)
E75. Staff training	01(*36*)	02(*6*)
E76. Staff performance and attitudes	01(*30*)	02(*12*)
E77. Staff relations	01(*13*)	02(*29*)
E78. Program management	01(29)	02(*13*)
E79. Other (specify _____)	01(*20*)	02(22)

For items 80–89, indicate whether each technique was used to gather data on program success by circling the appropriate number. If no data were collected, circle "not applicable."

	Were Used	*Were Not Used*	*Not Applicable*
E80. Locally designed tests	01(*31*)	02(*10*)	99(*1*)
E81. Nationally normed tests	01(*40*)	02(2)	99(*0*)
E82. Interviews (specify with whom _____ _____)	01(*18*)	02(*24*)	99(*0*)
E83. Questionnaires	01(*34*)	02(*7*)	99(*1*)
E84. Reports from project staff	01(*23*)	02(*19*)	99(*0*)
E85. Checklists	01(*26*)	02(*16*)	99(*0*)
E86. Site visits	01(*31*)	02(*11*)	99(*0*)
E87. Documentation	01(*35*)	02(*7*)	99(*0*)
E88. Archieval data	01(*19*)	02(*23*)	99(*0*)
E89. Other (specify _____)	01(*15*)	02(27)	99(*0*)

For items 90–94, indicate how the data were analyzed by circling the appropriate number. If no data were collected, or data not analyzed, circle "not applicable."

	Used Technique	*Did Not Use Technique*	*Not Applicable*
E90. Multivariant statistical procedures	01(*0*)	02(*42*)	99(*0*)
E91. Hypothesis-related statistical procedures (i.e., t-tests)	01(27)	02(*15*)	99(*0*)

E92. Simple descriptive statistical procedures	01(38)	02(4)	99(0)
E93. Documentation without statistical referents	01(37)	02(5)	99(0)
E94. Other techniques (specify _____ _____)	01(3)	02(39)	99(0)

Part III: Content and Format of Evaluation Report

E95. How were data presented in evaluation report?

- 3 01 All raw data presented as well as summary data and/or results of analyses
- 19 02 Raw data not presented in entirety, but summaries and/or results of analyses well documented
- 19 03 Summary data and/or results of analyses presented, little documentation
- 1 04 Summary data and/or results of analyses with no documentation
- 0 05 Raw data only presented
- 0 06 No data presented

E96. Were data included on all of the program's pre-specified objectives?

- 25 01 Yes
- 17 02 No

E97. If answer to item 96 is "no," were reasons for exclusions given? If answer to item 96 is "yes," check "not applicable."

- 5 01 Yes
- 10 02 No
- 27 99 Not applicable

E98. Did the evaluation report include interpretation of the results of the analyses?

- 1 01 No
- 22 02 Yes, but *very* brief interpretation
- 8 03 Yes, explanations fairly complete, requires some expansion
- 11 04 Yes, complete interpretation included
- 0 99 Not applicable if no data collected *or* no statistical analyses performed

E99. To what extent were assessment instruments described?

- 14 01 Described in full, or included in report
- 12 02 Described all essential aspects, although not described in full
- 15 03 Described in part
- 1 04 Not described
- 0 99 Not applicable if no data collected *or* no assessment instruments used

E100. Which of the following best describes the language of the evaluation report?

- 11 01 Highly technical; language very specialized, reader would need appropriate background to understand report
- 17 02 Somewhat technical; some specialized language used, but reader would understand most of report without any specialized background
- 14 03 Not technical; no specialized background needed to understand report

E101. Did the report include any comments on staff performance?

37 01 Yes
5 02 No

E102. Did the evaluator discuss parent and community involvement in the project?

39 01 Yes
3 02 No

E103. Which category best describes the target community?

1 01 Total Metropolitan Area
9 02 Urban Neighborhood
5 03 Suburban Community
9 04 Rural
18 98 No information

E104. What are the grade levels of the target group?

