

**PROGRAM EVALUATION IN EDUCATION: SCHOOL DISTRICT PRACTICE IN
BRITISH COLUMBIA**

by

TRISHA G. WILCOX

B.Ed.(Hons.), The University of Bristol, U.K., 1971
M.A., The University of the Americas, Mexico, 1980

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF EDUCATION**

in

THE FACULTY OF GRADUATE STUDIES

Department of Administrative, Adult, and Higher Education

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1989

© Trisha G. Wilcox, 1989

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Administrative, Adult & Higher Education
The University of British Columbia
Vancouver, Canada

Date April 20, 1989

ABSTRACT

This study addressed a problem frequently noted in the evaluation literature, namely the lack of an empirically derived knowledge base about how evaluation is practiced. It examined the practice of program evaluation in school districts in British Columbia.

An historical account of the development of the evaluation literature led to a critique of the way the field is ordered, with the result that issues in evaluation, rather than approaches to the evaluative task, were used to structure the framework for the empirical research described. Four general questions that made up this framework were "Evaluation – to what end?" "Evaluation – by what means?" "Evaluation – for whom and by whom?" and "Evaluation – with what conclusion?" The framework was applied to written evaluation reports produced in school districts across the province. The results of the content analysis of these reports provided a description of school district evaluation which has not hitherto been available and which, in turn, served as a means of adding to the existing evaluation knowledge base.

A further analysis of the numerous specific findings led to the identification of four salient aspects of program evaluation as practiced in British Columbia school districts. The aspects identified were stakeholder participation in the conduct of evaluations; the role of the evaluator; the purposes of evaluation, and the identification of evaluation criteria. When these aspects are considered together it is difficult to avoid the conclusion that program evaluation practices maintain and reinforce the status quo rather than challenge it.

Two kinds of outcomes of the study were seen as important. First, it provides a sound basis for the creation of guidelines for writing evaluation reports in school districts. A number of such guidelines were suggested. Second, the study suggests areas in which further research might usefully be conducted both to amplify the picture discovered in this study and to explore what, if any, role is played by evaluation in the adoption of change in school systems.

TABLE OF CONTENTS

Abstract	ii
List of Tables	vii
Acknowledgements	x
Chapter I. BACKGROUND AND PURPOSE OF THE STUDY	1
A. Background	1
B. Purpose	6
C. Overview of the Thesis	7
Chapter II. THE LITERATURE OF PROGRAM EVALUATION	8
A. The Development of the Evaluation Literature	8
B. The Ordering of the Evaluation Literature	17
1. Conceptions of Evaluation	17
2. Classification Schemes	22
C. From Approach-based Writing to Issue-based Investigation	31
D. Summary	33
Chapter III. THE FRAMEWORK AND RESEARCH DESIGN	34
A. The Framework	34
1. Identification of Basic Questions	35
2. Developing the Framework	37
3. Modifying the Specific Questions	39
B. Research Design	47
1. The Method of Content Analysis	47
a. Content Analysis	47
b. Reliability and Validity	50
2. Data Collection	53
3. The Recording of Content	56
4. Delimitations of the Study	57
C. Summary	59
Chapter IV. THE DEVELOPMENT OF RULES FOR THE CODING OF CONTENT	61
A. Evaluation — To What End?	62
1. How was Evaluation Defined?	62
a. Judgement [No Supporting Information]	63
b. Judgement [Some Supporting Information]	64
c. Judgement [Much Supporting Information]	64
d. Provision of Information [Some Judgement]	65
e. Provision of Information [No Judgement]	66
2. What were the Intents of the Evaluation?	66
a. Purpose	67
b. Difficulties in Coding Purpose Statements	72
c. Function	73
3. Why was the Evaluation Undertaken?	75
4. What was the Object Evaluated?	77

a. Type of Object Evaluated	77
b. Additional Characteristics of Objects Evaluated	79
B. Evaluation — By What Means?	83
1. What Kinds of Information Regarding Each Object were Reported?	84
a. Source of the Information	84
b. Nature of the Information	86
2. What Criteria were Used to Judge Merit and Worth? ...	90
a. Location of Criteria	91
b. Source of Criteria	94
c. Nature of Criteria	95
3. What Methods of Inquiry were Used in the Evaluation?	98
a. Approach Ascribed by Evaluators	98
b. Types of Data Collection Techniques	100
c. Number of Data Collection Techniques	103
C. Evaluation — For Whom and By Whom?	103
1. To Whom was the Report Submitted?	103
2. Who were the Designated Evaluators?	104
a. Designated Evaluators	105
b. Number of Evaluators	106
c. Advisory Structure	106
D. Evaluation — With What Conclusion?	106
1. What Recommendations (if any) were Made?	107
a. Assigning Recommendations to Action Categories	107
b. Assigning Recommendations to Target Areas	110
c. Difficulties in Coding Recommendations	117
E. General Information	119
F. Summary	120

Chapter V. THE CONTENT OF THE EVALUATION REPORTS	121
A. General Information	121
B. Reporting the Content: Frequencies	124
1. Evaluation — To What End?	124
a. How was Evaluation Defined?	124
b. What were the Intents of the Evaluation?	126
c. Why was the Evaluation Undertaken?	130
d. What was the Object Evaluated?	131
2. Evaluation — By What Means?	136
a. What Kinds of Information Regarding Each Object were Reported?	136
b. What Criteria were Used to Judge the Merit and Worth of the Object?	145
c. What Methods of Inquiry were Used in the Evaluation?	148
3. Evaluation — For Whom and By Whom?	151
a. To Whom was the Report Submitted?	151
b. Who were the Designated Evaluators?	152
4. Evaluation — With What Conclusion?	158

a. What Recommendations (if any) were Made?	158
C. Reporting the Content: Cross-tabulations	162
1. Expected Associations	163
2. Unexpected Associations	168
D. Summary of the Content of the Reports	171
 Chapter VI. SALIENT ASPECTS OF PROGRAM EVALUATION IN BRITISH COLUMBIA SCHOOL DISTRICTS	 177
A. The Identification of Salient Findings	178
B. Four Salient Aspects of the Findings of the Study	180
1. Stakeholder Participation	180
2. The Role of the Evaluator	182
3. Purposes of Evaluation	185
4. The Identification of Criteria	186
C. Conclusion: On Not Rocking the Boat	188
 Chapter VII. SUMMARY, CONCLUSIONS, AND IMPLICATIONS	 191
A. Summary	191
B. Conclusions	193
1. Usefulness of the Framework and Coding Categories ...	194
2. The Picture of Program Evaluation: On Not Rocking the Boat	196
C. Implications	198
1. Guidelines for Writing Evaluation Reports	199
2. Implications for Further Research	204
 REFERENCES	 209
 APPENDIX 1. Correspondence	 221
 APPENDIX 2. Reports Excluded in the First Stage of Analysis	 224
 APPENDIX 3. Coding Instrument	 226
 APPENDIX 4. Coding Categories Showing Frequencies Found in Data Base	 248

LIST OF TABLES

TABLE 2.1 — Authors of Classification Schemes and Evaluation Approaches Identified	24
TABLE 2.2 — Authors of Classification Schemes and their Bases for Comparison among Approaches	25
TABLE 4.1 — Coding Categories for Content Pertaining to the Definition of Evaluation	63
TABLE 4.2 — Coding Categories for Content Identifying the Functions and Purposes of Evaluation	68
TABLE 4.3 — Coding Categories Showing the Reasons for Evaluation	76
TABLE 4.4 — Coding Categories for Content Identifying the Type of Object Evaluated	78
TABLE 4.5 — Coding Categories for Content Referring to Additional Characteristics of Objects Evaluated	80
TABLE 4.6 — Coding Categories for the Identification of the Source of Information Reported	85
TABLE 4.7 — Coding Categories for Content Pertaining to the Nature of Information Reported	87
TABLE 4.8 — Coding Categories for Content Identifying the Location in the Document of Criteria for Judgement	91
TABLE 4.9 — Coding Categories for the Content Identifying the Source of Criteria for Information	94
TABLE 4.10 — Coding Categories for the Content Identifying the Nature of Criteria for Judgement	96
TABLE 4.11 — Coding Categories for Content Pertaining to Methods of Inquiry	99
TABLE 4.12 — Coding Categories for Content Identifying the Recipients of Evaluation Reports	104
TABLE 4.13 — Coding Categories for Content Identifying the Designated Evaluators	105
TABLE 4.14 — Coding Categories for the Analysis of Recommendations	108
TABLE 5.1 — Number of School Districts in each of Five Size Categories and Six Categories of Number of Reports Submitted	123