31 01 Elementary (K-6)
1 02 Intermediate (7-9)
0 03 Secondary (10-12)
3 04 Elementary and Intermediate
0 05 Intermediate and Secondary
1 06 Elementary, Intermediate, and Secondary
6 07 Other (specify _____)
0 98 No information

E105. Predominate race of target group:

0 01 Caucasian
0 02 Black
33 03 Mexican-American
1 04 Oriental
5 05 American Indian
0 06 Other (specify _____)
3 98 No information
0 99 Not applicable—no demographic data collected

E106. Indicate number of schools in district:

2 01 0-10
2 02 10-25
0 03 25-50
1 04 50-100
2 05 greater than 100
35 98 No information

E107. Indicate number of experimental schools involved in project:

8 01 One
6 02 Two
3 03 Three
1 04 Four
2 05 Five
9 06 Six or more
13 98 No information

E108. Did the evaluator make any recommendations for project modification?

38 01 Yes
4 02 No

For items 109–116 indicate in which areas recommendations were made by circling the appropriate numbers. If answer to item 108 was "no," circle "not applicable."

	Modifications Recommended	*No Modifications Recommended*	*Not Applicable*
E109. Program objectives	01(*11*)	02(*28*)	99(*3*)
E110. Program design	01(*16*)	02(*23*)	99(*3*)
E111. Program implementation	01(*7*)	02(*32*)	99(*3*)
E112. Program management	01(*11*)	02(*28*)	99(*3*)
E113. Staff training or background	01(*24*)	02(*15*)	99(*3*)
E114. Evaluation procedures	01(*26*)	02(*13*)	99(*3*)
E115. Project/Community Involvement	01(*18*)	02(*21*)	99(*3*)
E116. Other (specify _____ _____)	01(*14*)	02(*25*)	99(*3*)

E117. How extensive were the changes recommended by the evaluator?

5 01 Major revisions recommended
33 02 Minor revisions only recommended
4 03 No changes recommended

Part IV: Results of Evaluation

For items 118–132, indicate whether the evaluator judged the project to be successful, unsuccessful, or neutral in given areas. "Not applicable" means this component not included in program. "No information" means the evaluator did not consider this aspect of the program.

	Successful	*Neutral*	*Unsuccessful*	*No Info*	*Not Applicable*
E118. Cognitive Outcomes: anticipated	01(*32*)	02(*9*)	03(*1*)	98(*0*)	99(*0*)
E119. Cognitive Outcomes: unanticipated	01(*2*)	02(*3*)	03(*1*)	98(*35*)	99(*1*)
E120. Parent/Community relations: anticipated	01(*34*)	02(*2*)	03(*3*)	98(*3*)	99(*0*)
E121. Parent/Community relations: unanticipated	01(*2*)	02(*0*)	03(*1*)	98(*38*)	99(*1*)
E122. Staff relations: anticipated outcomes	01(*10*)	02(*4*)	03(*0*)	98(*26*)	99(*2*)
E123. Staff relations: unanticipated outcomes	01(*0*)	02(*0*)	03(*0*)	98(*40*)	99(*2*)
E124. Costs: anticipated costs	01(*0*)	02(*6*)	03(*0*)	98(*34*)	99(*2*)
E125. Costs: unanticipated costs	01(*0*)	02(*0*)	03(*0*)	98(*40*)	99(*2*)
E126. Affective outcomes: anticipated	01(*18*)	02(*2*)	03(*1*)	98(*18*)	99(*3*)
E127. Affective outcomes: unanticipated	01(*2*)	02(*1*)	03(*0*)	98(*36*)	99(*3*)
E128. Staff training	01(*29*)	02(*5*)	03(*2*)	98(*5*)	99(*1*)
E129. Staff performance and Attitudes	01(*23*)	02(*5*)	03(*2*)	98(*11*)	99(*1*)
E130. Other: Anticipated outcomes (specify: _____)	01(*20*)	02(*3*)	03(*3*)	98(*9*)	99(*7*)
E131. Other: Unanticipated outcomes (specify: _____)	01(*2*)	02(*0*)	03(*1*)	98(*21*)	99(*18*)

Part V: Rater's Reaction to Evaluation

E132. How would you characterize the evaluator's overall judgment of the project?

22 01 Outstanding
20 02 Average
0 03 Below average
0 98 No information

E133. After reading the evaluation report, how would *you* judge the quality of the project?

9 01 Outstanding
26 02 Average
7 03 Below Average

E134. How would *you* describe the evaluator's attitude toward the project?

5 01 Overly positive; evaluator tended to make project look good
34 02 Positive, but fair
3 03 Critical, but fair
0 04 Hyper-critical (fault-finding exercise)

E135. In *your* opinion, how useful was this evaluation report for analyzing program effectiveness?

7 01 Of no use; I couldn't make any decisions on the basis of this report
17 02 Of very limited use; the evaluation report provides limited amounts of information
12 03 Fairly useful; the report provides most of the important information I would need
6 04 Very useful; the report provides all the information I would need

Project ID _____
Rater _____
Year of Project 1970–71 _____
Date _____

AUDIT DATA SHEET

Part I: Personal Characteristics of Auditor

A1. Auditor's professional affiliation:

7	01	University (specify _____)
3	02	Government Agency (i.e., regional lab) (specify _____)
18	03	Private agency (specify _____)
11	98	No information

A2. Highest degree attained:

0	01	BA
1	02	MA or MS
12	03	Ph.D. or Ed.D.
0	04	Other (specify _____)
26	98	No information

A3. Auditor's major field of study:

3	01	Education
0	02	Psychology
0	03	Sociology
0	04	Political Science
0	05	Business Management
1	06	Other (specify _____)
35	98	No information

Part II: Methods Used by Auditor

A4. Did the evaluator(s) report a model used in guiding his efforts?

8	01	Yes
27	02	No
4	03	No, but auditor reports an evaluation model

A5. Did the auditor use the evaluator's model as a standard for judging the quality of evaluation?

4	01	Yes
8	02	No
27	99	Not applicable if evaluator(s) did not report a model.

For items 6–47, indicate which sources of data the auditor used as a basis for his report by circling the appropriate number. Indicate *all* sources used.

		Used By Auditor	*Not Used By Auditor*
A6.	Site visits	01(37)	02(2)
A7.	Interviews in general	01(29)	02(10)
A8.	Interviews with teachers	01(18)	02(21)
A9.	Interviews with project director	01(23)	02(16)
A10.	Interviews with project staff and other project administration	01(22)	02(17)
A11.	Interviews with other school district personnel	01(6)	02(33)

	Used By Auditor	*Not Used By Auditor*
A12. Interviews with students	01(7)	02(32)
A13. Interviews with parents and community members	01(11)	02(28)
A14. Interviews with evaluation team	01(15)	02(24)
A15. Other interviews, (specify _____ _____)	01(1)	02(38)
A16. Auditor-administered tests or questionnaires	01(1)	02(38)
A17. Measurement tools of evaluator in general	01(37)	02(2)
A18. Evaluator-administered tests	01(37)	02(2)
A19. Evaluator-administered questionnaires	01(17)	02(22)
A20. Evaluator interview checklists	01(9)	02(30)
A21. Evaluator observation checklists	01(14)	02(25)
A22. Other evaluator measurement tools (specify _____)	01(5)	02(34)
A23. Raw data gathered by the evaluator	01(33)	02(6)
A24. Raw data: test results	01(32)	02(7)
A25. Raw data: questionnaire results	01(21)	02(18)
A26. Raw data: interview records	01(5)	02(34)
A27. Raw data: observation records	01(15)	02(24)
A28. Raw data: project reports for staff	01(6)	02(33)
A29. Raw data: other documentation	01(12)	02(27)
A30. Project proposal	01(12)	02(27)
A31. Evaluation proposal	01(7)	02(32)
A32. Evaluation contract	01(5)	02(34)
A33. Interim evaluation report	01(10)	02(29)
A34. Final evaluation report	01(36)	02(3)
A35. Continuation proposal	01(9)	02(30)
A36. Instructional materials in general	01(18)	02(21)
A37. Instructional materials: textbooks	01(10)	02(29)
A38. Instructional materials: curriculum	01(10)	02(29)
A39. Instructional materials: instructional planning guides	01(6)	02(33)
A40. Other instructional materials: (specify _____)	01(2)	02(37)
A41. Data from teacher training sessions	01(18)	02(21)
A42. Project products (e.g., specially designed tests)	01(24)	02(15)
A43. Archival information in general	01(12)	02(27)
A44. Project files: news clippings	01(1)	02(38)
A45. Project files: interstaff communications	01(4)	02(35)
A46. Project files: cumulative records of students	01(2)	02(37)
A47. Project files: letters from community members, School Board Minutes, etc.	01(2)	02(37)
A48. Other project file information (specify _____)	01(6)	02(33)
A49. Project financial records	01(3)	02(36)
A50. Correspondence with staff	01(2)	02(37)
A51. Correspondence with evaluation team	01(3)	02(36)
A52. Other sources of information used by auditor (specify _____)	01(7)	02(32)