TABLE 5.2 – Definition of Evaluation	125
TABLE 5.3 – Purposes of Evaluation	127
TABLE 5.4 – Formative and Summative Functions of Evaluation	129
TABLE 5.5 – Reason for Evaluation	131
TABLE 5.6 – Object of the Evaluation	132
TABLE 5.7 – Permanence of Object Evaluated	133
TABLE 5.8 – Aspect of Object Evaluated	134
TABLE 5.9 – Organizational Aspects of Object Evaluated	135
TABLE 5.10 – Source of Information Reported about Each Object	137
TABLE 5.11 – Sources Used for the Provision of Information about Each Object ...	139
TABLE 5.12 – Nature of Information Reported about Each Object	141
TABLE 5.13 – Types of General and Specific Information Reported about Each Object	143
TABLE 5.14 – Areas of Focus of Reports Containing Both Opinion and Descriptive Information Reported about Each Object	144
TABLE 5.15 – Source and Nature of Criteria Used to Judge the Merit and Worth of the Object Evaluated (n=64)	146
TABLE 5.16 – Methods of Inquiry Made Explicit by Evaluators	149
TABLE 5.17 – Data Collection Techniques Used in the Evaluations	149
TABLE 5.18 – Number of Data Collection Techniques Used in the Evaluations	150
TABLE 5.19 – Recipients of Evaluation Reports	151
TABLE 5.20 – Position-based Source of Designated Evaluators	153
TABLE 5.21 – Source Groups of Designated Evaluators	154
TABLE 5.22 – Number of Designated Evaluators	155
TABLE 5.23 – Number of Members and Composition of Advisory Groups	157
TABLE 5.24 – Actions Recommended	158

TABLE 5.25 — Combined Action Recommended and Target Area of Recommendations	160
TABLE 5.26 — Combined Action Recommended and Combined Target Area of Recommendations	161
TABLE 5.27 — Definition of Evaluation and Number of Purpose Statements in which Improvement and Change, Decision Making, and Awareness and Knowledge were identified as Purposes	164
TABLE 5.28 — Definition of Evaluation and Recommendations	166
TABLE 5.29 — Definition and Function of Evaluation	166
TABLE 5.30 — Purpose and Function of Evaluation	167
TABLE 5.31 — Purpose of Evaluation and Recommendations	168
TABLE 5.32 — Type of Object Evaluated and Function of Evaluation	169
TABLE 5.33 — Type of Object Evaluated and Designated Evaluators when the Definition of Evaluation is Judgement	170
TABLE 5.34 — Designated Evaluators and School District (SD) Size	171

ACKNOWLEDGEMENTS

Without the help of colleagues and friends, completion of this dissertation would have remained Sisyphean. However, unlike the legendary rock, my academic boulder reached the top of the hill and remained there. For this reason I am indebted to a number of people.

I would like to thank the members of my research committee: Dr. Graham Kelsey, my research supervisor, for his substantive comments and his inimitable forbearance; and committee members, Dr. Ian Housego and Dr. Tom Sork, for their encouragement and suggestions for improving the thesis.

I would like to thank Charlotte Coombs for levering me out of a particularly deep rut and for spurring me on to scale the summit. I would like to thank Jay Handel for the time and effort he put into the computer assistance he gave me, and Lesley Bellamy for her constant support. Thanks also go to Marina Koskinen and to the annual and perennial inhabitants of South Staff Office Block for their warmth and understanding over the last few years. In addition, I would like to thank my parents for the financial care packages and for their unfailing patience.

Finally, I would like to acknowledge the school district personnel who responded to my initial inquiries and who submitted written evaluation reports for analysis.

Thank you.

CHAPTER I.

BACKGROUND AND PURPOSE OF THE STUDY

A. BACKGROUND

This study was an attempt to deal with a problem widely acknowledged in the literature, namely the lack of empirically based knowledge of the practice of program evaluation. The growing demand for accountability since the mid sixties has been reflected by a dramatic increase in program evaluations in the United States and Canada. This increase has been accompanied by a literature replete with discussions of conceptual approaches to evaluation and with a number of case studies of evaluations and evaluation practices. What the literature has rarely, if ever, included is any analytical description of evaluation practice, particularly in large-scale settings such as provincial jurisdictions.

Program evaluation did not emerge as a field of study in English-speaking North America until the mid 1960s, when civil rights and the education of children perceived as disadvantaged were major concerns of the United States Congress. A number of government projects designed to promote innovation and to provide equality of opportunity were introduced. The Elementary and Secondary Education Act (ESEA), passed in 1965, resulted in the introduction into the school system of numerous programs which were intended to overcome the injustices associated with poverty and race. To ensure that educators would become accountable for the federal funds they received, these programs included an evaluative requirement. Although this demand for evaluation did result in the completion of a large number of evaluative studies, rarely were they exemplary. Not only were there insufficient numbers of educators with the

necessary personal expertise in evaluation, but also appropriate strategies for program evaluation were lacking (Marcus and Stickney, 1981).

As a result of the perception that the methods of evaluation in use at the time were inadequate, scholars began to develop a number of alternatives (Scriven, 1967; Stake, 1967; Hammond, 1969; Provus, 1969, 1971; Stufflebeam *et al* 1971), and program evaluation evolved as a field of study. Additional perspectives continued to emerge. Among these were goal-free evaluation (Scriven, 1972, 1973); adversary evaluation (Owens, 1973); the naturalistic approach (Guba, 1978); the anthropological approach (Koppelman, 1979); connoisseurship and criticism (Eisner, 1976, 1979a); responsive evaluation (Stake, 1975a, 1975b, 1976a); illuminative evaluation (Parlett and Hamilton, 1976); the judicial method (Wolf, 1979) and a variety of strategies taken from the realms of such subject areas as geography and journalism (Smith, 1981a). More recently, the literature has reflected a trend away from the use of discrete approaches and towards the use of a variety of evaluative methods, depending on the particular evaluation context (Bryk and Light, 1981; Cronbach, 1982; Borich, 1983).

The burgeoning literature has spawned a plethora of both conceptions of evaluation and classification schemes. However, reports of empirical studies of evaluation have not been as prevalent in the literature as the discussions of conceptual approaches to the evaluative task. For the most part, those empirical studies that are reported have focussed on accounts of evaluation studies done for particular research purposes such as testing hypotheses, or on investigations of single evaluation studies done for particular client groups. Although the literature does offer generic suggestions for carrying out evaluation studies, there is minimal evidence of empirical research on the process of

evaluation as it actually takes place in practice.

One of these areas where little is known about the practice of program evaluation is in school districts. As practitioners involved in evaluating school district programs rarely write for publication, there is little information on program evaluation in school districts reported in the literature. Individual districts may have a great deal of information in their district offices, but such information is not made readily accessible in that their findings may be confidential and are often intended for internal use only.

In 1980, Smith observed:

We need research on evaluation; we especially need grounded, empirical studies of evaluation practice. We have almost no descriptive information on the practice of evaluation, few field studies of evaluation impact, and scant attention to the empirical study of evaluation method (385).

Now, eight years later, Smith's observation remains true. There is still a dearth of empirical research on evaluation practice. For example, in their search for research articles concerned with program evaluation in United States school districts over the last twenty years, Stufflebeam and Welch (1986) found that of the 150 articles they collected, only thirty-four presented the results of empirical research. The authors assert that the published research is spotty and inconclusive and that there is very little on the evaluation process per se. They echo Smith's call for a solid published research base, suggesting that although the conceptual development of the field should continue, it should "be buttressed by a stronger program of empirical inquiry" (1986:166). More particularly, Stufflebeam and Welch identify the need for research which addresses significant questions concerned with the evaluation process, such as the kinds of objects which are evaluated, the purposes of the evaluation, the audiences which are served by the evaluation, the evaluation methods, the people who participate

in the evaluation, the patterns of organization and funding for evaluation, as well as questions about the utilization of the evaluations together with the factors which are associated with the influence of the evaluation reports.

The empirical study of evaluation practice is necessary for both the improvement of practice and the advancement of theory (Smith, 1979, 1980, 1981b; Stufflebeam and Welch, 1986). Lawler claims that in order to contribute to theory and practice, research studies must be grounded in the workplace and must deal with issues pertinent to practitioners. He argues that research:

must help practitioners understand organizations in a way that will improve practice, and must contribute to a theoretical and scientifically useful body of knowledge about organizations (1985:2).

Lawler suggests that to do this researchers would be well advised to focus on the study of practice.

A detailed empirical examination of the evaluation process can potentially enhance both evaluation research and evaluation practice. It can enhance evaluation research by providing data on which hypotheses can be tested. It can also provide direction for further research by suggesting areas within which new hypotheses might fruitfully be generated; for example, hypotheses about the kinds of evaluation designs used, about the kinds of questions evaluations address, or about the place of evaluation in innovation. Shavelson (1988) suggests that educational research contributes to practice by challenging and changing the way practitioners think about problems and carry out their tasks. Hills and Gibson (1988) have suggested that practitioners would benefit from witnessing their actions and asking questions about the way they carry out their

tasks. Empirical data about how evaluation is done can enhance evaluation practice by providing evaluators with a broad view of common practice with which to compare their own. They might modify their evaluation practices accordingly and avoid pitfalls to which others have fallen prey. Such studies are also potentially useful to clients, acquainting them with a range of alternative processes from which to choose; and to administrators who are required to make decisions about evaluation policies and procedures.

To achieve such outcomes, it is necessary to raise the question of where and how to examine the evaluation process. The answer to *where* must clearly be one which permits the examination of a fairly wide range of evaluations. Since the interest for the present study is school district evaluation, and since all districts in a province are subject to the same provincial statutes, it is useful to think of one provincial jurisdiction as the locus of the examination.

As to *how* the process is to be examined, it is clear that there are problems in observing evaluation as practiced in an entire provincial jurisdiction. It is simply not feasible, at least within the resource limitations of a doctoral study, to observe the number and variety of activities associated with evaluation in seventy-five school districts. Most formal evaluations, however, culminate in a written report. These reports reflect both the evaluation processes and the nature of the objects of evaluation, as well as particular evaluator or client preferences about the kinds of information considered worthy of inclusion. These documents are a particularly useful source of information because they represent how the evaluators think about evaluation and have chosen to portray it. Ideas are usually thought through carefully before they are

committed to paper. Thus, the documents contain information that the evaluators have reflected on and have chosen to include. The evaluators have screened out what they consider to be irrelevant and in this way have illuminated the elements of evaluation that they consider important. It is plausible to argue, then, that an analysis of the written reports of evaluation would be a useful and feasible way of investigating empirically the practice of evaluation.

In order to examine program evaluation as reported in these documents, it was necessary to find a way to select data, focus and reduce them into a manageable form so that the data could be displayed and conclusions drawn. This requires a framework, both for examining the process of program evaluation as reported in the documents, and for comparing the accounts in order to identify similarities and differences. However, as no framework was available, it became necessary to develop one. This study, therefore, begins with the development of a framework and moves to its application. These stages are made explicit in the following statement of the purpose of the study.

B. PURPOSE

The purpose of the study was to examine how program evaluation, as documented, is done in school districts in British Columbia. The study consisted of three stages:

1. developing a framework for the analysis of evaluation practices;
2. using this framework to examine school district practices in British Columbia so as to provide a description of program evaluation in the province; and
3. using this knowledge of program evaluation practice as a means of adding to the existing evaluation knowledge base.

C. OVERVIEW OF THE THESIS

Chapter I has sketched the background to, and the purpose of, the study. Chapter II contains a review of the literature on program evaluation. A history of the development of the evaluation literature is given and various ways of ordering the different evaluation approaches are discussed. The chapter ends with a list of questions that changes the focus of the discussion from approaches to evaluation to issues in evaluation. The first part of Chapter III outlines how four general questions and the list of questions from Chapter II became the framework for the study. The second part of Chapter III describes the method of analyzing written communication known as content analysis and provides the research design. Chapter IV explains how coding categories were inductively derived from the content of the reports; it explains the rules for the coding of content and shows how the program evaluation reports were assigned to these categories. Chapter V outlines the detailed findings of the study. Chapter VI takes a broader view of these findings and argues that evaluation in British Columbia school districts seems to serve some purposes more than others. Chapter VII provides a summary, conclusions and implications for theory and practice, and gives suggestions for further research.

CHAPTER II.

THE LITERATURE OF PROGRAM EVALUATION

“Evaluation is a very peculiar breed of cat. The considerable charm of each of a dozen radically different models for it, . . . can only be explained by the fact that it is a chimerical, Janus-faced and volatile being” (Scriven, 1983:256).

The creation of a framework for the analysis of evaluation practice cannot begin *in vacuo*. An understanding of the literature of the field is necessary not only for better reader comprehension, but also, and more importantly, as a basis for framework building. Accordingly, this chapter begins with a history of the evaluation literature, tracing its progression from the 1920s when it was small and narrowly focussed to its present voluminous and wide-ranging state. The second section reviews the literature by discussing various conceptions of evaluation, and by describing a number of classification schemes, each of which provides a different way of ordering the field. In the final section, a change in emphasis from approach-based evaluation to issue-based evaluation is suggested as a useful way of identifying a framework for structuring empirical research in program evaluation.

A. THE DEVELOPMENT OF THE EVALUATION LITERATURE

The North American literature on program evaluation has grown considerably over the last half century. In contrast with the narrowness of its beginnings in the 1920s and 1930s when the testing movement was flourishing, when evaluation was equated with testing, and when the literature focussed on ways of measuring individual differences, the recent literature shows that evaluation has evolved into “a field of intellectual

endeavor complete with its own theorists, controversies, journals and conferences” (House, 1986:5).

The work of Tyler (1942, 1949) was instrumental in initiating the change in the way that evaluation was conceptualized and carried out. The seminal work, *Appraising and Recording Student Progress*, (Smith and Tyler, 1942) contained a discussion of The Eight Year Study which began in 1933 and was designed to encourage schools to innovate without jeopardizing students' chances to be admitted to college. Tyler's approach to educational evaluation centred on objectives and measured the outcomes of schooling in terms of the attainment of these objectives. At this time also, school accreditation¹ increased in prominence (Madaus, Scriven and Stufflebeam, 1983). In contrast with the emphasis on outcomes expressed in Tyler's work, accreditation tended to emphasize process.

These three foci, standardized testing, the attainment of performance objectives and accreditation continued to be important in the field of educational evaluation. However, it was not until a quarter of a century later that other methods of evaluating educational programs achieved prominence. Educators had other priorities; for example, in the years following the second world war, they were preoccupied with accommodating and coping with the increasing numbers of students entering the school system. Even at the end of the 1950s when Sputnik 1 was launched and resources were poured into math and science programs, there were no substantial attempts to assess their results. In 1963, Cronbach published an article in which, for the first

¹ Accreditation is the means by which approval is granted to an institution by a governing organization. The approval usually carries authorization for the institution to continue to offer a particular program or programs.

time, the suggestion was made that evaluation be reconceptualized and viewed as a process for gathering useful data to be used in the revision and improvement of programs. However, it was not until two years later that this publication was recognized as a landmark and the literature began to reflect the increasing emphasis on the importance of program evaluation in education.

As noted in Chapter 1, legislative requirements (ESEA, 1965) resulted in the annual evaluation of programs funded under the Act. The educational community, however, was not prepared; individuals from a variety of fields became involved but their actions tended to reflect their different disciplinary backgrounds rather than the demands of the actual evaluation situation. Given the diversity of backgrounds of those involved in evaluation, as well as the lack of available information about how to evaluate programs, it is hardly surprising that there was ambiguity, uncertainty, and disagreement in the field with regard to what evaluation involved and how it should proceed. The literature, which increased dramatically during the years following the ESEA legislation, reflected this diversity and turmoil. Attention was drawn to the proliferation of conceptions and technologies by the introduction of collections of seminal works such as *Educational Evaluation: Theory and Practice* (Worthen and Sanders, 1973) which, by exposing the many diverse approaches, pointed out the lack of underlying coherence within the field.

Another factor which was instrumental in the expansion of the evaluation literature of the 1960s was the development of social science methods, which, together with the advances in computer technology, enabled large-scale, relatively complex studies to be carried out. These government-sponsored studies were intended to provide legislators

with sufficient data to enable decisions to be based on valid empirical evidence about the effects of social programs (Rivlin, 1971). Of particular significance in the literature at this time was an article written by Campbell (1969), in which he suggested that government policy making could best be viewed as social experimentation. If policy reforms were seen as experiments, then policy makers could institutionalize or abandon policies according to the results of the studies. Campbell's work was of great importance in its own right and also served to give rise to a number of alternative approaches to evaluation in which reliable data could be collected without applying all the controls of a laboratory situation. Experimental designs (Campbell and Stanley, 1966) were extremely important but among the proponents of the experimental approach, quasi-experimental designs gradually became more popular (Riecken and Boruch, 1974; Cook and Campbell, 1979). However, because of the difficulty in controlling societal experiments and the ethical questions they raised, and the emphasis on outcomes to the exclusion of all else, the literature had begun to reflect a dissatisfaction with both experiments and quasi-experiments in evaluation (Guba, 1969), and a number of alternatives began to appear.

In "The countenance of educational evaluation" (1967), Stake argued that evaluation involved much more than the analysis of information about program outcomes. He suggested, alternatively, that the evaluation process was one of description and judgement, and provided a framework which evaluators could use to collect information on "antecedents" (conditions prior to the introduction of the program), "transactions" (everything that comprises the program in action) and "outcomes" (program results).

Stufflebeam *et al* (1971) also objected to the emphasis on outcomes. As an alternative,

Stufflebeam suggested the Context-Input-Process-Product model (CIPP). He viewed evaluation as a continuing process intended to provide information to decision-makers. This process involved answering questions about "context" (to identify objectives to be met); "input" (to identify a feasible way of meeting these objectives); "process" (to identify the strengths and weaknesses of the chosen program during implementation); and about "product" (to find out if the objectives are met and the program effective).

Scriven (1967) separated the role of evaluation in program improvement from the role of evaluation in judging the merit and worth of a program. He coined the term "formative" for the former and "summative" for the latter. Another contribution made by Scriven which also seems to counter the emphasis on the dependent variables of the experimental approach was his conception of "goal-free" evaluation (Scriven, 1972, 1973). Here, there is no *a priori* identification of goals or outcomes; any number of consequences of a program, both intended and unintended, can be identified.