Part III: Perceived Quality of Evaluation Procedures

Indicate which of the following components of an evaluation were discussed in the audit report. For each discussed, indicate the quality of that component as judged by the auditor using the following scale:

+3 High quality (done exceedingly well, could not be improved, auditor agreed strongly with evaluator's method)
+2 Above average quality (done well, auditor agreed with evaluator's method, possibly some minor improvements suggested)
+1 Average quality (evaluator's method acceptable to auditor, but could be improved considerably)
0 No judgment made (auditor neutral; auditor commented a task was performed, but made no judgments)
−1 Below average quality (done poorly, but acceptable to some degree; auditor disagreed somewhat with evaluator, noted several faults)
−2 Poor quality (done poorly, not acceptable; auditor has more fundamental disagreements with evaluator)
−3 Very poor quality (method inappropriate, important component missing, mental disagreement)

Components of Evaluation	*Perceived Quality*	$\bar{X}$	N
A53. General design	−3 −2 −1 0 +1 +2 +3	1.22	36
A54. Appropriateness of design	−3 −2 −1 0 +1 +2 +3	1.50	24
A55. Implementation of design	−3 −2 −1 0 +1 +2 +3	.90	21
A56. Identification of goals	−3 −2 −1 0 +1 +2 +3	.67	24
A57. Assessment techniques	−3 −2 −1 0 +1 +2 +3	.94	36
A58. Testing instruments	−3 −2 −1 0 +1 +2 +3	.77	35
A59. Administration of tests	−3 −2 −1 0 +1 +2 +3	.65	20
A60. Scoring of tests	−3 −2 −1 0 +1 +2 +3	.75	12
A61. Questionnaires	−3 −2 −1 0 +1 +2 +3	.61	18
A62. Observation schedule	−3 −2 −1 0 +1 +2 +3	1.10	10
A63. Observation checklists	−3 −2 −1 0 +1 +2 +3	.71	14
A64. Interview techniques	−3 −2 −1 0 +1 +2 +3	1.00	4
A65. Sampling procedures	−3 −2 −1 0 +1 +2 +3	.08	12
A66. Other documentation	−3 −2 −1 0 +1 +2 +3	.53	17
A67. Accuracy of gathered data	−3 −2 −1 0 +1 +2 +3	1.75	28
A68. Analysis of raw data	−3 −2 −1 0 +1 +2 +3	.30	33
A69. Appropriateness of data collection techniques	−3 −2 −1 0 +1 +2 +3	.97	32
A70. Interpretation of results of analysis	−3 −2 −1 0 +1 +2 +3	.32	34
A71. Format of evaluation report	−3 −2 −1 0 +1 +2 +3	.44	32
A72. Format of tables used to present data	−3 −2 −1 0 +1 +2 +3	.10	19
A73. Documentation of results	−3 −2 −1 0 +1 +2 +3	−.21	24
A74. Objectivity of report	−3 −2 −1 0 +1 +2 +3	1.64	11

Components of Evaluation	*Perceived Quality*	$\bar{X}$	N
A75. Appropriateness for intended audience	−3 −2 −1 0 +1 +2 +3	0.0	14
A76. Completeness of report	−3 −2 −1 0 +1 +2 +3	0.00	28
A77. Language	−3 −2 −1 0 +1 +2 +3	.36	11
A78. Oral presentations	−3 −2 −1 0 +1 +2 +3	0.00	2
A79. Other (specify _____ _____)	−3 −2 −1 0 +1 +2 +3	−1.83	6
A80. Formative evaluation activities	−3 −2 −1 0 +1 +2 +3	.47	15
A81. Monitoring of program implementation	−3 −2 −1 0 +1 +2 +3	.40	10
A82. Monitoring of progress toward objectives	−3 −2 −1 0 +1 +2 +3	.67	9
A83. Interim evaluation reports	−3 −2 −1 0 +1 +2 +3	1.00	7
A84. Recommendations for en-route improvements	−3 −2 −1 0 +1 +2 +3	−1.00	2
A85. Other (specify _____ _____)	−3 −2 −1 0 +1 +2 +3		0
A86. Other component of evaluation (specify) Criteria for success _____ _____)	−3 −2 −1 0 +1 +2 +3	−2.00	1