Implicit in the alternatives to non-experimental approaches to evaluation were implications for data collection. With the focus on process as well as outcomes, evaluators were required to obtain the descriptive information they needed from observation, documents or interviews, i.e., there was a qualitative rather than quantitative emphasis. This was a significant break from adherence to existing scientific traditions. Other writers continued this trend by contributing ideas for evaluation which depended on qualitative data collection. In their evaluation text *Beyond the Numbers Game*, Hamilton *et al* (1977) comment on this trend which they call "a paradigm shift from an evaluation methodology valuing numeracy to one valuing literacy," (p.4). Their discussion of illuminative evaluation is illustrative of this shift.

Illuminative evaluation does not judge, nor does it measure or predict; it is concerned instead with description and interpretation. In the same vein, Eisner (1976) voiced the opinion that looking for laws to explain or control behaviour was inappropriate in education. He suggested that, instead, educational evaluation should take the form of "connoisseurship" and "criticism." Eisner expressed the belief that the evaluator as connoisseur can heighten awareness of the quality of life and discriminate among the subtle elements of the educational situation, while the evaluator as critic can illuminate the situation so that others can understand the implications of what they observe. There are numerous references to the value of qualitative methods in the literature of the late 1970s and early 1980s (Willis, 1978; Patton, 1980, Guba and Lincoln, 1981; Hatch, 1983).² Although there is evidence that the qualitative/quantitative debate has yet to be resolved (Shapiro, 1986), it is becoming increasingly evident that resolution of this debate is less important than it was because evaluators are acknowledging now, more than ever before, that the conduct of evaluations depends on the kinds of evaluation questions asked and the particular context within which the evaluation is to take place.³

² At this time also there was considerable discussion about the utilization of evaluation results stemming from concern that these results were not being used (Patton, 1978; Alkin, Daillak and White, 1979; Braskamp and Brown, 1980; Ciarlo, 1981; Leviton and Hughes, 1981; Weiss, 1982; Leviton and Boruch, 1983; Conner, Altman and Jackson, 1984). However, as it is the conduct of evaluations rather than use of their results that forms the focus of the present study, utilization is not discussed further in this review.

³ The qualitative/quantitative debate and the transition from focussing on evaluation approach to focussing on evaluation context has been recorded in the numerous journals that began to appear in the mid 1970s. These include: *Studies in Educational Evaluation* (first published in 1975); *Evaluation in Education: An International Review Series* (first published in 1977 and now the *International Journal of Educational Research*); *Evaluation Quarterly* (first published in 1978 and published as *Evaluation Review* in 1980); and *Educational Evaluation and Policy Analysis* (first published in 1979). The series *New Directions for Program Evaluation* began in 1978; and in Canada, the *Canadian Journal of Program Evaluation* was first published in 1986.

In 1981, one of the most significant documents published in the field was released. This document, *The Standards for Evaluation of Educational Programs, Projects and Materials* (Joint Committee, 1981) was developed by a heterogeneous committee consisting of members of twelve organizations which represented both evaluator and client groups in the United States. The standards, produced over a four year period, were intended to guide and improve the practice of evaluations, as well as to provide a means of judging their success. The thirty standards were grouped in four classes: utility, feasibility, propriety and accuracy. Utility standards reflect the consensus that emerged in the late 1960s and early 1970s that evaluations should become more responsive to the needs of clients and, as a result, should become more useful to them. Feasibility standards recognize that evaluations must be workable in everyday settings and that they must be cost-effective. Propriety standards ensure that the rights of those involved in, or affected by, the evaluations are protected; and accuracy standards emphasize the technical adequacy of the evaluations. At about the same time as the development of these standards evidenced the maturing of the field and progress towards the professionalization of evaluators, another gradual change was occurring. This time it was in recognizing the complexity of the settings in which evaluations take place.

Appreciation of how these settings and the contextual variables which represent them affect the design and conduct of evaluations became more noticeable in the literature of the late 1970s and early 1980s. Writers such as Smith (1979) and Cronbach (1982) pointed out that evaluation is a service-oriented, client focussed endeavour which usually takes place in particular institutional and political contexts. Occurring in such contexts, evaluations usually require a special kind of consideration, one that does not

come easily from the use of traditional scientific methods. Cronbach suggested that the evaluator's aim is not to diminish or control for the effects of these contexts, but to use the in-context information to illuminate the evaluation situation so that ultimately, the program can be improved.

Smith argued that given the reality of institutional and political contexts, evaluators have to be able to function within them. He made the point that evaluators have to deal with the complexity of management, policy, value and economic questions (1981b, 1982). Kosecoff and Fink maintained that knowledge of evaluation theory, research design, statistics and psychometrics was no longer sufficient, since evaluators:

must negotiate with the scores of people in public and private agencies who finance evaluation studies, put the programs together, and participate in them. [They] must also organize and administer projects, speak to groups of people, and write proposals, reports and budgets. Finally [they] must frequently play the role of politician and philosopher (1982:15).

Bryk and Light suggested that evaluation designs should "mesh with the program environment and the social and political context in which they are immersed" (1981:30), and Borich viewed the approach chosen for any evaluation as essentially:

a function of the problem, the client for whom the evaluation is being conducted, the values inherent in the program context, and the amount of time and money that can be devoted to it (1983:61).

Appreciation of the importance of contextual variables led to the conclusion that there was no one best way to do evaluations (Patton, 1980; Joint Committee, 1981), not even "for an inquiry into a particular program, at a particular time, with a particular budget" (Cronbach, 1982:321).

In *Toward Reform of Program Evaluation* (1980), Cronbach *et al* argued that because the evaluation field is beset with problems it should be reconceptualized. Their arguments and recommendations for this transformation are given in the form of ninety-five theses which cover issues such as qualitative/quantitative data gathering, flexibility of evaluation design, communication of evaluation results and the characteristics of evaluators. The authors view evaluation as a political process and suggest, therefore, that program improvement can be facilitated by enlightening those involved. Cronbach *et al* advocated flexibility in evaluation design and suggested that evaluators choose whatever approaches seem appropriate given both practical and political considerations.⁴

Such advocated flexibility, such attention to the practicalities and politics of context, was not evident in the earlier literature of the field whose evolution has been described in the preceding pages. The brevity of the description has allowed the main lines of the evolution to become apparent, but has perhaps obscured an important element - that of the complexity which has accompanied growth. The world of evaluation has been described by Patton as one "filled to overflowing with uncertainty, ambiguities, competing perspectives, and conflicting roles" (1982:32). The confusion that results from such a world can make life difficult for the evaluator and the student of evaluation alike. Making sense of that confusion has been a task undertaken by a number of writers and thus, their work needs to be seen as an important aspect of the literature of the field.

⁴ A fuller discussion of these considerations, not considered appropriate for the present chapter, is contained in Cronbach (1982).

B. THE ORDERING OF THE EVALUATION LITERATURE

There have been many attempts at ordering or making sense of the literature on evaluation, so many in fact that it is useful for some order to be imposed on those attempts in turn. The following pages deal first with authors who have discussed the field in terms of underlying conceptions and, second, with those who have proposed detailed classifications of the literature.

1. Conceptions of Evaluation

The question "What is Evaluation?" is a difficult one in that it has numerous answers depending on the particular point of view adopted by the respondent. Rather than clarifying the definitional issue, it can lead to more questions about what evaluation is and how it should proceed. In general parlance, evaluation and such terms as assessment, measurement, testing or research⁵, are often used interchangeably. There are various ways of answering the question, one of which is to adopt a definition according to the approach which is emphasized. For example, the definition underlying evaluations which use experimental or quasi-experimental designs is that evaluation is the application of social science methods to determine relationships which can be described and analyzed quantitatively. The definition underlying evaluations which are designed to measure the extent to which *a priori* goals and objectives are met is that evaluation is the determination of the attainment of goals and objectives. Thus, definitions depend on particular conceptions of evaluation.

⁵ The confusion over the distinction between research and evaluation is compounded by a substantial portion of the literature being devoted to "evaluation research," (Suchman, 1967; Weiss, 1972; Hoole, 1978; Conner, 1981; Rutman, 1984; Rossi and Freeman, 1985). Indeed, Rossi and Freeman observe that the terms evaluation and evaluation research may be used as synonyms.

There is evidence in the literature that evaluation can be conceived in a number of ways; for the purpose of this discussion, evaluation is considered as a science, an art, or as naturalistic inquiry. As there are assumptions underlying these approaches which are not always made explicit, this section also includes a brief description of the "utilitarian" and "intuitionist/pluralist" assumptions identified by House (1978).

Rossi and Freeman provide a discussion of the differences between the "scientific posture" and, in their terms, a "pragmatic" one (1985:34). They draw attention to the differences between these postures by contrasting the position taken by Cook and Campbell (1979) with that of Cronbach (1982). The former position is based on the belief that decisions about programs should be based on a continuing process of testing ways by which programs may be improved. Thus, an experimental approach is advocated. In order to remain within the scientific tradition the evaluator/researcher must ensure that the requirements of science are met, i.e., that the evaluation meets the research standards of other scientific researchers. The latter position, the pragmatic one, is in contrast, based on the belief that evaluation is an art. The major proponent of the conception of evaluation as an art is Eisner (1979a). For him, evaluation involves both connoisseurship and criticism. Connoisseurship is "the art of appreciating what is educationally significant" (1979a:ix); while criticism is "the art of disclosing the qualities of events or objects that connoisseurship perceives" (1979a:197). Characteristics of the artistic approach to inquiry include its use of metaphor and expressive forms of communication; belief in the persuasiveness of personal vision and personal interpretation; and its emphasis on the creation of images that have use and meaning for others (Eisner, 1981, 1985).

According to Rossi and Freeman the artistic view (the pragmatic posture) differs from the scientific view (posture) in that rather than meeting a set of research standards, the evaluator's prime intent is to provide "maximally useful" (Cronbach, 1982) information to those involved, while working within the constraints of a particular evaluation context. This is not to imply that experiments are never appropriate; it suggests instead that those evaluations which do not meet the accepted standards of scientific research are still valuable if they provide useful information to clients.⁶

Rossi and Freeman hold a middle view between those evaluators who strive to maintain scientific standards and those who are prepared to relinquish strict allegiance to these standards in order to provide maximally useful information. Rossi and Freeman acknowledge that there are aspects of both art and science in evaluation research and that evaluation can meet scientific standards at the same time as providing clients with useful information:

In a sense we are like religious reformists who are deeply respectful of the orthodox roots of their enterprise but recognize the realities of the real world in which they are operating (1985:37).

In the final analysis, evaluation for Rossi and Freeman is a social science activity which involves the systematic application of social science research procedures. Use of these procedures results in the collection of reliable and valid evidence which is used for the purpose of judging and improving the program.

Although Rossi and Freeman refer to the "scientific" and "pragmatic" postures in their

⁶ Boruch, McSweeney and Soderstrom (1978) have compiled a list of evaluations which have taken the form of experiments and which, as well as meeting scientific standards, have provided useful information.

discussion of the field of evaluation, they make only a passing reference to a third, the "naturalistic" posture, proposed by Guba and Lincoln in their 1981 publication *Effective Evaluation*. The view of evaluation held by Guba and Lincoln is not shared by Cook and Campbell, Cronbach, or Rossi and Freeman.

Guba and Lincoln (1981) illustrate the differences between scientific inquiry and naturalistic inquiry by showing how each deals with reality, the relationship between inquirer and subject, and the nature of truth. Scientific inquiry, they suggest, is based on belief in the existence of one reality which can be fragmented into discrete parts and which can be discovered by the examination of a limited number of variables at any one time. The inquirers (or researchers) remain independent and objective. They do not affect the phenomenon under study, nor does the phenomenon affect them in any way. In order to arrive at generalizations, similarities among phenomena must be identified.

Naturalistic inquiry, on the other hand, is based on belief in the existence of multiple realities which are divergent and which interrelate to form a pattern of "truth." As interaction between inquirers and the phenomenon under study cannot be avoided, inquirers cannot remain independent. As each situation is different from every other (given the complex nature of the multiple realities and the relationship between the inquirer and the phenomenon), differences, rather than similarities, provide a focus. Thus, generalization (which results from the identification of similarities) cannot be the aim of naturalistic inquiry which aims, instead, at an understanding of the unique nature of phenomena.

Thus, Guba and Lincoln suggest that effective evaluation can result only from an appreciation of the multiple realities of stakeholders, realities which are a result of their individual experiences and which depend on their own particular value perspectives. If as Guba and Lincoln suggest, evaluations are value-oriented, then, ultimately, a program can be evaluated in terms of "the values, needs and concerns of those from whom the impetus for the program originated" (Borich, 1983:62). Criteria for evaluation, therefore, tend by this view to be context specific and based on a plurality of values.

Utilitarian assumptions (House, 1978), which underlie the scientific conception of evaluation, place emphasis on the development of explicit criteria and specific procedures in order to maintain objectivity. If criteria and procedures are made explicit, then others using them are likely to make similar observations and arrive at similar conclusions. Utilitarian approaches assume consensus on what is considered to be of value, e.g., the significance of results in experimental studies, the attainment of objectives in studies designed to determine if those objectives have been attained; the choice of one alternative from a range of alternatives in studies centred on key management decisions; and the choice of a superior package of materials in those studies which centre on the needs and wants of consumers. In these evaluations, findings are combined in an attempt to arrive at a single judgement of value.

In contrast, the intuitionist/pluralist assumptions which underlie the artistic and naturalistic conceptions of evaluation emphasize divergence and subjectivity. Multiple viewpoints may be included and criteria for determining the value of a program are derived from an individual's own experience, knowledge, perceptions and opinions.

Intuitionist/pluralist approaches rely on the evaluators' knowledge and experience rather than on the specification of criteria and procedures. For example, studies which use the expertise of a particular evaluator depend on that person's knowledge and experience for judgement rather than on the specification of criteria and procedures. Similarly, those studies that centre on the participants make the assumption that the participants' knowledge of the program in context is paramount, and as such the perceptions and interpretations of those involved in the program are of most value.

This discussion of competing conceptions was intended to illustrate the point that the evaluator has a number of alternative basic conceptions available. Although basic conceptions inevitably underlie more detailed classification schemes, there are enough of the latter to warrant separate discussion. They do, moreover, amplify the picture of evaluation to an extent not completely achieved by the discussion of basic conceptions.

2. Classification Schemes

The classification schemes to be discussed in this section were put forward by Worthen and Sanders (1973), Popham (1975), Stake (1976b), Gardner (1977), Glass and Ellett (1980), House (1980), Stufflebeam and Webster (1980), Talmage (1982), and Pietro (1983). The presentation of these schemes illustrates how people writing in the field think about evaluation in different ways. For ease of understanding and classification, the authors have reduced evaluation approaches to "ideal types" (House, 1980). Each "ideal type" represents the basic elements of an approach, i.e., its proponent's conception of evaluation and the technology suggested by that proponent as the means of operationalizing that concept in practice. The authors of each scheme treat each approach as a whole, as a single entity in its entirety. As such each is grouped with

and distinguished from others on the basis of observed similarities and differences. The authors of the classification schemes have identified and labelled these approaches in different ways and have compared them on different dimensions. The remainder of this section contains a discussion of the classification schemes. The approaches classified are listed in Table 2.1, while the dimensions used for classifying them are listed in Table 2.2.

The major bold-faced divisions of Table 2.1 are the names of the authors of the classification schemes. Below each author are listed the names of the approaches to evaluation identified by those authors in their particular classification schemes. In accordance with the divisions made by the authors, both major and minor headings are used in some cases. The brief description which follows is intended to draw attention to the different ways that the authors have identified and grouped the approaches.

Worthen and Sanders (1973) suggest that their eight approaches can be classified in three major groups according to whether they emphasize the critical nature of judgement (judgemental approaches); the importance of data-collection and storage for the use of decision-makers (decision-management approaches); or the attainment of behavioural objectives (decision-objective approaches). Popham (1975) goes further by suggesting that there is a distinction between those judgemental approaches which emphasize intrinsic criteria and those which emphasize extrinsic criteria. He describes four general classes of approach: goal-attainment (similar to the decision-objective approaches of Worthen and Sanders); judgemental, with emphasis on intrinsic criteria, i.e., professional judgement; judgemental, with emphasis on extrinsic criteria; and decision-facilitation (similar to the decision-management approaches of Worthen and

TABLE 2.1

Authors of Classification Schemes and Evaluation Approaches Identified

Worthen and Sanders (1973)*Judgemental*

Scriven

Stake

personal judgement

Decision-Management

Stufflebeam

Alkin

Decision-objectives

Provus

Hammond

Tyler

Popham (1975)

judgement-extrinsic criteria

judgement-intrinsic criteria

decision facilitation

goal attainment

Stake (1976b)

student gain by testing

institutional self-study

blue-ribbon panel

transaction-observation

management analysis

social policy analysis

instructional research

adversary

goal-free

Gardner (1977)

goal-free/responsive

professional judgement

decision-oriented

measurement

assessed congruence between

performance and objectives

Glass and Ellett (1980)

description or portrayal

jurisprudence

rational empiricism

assessed progress towards goals

decision theory

systems management

applied science

Stufflebeam and Webster (1980)*Politically Oriented*

politically controlled

public relations inspired

Questions Oriented

objectives-based

accountability

experimental research

testing programs

management information

systems

Values Oriented

client-centered

connoisseur-based

consumer-oriented

decision-oriented

policy studies

accreditation/certification

House (1980)

systems analysis

behavioural objectives

decision-making

goal-free

art criticism

professional review

quasi-legal

case study/transaction

Talmage (1982)

experimentation

eclecticism

description

benefit/cost analysis

Pietro (1983)

goal-based

decision-making

goal-free

expert judgement

naturalistic

TABLE 2.2

Authors of Classification Schemes and their Bases for Comparison among Approaches