Part IV: Scope of Audit Report

For items 87–93, indicate whether or not the auditor agrees with the evaluator's assessment of the following:

	Did Agree	*Did Not Agree*	*No Info*	*Only auditor's consideration*
A87. Student learning (cognitive)	01(29)	02(5)	98(5)	99(0)
A88. Student attitudinal changes (affective)	01(14)	02(0)	98(25)	99(0)
A89. Parent/Community involvement	01(26)	02(3)	98(8)	99(2)
A90. Staff training and attitudes	01(24)	02(1)	98(12)	99(2)
A91. Unanticipated outcomes	01(3)	02(0)	98(35)	99(1)
A92. Program management	01(18)	02(2)	98(14)	99(5)
A93. Process objectives	01(15)	02(2)	98(18)	99(4)

A94. Did the audit report include a consideration of alternative strategies for conducting an evaluation?
If YES, Specify _____.

28 01 Yes
11 02 No

A95. Did the audit report include a discussion on the evaluator's recommendations for the future?

17 01 Yes
22 02 No

A96. Did the audit report include comments on the relationship between the evaluator and project director?

1 01 Yes
38 02 No

A97. Which category best characterizes the relationship between the evaluator and the project director as described by the auditor?

0 01 Overly close, resulted in evaluator bias
1 02 Cordial, but business-like
0 03 Less than cordial, but business-like
0 04 Hostile
0 05 Evaluator and project director same person
38 99 Not applicable if audit report included no comments about evaluator-project director relationship.

For items 98–101 indicate which of the following kinds of cost information was reported.

	Yes	*No*
A98. Evaluation costs	01(2)	02(37)
A99. Program costs	01(5)	02(34)
A100. Cost-benefit analysis	01(0)	02(39)
A101. Other (specify _____ _____)	01(0)	02(39)

A102. Did the audit report include any recommendations for future modifications?

38 01 Yes
1 02 No

For items 103–119 indicate in which areas recommendations were made. If *no* recommendations were made by the auditor, circle "Not Applicable."

	Recommended	*Not Recommended*	*Not Applicable*
A103. Formative evaluation activities	01(14)	02(25)	99(0)
A104. Program Management	01(10)	02(29)	99(0)
A105. Evaluation design	01(34)	02(5)	99(0)
A106. Staff attitudes and relations	01(0)	02(39)	99(0)
A107. Format of evaluation report	01(21)	02(18)	99(0)
A108. Community and parent involvement	01(8)	02(31)	99(0)
A109. Evaluation of long-term effects	01(9)	02(30)	99(0)
A110. Provisions for unanticipated outcomes	01(1)	02(38)	99(0)
A111. Data collection techniques	01(28)	02(11)	99(0)
A112. Data analysis techniques	01(24)	02(15)	99(0)
A113. In-service training	01(9)	02(30)	99(0)

	Recommended	*Not Recommended*	*Not Applicable*
A114. Dissemination of evaluation information	01(*10*)	02(*29*)	99(*0*)
A115. Program design	01(*10*)	02(*29*)	99(*0*)
A116. Program implementation	01(*2*)	02(*37*)	99(*0*)
A117. Program materials	01(*9*)	02(*30*)	99(*0*)
A118. Process objectives	01(*9*)	02(*30*)	99(*0*)
A119. Other (specify _____ _____)	01(*14*)	02(*22*)	99(*3*)

Part V: Rater's Reaction to Audit Report

A120. How would you characterize the auditor's judgment of the project?

17 01 Outstanding
21 02 Average
1 03 Below Average

A121. After reading the audit report, how would *you* describe the auditor's attitude with respect to the evaluation?

4 01 Overly positive, the auditor tended to make the evaluation look good
21 02 Positive, but fair
14 03 Critical, but fair
0 04 Hyper-critical (fault-finding exercise)

A122. In *your* opinion, how useful was this audit for analyzing evaluation effectiveness?

1 01 Of no use; I couldn't make any decisions on the basis of this report
18 02 Of limited use; the audit report provided limited amounts of information
17 03 Fairly useful; the report provided most of the important information I would need to make decisions
3 04 Very useful; the report provided all the information I would need

A123. On the basis of the audit report, how would you rate the program?

6 01 Outstanding
30 02 Average
3 03 Below Average

A124. On the basis of the audit report, how would you rate the auditor's opinion of the evaluation?

11 01 Outstanding
19 02 Average
9 03 Below Average

PROJECT DIRECTOR QUESTIONNAIRE

Project Director _____
Project Name _____
Grant Number _____
Year of Project 1970 to 1971 ____

Characteristics of Evaluation Team

PD1. To whom was the evaluation contracted?

17 01 single individual
24 02 evaluation "team" (i.e. an organization or more than individual) (specify _____)
1 98 not known

PD2. If there was an evaluation "team" which of the following best characterizes its composition?

19 01 several individuals working under the supervision of principal evaluator
3 02 several evaluators with distinct responsibilities
2 03 other (specify _____)
7 98 not known
11 99 not applicable

> A distinction is sometimes made between internal and external evaluators. An internal evaluator is typically a member of the project staff or a permanent employee of the school district. An external evaluator is an outside consultant called in exclusively for the evaluation of one project.

PD3. What was the position of the evaluator(s) with respect to the project?

16 01 internal only
20 02 external only
6 03 both internal and external components
0 04 other (specify _____)
0 98 not known

Characteristics of Evaluator

> In answering questions 4 through 8 please consider the sole evaluator or, in the case of a team, the principal evaluator.

PD4. Professional affiliation of the evaluator

12 01 university (specify _____)
4 02 government agency (specify _____)
12 03 private agency (specify _____)
11 04 project staff member (specify _____)
3 05 other (specify _____)
0 98 not known

PD5. What do you believe was the highest degree obtained by the evaluator?

3	01	BA or BS
18	02	MA or MS
21	03	PhD or EdD
0	04	other (specify _____)
0	98	not known

PD6. What do you believe was the evaluator's principal field of study?

20	01	education
11	02	psychology
0	03	sociology
1	04	political science
2	05	business management
5	06	other (specify _____)
2	98	not known
1	99	not applicable if only one candidate considered

Selection of Evaluator

PD7. If there was more than one candidate for the position of evaluator, who made the final selection?

4	01	district office
8	02	project director
5	03	district superintendent
13	04	selection committee (specify composition _____ _____)
7	05	other (specify _____)
0	98	not known
5	99	not applicable if only one candidate considered

PD8. How was the evaluator who was finally selected for the project located? (please check all that apply)

18	01	recommendation
4	02	school district evaluation unit or equivalent
10	03	known from previous professional association (specify _____)
0	04	state employee (specify _____)
1	05	private agency located by (specify _____ _____)
7	06	other (specify _____)
2	98	not known

PD9. Please indicate the time at which evaluation activities for the year '70 to '71 began.

32	01	before the program started for school year 1970 to 1971
8	02	during first half of the school year
1	03	during second half of the school year
0	04	after completion of 1970 to 1971 program
0	05	other (specify _____)
1	98	not known

Formative Activities

> In formative (en-route) evaluation activities, the evaluator studies the implementation of the program and the progress being made throughout the year, not just at the end.

PD10. Were formative evaluation studies conducted for your project?

- 2 01 no formative studies
- 23 02 some formative studies
- 16 03 extensive formative studies
- 1 98 not known

PD11. In what way was the information from formative evaluation studies useful? (please check all that apply)

- 26 01 confirmatory (confirmed what we thought was the situation) (yes—26 no—13)
- 23 02 informative (discovered problems or progress of which we were not aware) (yes—23 no—16)
- 32 03 guided project staff in making modifications in the program (yes—32 no—7)
- 3 04 other (specify _____) (yes—3 no—36)
- 0 05 not useful (yes—0 no—39)
- 1 99 not applicable because en-route evaluation studies were not performed
- 2 98 (no answers)

PD12. Which procedure was used most often in developing criteria for program objectives?