Worthen and Sanders (1973)

definition
 purpose
 key emphasis
 role of evaluator
 relationship to objectives
 relationship to decision-making
 types of evaluation
 constructs proposed
 criteria for judging evaluation
 implications for design
 contributions
 limitations

Popham (1975)

main emphases

Stake (1976b)

purpose
 key elements
 purview emphasized
 protagonists
 cases, examples
 risks
 payoffs

Gardner (1977)

principal focus
 examples
 basic assumptions
 advance organizers
 nature of expected outcomes and
 mode of interpretation

Glass and Ellett (1980)

alternate conceptions

Stufflebeam and Webster (1980)

advance organizers
 purpose
 source of questions
 main questions
 typical methods
 pioneers
 developers

House (1980)

major audiences
 focus of assumed consensus
 methodology
 outcome
 typical questions

Talmage (1982)

philosophical base
 disciplinary base
 focus of methodology
 methodology
 variables
 control or comparison group
 participants' role in evaluation
 evaluator's role
 political pressures
 focus of evaluation report

Pietro (1983)

major purpose
 typical focus questions
 methodology

Sanders). Stake (1976b) lists nine approaches: test-based student assessment, institutional self-study, blue-ribbon panels, transactional/observational approaches, approaches based on management analysis, on social policy analysis, and on instructional research and those identified as adversary and goal-free approaches. Gardner (1977) identifies five approaches which he bases on definitions of evaluation, while Glass and Ellett (1980) identify seven approaches, also derived from alternate definitions of evaluation. Stufflebeam and Webster (1980) identify thirteen approaches and suggest that some of these are not *bona fide*. Of the three general categories of evaluation studies that they put forward, they suggest that the first two, containing politically-oriented and questions-oriented studies, are not "true" evaluation approaches in the sense that they are not intended to assess or improve the worth of the object being evaluated. The third category, containing approaches that are values-oriented, is their category of "true" evaluation approaches. House (1980) provides a taxonomy of eight major approaches, while Talmage (1982) identifies four approaches, and Pietro (1983) identifies five.

In the same way that Table 2.1 indicated the number and diversity of evaluation approaches, Table 2.2 illustrates the number and diversity of ways by which evaluation approaches may be compared. Worthen and Sanders (1973) identify twelve bases for comparison, and Talmage (1982) identifies ten. A more general view is provided by Popham (1975) and Glass and Ellett (1980) who distinguish among approaches on the basis of main emphases and alternate conceptions respectively. Pietro (1983) identifies three bases for comparison while Gardner (1977) compares the approaches he identifies on five bases: principal focus; examples of the approach in practice; basic assumptions; advance organizers such as research design, evaluator role, method, and type of

communication and feedback; and the nature of expected outcomes and the way that the data are interpreted. Stufflebeam and Webster (1980) also use advance organizers, but in addition, they distinguish among the approaches on the basis of such variables as their developers and the source and type of evaluation questions asked. House (1980) also focusses on the questions asked, on methods, and, like Gardner, on outcomes. The remaining bases for comparison are listed in Table 2.2. In the same way that there are commonalities and differences among the approaches identified, there are also similarities and differences among the bases for comparison chosen by the authors.

A brief discussion will illustrate the potential for confusion when authors of classification schemes label approaches in different ways and compare them on various dimensions. There are two particular problems which face the prospective evaluator. The first is that many of the approaches are closely allied to particular methods. This may suggest to the unwary that one approach rather than another would be appropriate. Illustrated in Table 2.1, for example, are approaches with labels indicative of particular methods such as blue-ribbon panel, student gain by testing, measurement, professional judgement, applied science, jurisprudence, art criticism and benefit/cost analysis. The second potential problem is that although they are classified as discrete entities, these approaches are not mutually exclusive. Authors have classified the approaches on the basis of different, implicit criteria with the result that some which are labelled in the same way may not include the same identifying characteristics, and some which have different labels may contain similar characteristics. Thus, similarly labelled groupings, based as they are on different criteria, may include or exclude approaches which have been subsumed by other classification schemes. Indeed, some

approaches may not appear at all.

For example, both House (1980) and Stake (1976b) use the descriptor "transaction" in reference to broadly conceived approaches. House uses the descriptor "transaction" and "case study" interchangeably while Stake identifies a "transaction-observation" approach that includes the particular transactional approach suggested by Rippey (1973),⁷ which, in turn, can be easily distinguished from responsive, case study or illuminative evaluation. Responsive, illuminative and the transactional approach of Rippey are subsumed within the naturalistic approach of Pietro (1983), which cannot be the same as the naturalistic approach of Guba and Lincoln (1981) as Pietro identifies five major approaches and Guba and Lincoln identify two. Responsive, transactional and illuminative evaluation are found within the client-centered studies of Stufflebeam and Webster (1980), while although Glass and Ellett (1980) include Stake's work in the approach labelled "description or portrayal," they make no mention of transactional or illuminative evaluation. Stake, House and Pietro use the goal-free label to describe the goal-free approach (as first described by Scriven, 1973), while Gardner (1977) writes that goal-free and responsive evaluation are similar enough to be considered as a single approach. Popham (1975) includes goal-free evaluation in his general class of judgemental approaches emphasizing extrinsic criteria, while Glass and Ellett include it in the approach they call "rational empiricism." Adversary evaluation (Stake), jurisprudence (Glass and Ellett) and quasi-legal evaluation (House) are three labels for an approach, which, in the eyes of these writers is exclusive and is therefore not subsumed by any other categories. None of the remaining classification schemes, however, has a similar category, although "judicial proceedings" and "adversary

⁷ "Transaction," according to Rippey (1973) has a very specific focus, i.e., the effects of change on the role incumbents of a system.

reports" are included by Stufflebeam and Webster (1980) as typical methods, the former under their approach labelled "policy studies," and the latter under their "client-centered" approach. Some authors have generated approaches which are not emphasized in other schemes. For example, Talmage (1982) is the only writer to suggest benefit/cost analysis as a discrete approach⁸ and Stufflebeam and Webster (1980) are the only writers to base their classification scheme on the notion that some approaches are not "true" evaluation approaches.

In some cases, however, there does appear to be some consensus among authors who have identified similar approaches and to have organised them in similar ways. For example, the reader can assume that decision-objectives (Worthen and Sanders, 1973), assessment of congruence between performance and objectives (Gardner, 1977), objectives-based (Stufflebeam and Webster, 1980), behavioural objectives (House, 1980), goal attainment (Popham, 1975), assessment of progress towards goals (Glass and Ellett, 1980), and goal-based (Pietro, 1983) are all concerned with determining the extent to which progress towards meeting specific objectives and goals has occurred.

Similarly, in terms of bases for comparison (Table 2.2), those who have produced the various classification schemes have compared the approaches on a number of different dimensions and, although there are numerous differences among these bases, there are also some commonalities. For example, "purpose" is listed in four schemes (Worthen and Sanders (1973), Stake (1976), Stufflebeam and Webster (1980) and Pietro (1983)). In terms of differences among bases for comparison, Talmage (1982) provides the only

⁸ However, some publications are devoted solely to the utility of cost/benefit analysis or cost effectiveness in evaluation (Levin, 1975; Rothenberg, 1975; Thompson, 1980; Catterall, 1985).

scheme which points to social science disciplines as a source of comparison for the approaches.

The unit of analysis which has formed the basis for discussion in this section is the "evaluation approach." Although the notion of "evaluation approach" may be useful in academic discussions of the similarities and differences between and among evaluation concepts and their associated methods, classifying approaches in the ways that have been illustrated above may not be particularly useful for the evaluator engaged in the practice of evaluation or for the evaluation researcher. The schemes are useful in that they illustrate the general similarities and differences among approaches and in that they place some order on a very confusing field. However, the apparent lack of consensus about what constitutes a particular evaluation approach and about the ways these approaches may be compared makes discussion of approaches and schemes for ordering them less useful than initially appeared to be the case. The schemes provide a way of framing the field and illustrate the kinds of alternative methods available to evaluators, but are not particularly useful to evaluators seeking appropriate designs to answer particular questions in particular contexts. In their interpretation and in their application these schemes have limited utility.

Each of the schemes imposes order on the evaluation field; collectively, however, they are likely to leave the reader sensing some confusion, probably very little different from that which motivated the authors to generate their systems of classification in the first place. This confusion may stem, not only from the amount of literature and from the existence of competing basic conceptions, but also from a lack of knowledge about exactly how evaluations are carried out in practice. A focus on the *how* of evaluation

practice may yield an alternative way of looking at the field, which, rather than placing emphasis on underlying conceptions of evaluation would focus attention on how evaluation is done. The following section explores this possibility by moving away from the conceptualization of evaluation and considers, instead, the bases from which evaluation may be examined in practice.

C. FROM APPROACH-BASED WRITING TO ISSUE-BASED INVESTIGATION

The point was made in the previous section that it may be useful to deemphasize the approach as the unit of analysis and to find another way of facilitating an understanding of evaluation in practice. A viable alternative would have to be relevant to both evaluators and their clients. Ideally, it would aid those involved in the evaluation process in carrying out evaluations in particular contexts. Asking basic questions about evaluations would seem a reasonable place to start. It is likely that if the questions were well chosen, the answers to them would facilitate the identification of a wider range of underlying issues in evaluation than simply those concerned with the methodological issues (around which so much of what has been termed approach-based writing turns).

It is almost certainly true that for the evaluator him or herself, questions of method are paramount. But to consider evaluation as practiced is to consider a wider range of participants than just the evaluator. There are those who have called for the evaluation (the clients), those whose programs are to be evaluated, and the general tax-paying public besides. The range of relevant issues then becomes greatly expanded — issues of purpose, outcome, or personnel, as well as of method. As suggested above, the identification of basic questions may well serve as a useful point of entry into this

confusing field.

Few authors in the corpus of scholarly literature on evaluation have approached the field in this way, but Nevo (1983, 1986) provides a starting point. He suggests that asking questions about evaluation enables readers, evaluators and researchers to organize and develop their own perceptions of evaluation rather than piously adopting one evaluation approach or another. He suggests that the evaluator or prospective evaluator would benefit greatly from being given the opportunity to come to terms with the key issues in the field and thus, to become sufficiently well informed to develop his or her own conception of the field and of the various ways that evaluations could be undertaken. Thus, Nevo (1983:118) provides an alternative to classification schemes by identifying ten questions that serve to draw attention to evaluation issues rather than evaluation approaches. The questions are as follows:

1. How is evaluation defined?
2. What are the functions of evaluation?
3. What are the objects of evaluation?
4. What kinds of information should be collected regarding each object?
5. What criteria should be used to judge the merit and worth of an evaluated object?
6. Who should be served by an evaluation?
7. What is the process of doing an evaluation?
8. What methods of inquiry should be used in evaluation?
9. Who should do evaluation?
10. By what standards should evaluation be judged?

Six of these questions, however, imply prescription. These are the "should" questions. Seeking answers to normative questions often presupposes that the "is" and "are" questions have been answered. However, given the dearth of empirical studies on how evaluation is actually carried out (Smith, 1980; Stufflebeam, 1981; Stufflebeam and Welch, 1986), such answers cannot be assumed. The list of questions is useful in that

it changes the emphasis from an approach-based view of evaluation to an issue-based one.

D. SUMMARY

This chapter has attempted to illustrate the complexity and diversity of the field as reflected in the voluminous evaluation literature. It outlined the development of this literature. It described different conceptions of evaluation and various classification schemes which have been used to place some order on the literature of the field. The final section suggested that one way of facilitating the empirical study of evaluation practice is to change the emphasis from approach-based to issue-based enquiry.

According to one author (Nevo, 1983), certain questions suggest themselves as appropriate for such an issue-based inquiry. By extension, the list of questions might also provide a point of departure for the development of a framework (the absence of which was noted in Chapter 1) with which to examine program evaluation in practice. The development of this framework and the design of the empirical study of evaluation practice undertaken here are discussed in the next chapter.

CHAPTER III.

THE FRAMEWORK AND RESEARCH DESIGN

The purpose of this study was to examine how program evaluation, as documented, is done in school districts in British Columbia. This involved a number of tasks:

1. designing a framework for the analysis of evaluation practices;
2. identifying a data base representative of school district evaluation practices;
3. using the framework to examine this data base;
4. displaying the results and analyzing the findings so as to provide a comprehensive account of school district evaluation processes; and
5. using this detailed empirical knowledge of evaluation practice to add to the existing evaluation knowledge base.

This chapter deals with the first two of these tasks. The development of the framework is described in the first part of this chapter. The identification of a data base is discussed in the second part which focusses on the research design of the study.

A. THE FRAMEWORK

In order for the reader to follow the description of the development of the study's framework, it is helpful to recapitulate what was said in Chapter I about the rationale for the study and the kind of data used. It will be recalled that the quest for some detailed empirical knowledge about evaluation in school districts was signalled as urgent by a number of writers, most recently Stufflebeam and Welch (1986). As was noted in Chapter I, evaluation documents are a useful source of information because they

represent the aspects of the evaluation that evaluators perceive to be important and worthy of inclusion. Thus, the framework developed was one which from the start was designed for use with written data.

The framework was developed in three stages. The first was the identification of four basic questions. The second was combining these with the already-noted questions suggested by Nevo (1983) to form a preliminary framework. The third involved modifications to the specific questions in the framework. These stages are now described in greater detail, and the complete framework of four general questions and ten specific questions is provided.

1. Identification of Basic Questions

Fundamental to the *raison d'être* of the study was the lack of any broad-based, large-scale empirical description of evaluation as practiced. Description, however, cannot begin without some starting point, some organizing question or concept. It was shown in Chapter II that little was available in the existing literature which might be used as such a starting point, except, that is, for the questions suggested by Nevo (1983). The researcher was thus faced with a situation in which (1) no framework was available for approaching such a description, and (2) she herself was a knowledgeable practitioner in the field to be described. In this situation, it seemed sensible to begin with some basic questions. The questions found to be most useful are described in the following paragraphs.

An evaluation takes place in a given context; it is directed at an object and designed to fulfil certain intents. The way a particular evaluation is defined depends on the

conception of evaluation operative at that time (this conception is usually shared by evaluator and client and can determine the desired end-product of the evaluation and the way in which the evaluation is carried out); the characteristics of the object to be evaluated; and the particular intents identified. It is the fulfillment of these intents that is the desired end product of many evaluation studies. Thus, the question "Evaluation — to what end?" can be argued to be fundamental to the enterprise.

Once the evaluation has been conceived, it is important to determine how the required information may best be collected. Methods of inquiry, data collection techniques and other activities associated with the collection and compilation of information must be determined. Thus, the question "Evaluation — by what means?" is also a basic question about any evaluation.

Evaluations are always conducted *for* some individual or group *by* some individual or group. The recipient of the evaluation may have an important bearing on the way the evaluation is conducted, what is reported, and how it is reported. The evaluator(s) may also be an important variable in the evaluation process. Thus, a third basic question to be asked is "Evaluation — for whom and by whom?"

Often evaluators or clients will suggest that steps be taken in order to modify certain aspects of the object of the evaluation. Sometimes evaluators include such recommendations in the final section of their reports and sometimes recommendations suggested by clients are attached to the document once it has been submitted by the evaluators. It is for this reason that a fourth general question, "Evaluation — with what conclusion?" seems important.

Thus, the four questions: "Evaluation — to what end?" "Evaluation — by what means?" "Evaluation — for whom and by whom?" and "Evaluation — with what conclusion?" form the basic structure of the framework for the present study. The development of the necessary detail within that basic structure was achieved by considering the so-called "issue-based" questions proposed by Nevo (1983).

2. Developing the Framework

Nevo (1983) suggested that his ten questions could provide a first framework for the collection of empirical data about program evaluation. The questions are as follows:

1. How is evaluation defined?
2. What are the functions of evaluation?
3. What are the objects of evaluation?
4. What kinds of information should be collected regarding each object?
5. What criteria should be used to judge the merit and worth of an evaluated object?
6. Who should be served by an evaluation?
7. What is the process of doing an evaluation?
8. What methods of inquiry should be used in evaluation?
9. Who should do evaluation?
10. By what standards should evaluation be judged?

For the present study, Nevo's questions were reordered in light of the four more general basic questions identified in the first section above. This reordering, which was intended to result in a list of questions more useful for the task of examining evaluation practice, is described below.

The first general question "Evaluation — to what end?" is concerned with the definition of evaluation, its intents, and the objects being evaluated. In addition, the question about evaluation standards is included because it has implications for the conceptualization of evaluation. Thus, the questions to do with definition, function, object and standards would be appropriate here. The first part of the list would

therefore read:

1. Evaluation — to what end?
 - a. How is evaluation defined?
 - b. What are the functions of evaluation?
 - c. What are the objects of evaluation?
 - d. By what standards should evaluation be judged?

The second general question “Evaluation — by what means?” is concerned with how the evaluation process is carried out. Four of Nevo’s questions are relevant here: focussing on kinds of information required, choice of criteria for judgement, process and methods of inquiry. The second part of the list would therefore read:

2. Evaluation — by what means?
 - a. What kinds of information should be collected regarding each object?
 - b. What criteria should be used to judge the merit and worth of an evaluated object?
 - c. What is the process of doing an evaluation?
 - d. What methods of inquiry should be used in evaluation?

The third general question “Evaluation — for whom and by whom?” refers to the audiences or the clients of the evaluation, and to the evaluators themselves. Two questions from Nevo’s list refer to these aspects, hence the third part of the list would read:

3. Evaluation — for whom and by whom?
 - a. Who should be served by an evaluation?
 - b. Who should do evaluation?

The fourth part of the list would include the last of the basic questions given in the section on identifying the basic questions above. The question is not in Nevo’s list but was added because many evaluations result in recommendations. Thus, the final general question and the final specific question would read as follows:

4. Evaluation — with what conclusion?:
What recommendations (if any) are made?

3. Modifying the Specific Questions

For the purpose of this study, one major problem encountered with some of the questions on Nevo's list was that they were worded in normative terms. If the questions are to be used as a framework for the collection of empirical data, then, asking "should" questions is appropriate only for ascertaining the values people hold about evaluation. In the present case, the intent was to explore the actual practice of evaluation, so non-normative questions were needed. Accordingly, the "should" form of six of Nevo's questions was recast using the simple is/are form.