- 15 01 criteria accepted directly from the objectives as stated in the project proposal
- 1 02 the evaluator(s) selected criteria
- 21 03 criteria developed jointly by the evaluator(s) and project staff
- 0 04 criteria based on pilot studies
- 4 05 other (specify _____)
- 1 98 not known

PD13. Approximately how many on-site visits did the evaluator(s) make in the last school year?

- 0 01 none
- 11 02 1—5
- 31 03 more than 5 (specify if possible _____ _____)
- 0 98 not known

PD14. Who performed the following activities? Please circle the number under the appropriate column heading. Indicate "not known" only if the activity *was* performed but it is not known by whom. If the activity was *not* performed please circle "99." (Distinction between "internal" and "external" is given on page 2.)

Evaluation Activity	*external only*	*internal only*	*both*	*not known*	*activity not performed*
PD14 a) program observation	01(8)	02(11)	03(23)	04(0)	99(0)
PD14 b) student testing	01(5)	02(31)	03(5)	04(0)	99(1)

Evaluation Activity	*external only*	*internal only*	*both*	*not known*	*activity not performed*
PD14 c) examination of records	01(*9*)	02(*13*)	03(*20*)	04(*0*)	99(*0*)
PD14 d) data analysis	01(*21*)	02(*10*)	02(*10*)	04(*0*)	99(*1*)
PD14 e) data reporting	01(*12*)	02(*18*)	03(*12*)	04(*0*)	99(*0*)

Evaluation Reporting

PD15. How were the evaluation reports distributed by the evaluator?

29 01 all reports sent to project director
5 02 reports sent to federal sponsors and project director
2 03 reports sent to a variety of interest groups
5 04 other (specify _____)
1 98 not known

PD16. Which groups received the final evaluation reports?

	received report	*did not receive report*	*not known*
a) project director	01(*41*)	02(*0*)	98(*1*)
b) teachers	01(*29*)	02(*4*)	98(*9*)
c) parents	01(*15*)	02(*12*)	98(*15*)
d) students	01(*1*)	02(*27*)	98(*14*)
e) school board	01(*33*)	02(*5*)	98(*4*)
f) federal sponsors	01(*41*)	02(*0*)	98(*1*)
g) news media	01(*7*)	02(*18*)	98(*17*)
h) others (specify _____)	01(*29*)	02(*0*)	98(*13*)

PD17. Did you or your representative assist in writing the *final* evaluation report?

15 01 yes, I assisted
3 02 yes, my representative assisted (specify who _____)
2 03 other (specify _____)
22 04 no

PD18. Were any reports specially written for the community or other special groups?

23 01 yes
17 02 no
2 98 not known

PD19. If yes, what purpose was served by having the reports specially written? (please check all that apply)

15 01 reports were condensed (yes—15 no—6)
19 02 reports were in non-technical language (yes—19 no—3)
12 03 reports were in a language other than English (yes—12 no—10)
15 04 reports were aimed to increase community's confidence in the project (yes—15 no—7)
20 99 (if question 18 is '02' or '98', '99' is checked)

General Reactions to Evaluation

PD20. Please rate the extent to which the evaluation was of *assistance to you* in the indicated areas. Circle "99" if the activity was not performed.

	not useful	*moderately useful*	*useful*	*very useful*	*not performed*	*AV*
a) identifying possible problem areas	01(2)	02(8)	03(11)	04(19)	99(2)	3.17
b) preparing reports	01(1)	02(8)	03(14)	04(16)	99(3)	3.15
c) helping to develop criteria	01(2)	02(10)	03(13)	04(15)	99(2)	3.02
d) general recommendations for program changes	01(2)	02(8)	03(13)	04(17)	99(2)	3.12
e) changing the pace of instruction	01(6)	02(14)	03(14)	04(6)	99(2)	2.50
f) changing teaching procedures	01(3)	02(14)	03(14)	04(9)	99(2)	2.65
g) changing in-service activities	01(8)	02(3)	03(15)	04(14)	99(2)	2.87
h) changing personnel	01(20)	02(10)	03(3)	04(1)	99(8)	1.56
i) changing community relations activities	01(12)	02(9)	03(15)	04(2)	99(4)	2.18
j) changing program management	01(13)	02(7)	03(11)	04(7)	99(4)	2.32
k) changing instructional materials	01(9)	02(9)	03(15)	04(7)	99(2)	2.50

PD21. As a project director which would you prefer?

14 01 an internal evaluator only
1 02 an external evaluator only
25 03 both an internal and an external evaluator
0 04 no evaluator(s)
1 05 no preference
1 06 other (specify _____)

PD22. Please indicate the extent of your agreement with the following statements by circling one response:

	strongly disagree	*disagree*	*no opinion*	*agree*	*strongly agree*	
a) The evaluation for 70–71 influenced decisions that were made to modify the program during the year 70–71	01(5)	02(8)	03(1)	04(21)	05(6)	98(1)
b) The final evaluation report for 70–71 influenced decisions that were made to modify the program the following year 71–72	01(4)	02(3)	03(0)	04(23)	05(11)	98(1)

PD23. Please indicate how influential each of the following sources of information was in your decision making:

	very influential	*influential*	*somewhat influential*	*not very influential*	*not at all*
a) The evaluation for 70–71 influenced decisions that were made to modify the program during the year 70–71	01(*7*)	02(*13*)	03(*7*)	04(*10*)	05(*4*) 98(*1*)
b) The final evaluation report for 70–71 influenced decisions that were made to modify the program the following year (71–72)	01(*12*)	02(*17*)	03(*6*)	04(*5*)	05(*1*) 98(*1*)
c) Recommendations of federal monitors influenced decisions to modify the program the following year (71–72)	01(*8*)	02(*9*)	03(*9*)	04(*6*)	05(*9*) 98(*1*)
d) Recommendations of auditors influenced decisions to modify the program the following year	01(*8*)	02(*18*)	03(*11*)	04(*3*)	05(*2*) 98(*0*)

PD24. Please indicate the degree of correspondence between the federal monitors' recommendations concerning your program and those provided by the following sources:

	very strong correspondence	*strong correspondence*	*average correspondence*	*weak correspondence*	*no correspondence*
a) The federal monitors' recommendations corresponded with observations in the evaluation report	01(*3*)	02(*9*)	03(*19*)	04(*3*)	05(*5*) 98(*3*)
b) The federal monitors' recommendations corresponded with observations in the audit report	01(*3*)	02(*11*)	03(*14*)	04(*5*)	05(*6*) 98(*3*)

Appendix 2:
CODE VALUES FOR DICHOTOMIZED VARIABLES

DICHOTOMIZATION OF ORIGINAL VARIABLES

Variables	Coding	
A63 through A80	[-3, -2, -1, 0] = 0	[1, 2, 3] = 1
A120, A123, A124	[02, 03] = 0	[01] = 1
E1	[01, 98] = 0	[02] = 1
E3	[01, 98] = 0	[02, 03, 04, 05] = 1
E16	[98] = 0	[01, 02, 03, 04] = 1
E21, E30	[02] = 0	[01] = 1
E62 through E65	[98] = 0	[01, 02] = 1
E76 through E79	[02] = 0	[01] = 1
E80 through E88	[02, 99] = 0	[01] = 1
E91	[02, 99] = 0	[01] = 1
E96	[02] = 0	[01] = 1
E98	[01, 02, 99] = 0	[03, 04] = 1
E99	[03, 04, 99] = 0	[01, 02] = 1
E100	[03] = 0	[01, 02] = 1
E101, E102	[02] = 0	[01] = 1
E103	[98] = 0	[01, 02, 03, 04] = 1
E107	[98] = 0	[01, 02, 03, 04, 05, 06] = 1
E120, E126, E129, E132	[98, 99, 02, 03] = 0	[01] = 1
E133	[02, 03] = 0	[01] = 1
E135	[01, 02] = 0	[03, 04] = 1
PD1	[1, 98] = 0	[02] = 1
PD3	[1] = 0	[02, 03, 04, 98] = 1
PD5	[01, 02, 04, 98] = 0	[03] = 1
PD10	[01, 02, 98] = 0	[03] = 1
PD11a, PD11b, PD11c	[98, 99, 00] = 0	[01] = 1
PD13	[01, 02, 98] = 0	[03] = 1
PD20a through PD20k	[01, 02, 99] = 0	[03, 04] = 1
PD22a, PD22b	[01, 02, 03, 98] = 0	[04, 05] = 1
PD23a, PD23b	[03, 04, 05, 98] = 0	[01, 02] = 1