In addition, not all of Nevo's questions were mutually exclusive, and some of the terms used in the questions were ambiguous. Consequently, modifications were made to a number of questions. Details of these modifications, together with the rationales for making them, are provided in the following paragraphs. The modified list of questions which forms the framework for the study and which subsequently was used to structure the initial analysis of the data, is given at the conclusion of the section.

Evaluation – to what end?

1. How is evaluation defined? (retained)

In Chapter II, attention was drawn to the number of conceptions of evaluation extant in the literature which can determine the definition of evaluation adopted by evaluator and client. The point was made that some of these conceptions can be equated with a particular methodological approach. As methods of inquiry are addressed in a later question, the issue of definition to be investigated in this study did not include such methodological definitions of evaluation. The way definition of evaluation was to be interpreted in this study, however, derived from

contrasting conceptions of evaluation evident in the literature. Inherent in the work of Stufflebeam *et al* (1971) and Stufflebeam (1975), and (with some difference in emphasis), in the work of Cronbach *et al* (1980), is a focus on the information gathering aspects of evaluation. This is different from the judgemental focus placed on the process by members of The Joint Committee on Standards for Evaluation (1981) or by Guba and Lincoln (1981). Thus, in the present study, evaluation can be defined as *judgement*, in which the evaluators were responsible for the determination of merit and worth; or as *provision of information*, in which evaluators reported findings but did not make judgements. Thus, evaluators can be judges or information brokers.

2. What are the functions of evaluation? (reframed)

Function can be interpreted in two ways. It can be interpreted in the general terms of the formative or summative nature of the evaluation process, i.e., formative evaluation is usually intended to improve a program in process while summative evaluation is intended to result in a judgement about the success of a program. Function can also be interpreted in terms of the *purposes* of the study, i.e., espoused intents such as to ensure accountability, to inform decision-making, or to improve the program. Although functions (formative and summative) and purposes (decision-making or improvement) are related, they are different enough to warrant separate treatment in this study. It is for this reason that a question about the intents of an evaluation (i.e., function and purpose), rather than its function only is deemed appropriate for inclusion here. The question was amended to read "What are the intents of the evaluation?"

3. Why was the evaluation undertaken? (added)

In one sense, the previous question "What are the intents of evaluation?" posed a

“why” question in terms of the question “Evaluation — to what end?” There is, however, another sense in which the “why” question can be asked, namely, “Evaluation — from what circumstances?” The question “Why was the evaluation undertaken?” was therefore added in order to ensure that both senses of “why” were included. The usefulness of its inclusion was confirmed by the preliminary examination of the data, at which time it was noted that mention was often made of the circumstances surrounding the initiation of each evaluation.

4. What are the objects of evaluation? (reworded)

Educational evaluators usually focus on the evaluation of personnel, materials, projects or programs. Clear identification of what is to be evaluated is crucial to keeping the evaluation focussed. In order to remove the possibility that the “object” of an evaluation might be interpreted erroneously as the function or purpose of the evaluation, this question was reworded to ensure that the “object” of an evaluation refers only to that which is evaluated. The question was changed to read “What is the object evaluated?”

5. By what standards is evaluation judged? (excluded)

Evaluation reports are produced by evaluators to inform their clients. Standards for the judgement of the quality of an evaluation often depend on the standards of these particular client groups. As these standards were not evident from the data used in this study, this question was excluded.

Evaluation — by what means?

6. What kinds of information are collected regarding each object? (retained)

The amount and type of information collected is determined primarily from a consideration of the aspects of the object which are to be evaluated. Information

can be collected on a wide variety of topics ranging from, for example, the stated goals of a program or its implementation, to its outcomes and information about stakeholder perceptions of its value.

7. What criteria are used to judge the merit and worth of an evaluated object? (retained)

Choosing criteria for judgement is one of the most crucial of evaluative tasks. Those evaluators who function as information brokers rather than as judges are able to avoid such choices by leaving responsibility for judgement to their clients, or by working with them in order to determine appropriate criteria in any given evaluation situation. Taking responsibility for judgement, however, means taking responsibility for the interpretation of the collected information in light of certain criteria. The identification of these criteria is often difficult in that evaluators and clients have their own implicit ideas of what should be looked at in order to determine what is of value and what is not. Moreover, they do not always make their choices explicit.

8. What is the process of doing an evaluation? (excluded)

The evaluation process may be considered in three parts. The first, the conceptual, usually involves some degree of communication between evaluator and client about how the evaluation is to be conceived (or defined); about the nature of the object to be evaluated; and about the purposes to be fulfilled. The second, the technical, involves how the information is to be gathered; and the third, the interpretive, involves the interpretation of the gathered information, the assigning of value and, where appropriate, recommendations for further action.

For the first of these three parts, the conceptual part, information is provided by answers to the first four questions on this modified list (How is evaluation

defined? What are the intents of evaluation? Why was the evaluation undertaken? and What are the objects evaluated?). Answers to two other questions (What kinds of information are collected regarding each object? and What methods of inquiry are used in evaluation?), provide information on the second, technical, part of the evaluation process; while an answer to What criteria are used to judge the merit and worth of an evaluated object? and to the final specific question, What recommendations (if any) are made? provide information on the third and final interpretive part of the evaluation process. Thus, there was no need to retain a separate question about process.

9. What methods of inquiry are used in evaluation? (retained)

Choice of methods of inquiry occurs during the second, technical part of the evaluation process. In Chapter II it was noted that evaluation methods or models were often linked with particular conceptions of evaluation. In general terms, methods of inquiry refer to the approaches chosen by the evaluators. They may also correspond to a particular approach or model suggested by a writer in the evaluation field. In specific terms, the technical activities of data collection and analysis are subsumed within this question.

Evaluation – for whom and by whom?

10. Who is served by an evaluation? (reframed)

If an evaluation is intended to inform a particular client group (e.g., school trustees) for a specific purpose (e.g., to determine whether a program should continue to be funded), then determining who is served by the evaluation does not pose a problem. If, however, those served by an evaluation include all those involved in or affected by the program or the evaluation of that program (the

stakeholders⁹), or if the intended audiences are members of the policy-shaping community, then determining who is served by the evaluation may be difficult. Different evaluation reports may be submitted to different audiences. It is not always possible to identify various audiences from evaluation documents nor is it always possible to discern whether more than one evaluation report was produced to inform more than one audience. In order to make clear that those served by the evaluation are those to whom the evaluation reports are submitted the question was changed to read "To whom is the report submitted?"

11. Who does evaluation? (reframed)

The point that responsibility for evaluation (in terms of the assignment of value) may be given to the evaluator or to the client has already been made. It is for this reason that the question was reframed in order to ascertain who the designated individuals are, i.e., who are listed as evaluators or who are the authors of the evaluation report. The question was changed to read "Who are the designated evaluators?"

Evaluation -- with what conclusion?

The single question pertinent here was added to Nevo's list in the previous section. It reads, "What recommendations (if any) are made?"

In sum, Nevo's questions were recast in non-normative form. Three of them were

⁹ Stakeholders are those individuals and group members who are affected in some way by the program to be evaluated, its evaluation, or the outcomes of that evaluation. In this way, each individual has a "stake" in the evaluation. Not all groups have equivalent stakes, however, as the extent of each stake varies according to the nature of individual interests and formal or informal position within the school district. The term "stakeholder" may be attributed to the U.S. National Institute of Education (NIE) in that the stakeholder approach to evaluation was developed there in the late 1970s.

reframed, two were deleted, and one was reworded in order to clarify its meaning. Two questions were added to Nevo's list. As this framework was intended for the examination of the way specific evaluations are carried out in practice, one further minor modification was necessary. Written in the way they are above, the questions have general applicability. For the purpose of this study, however, it was necessary that they be applicable to specific examples of evaluation studies. The use of the definite article in questions #2, #5, #6 and #7, and the change from the plural to the singular verb and noun form in #4 achieved this end.

The four basic questions within which the ten specific questions were clustered, formed the initial framework for this study. Its purpose was to serve as a guide for approaching the examination of evaluation practice. The framework is given overleaf.

A Framework for the Analysis of Evaluations

Evaluation – to what end?

1. How is evaluation defined?
2. What are the intents of the evaluation?
3. Why was the evaluation undertaken?
4. What is the object evaluated?

Evaluation – by what means?

5. What kinds of information are collected regarding the object?
6. What criteria are used to judge the merit and worth of the object?
7. What methods of inquiry are used in the evaluation?

Evaluation – for whom and by whom?

8. To whom is the evaluation report submitted?
9. Who are the designated evaluators?

Evaluation – with what conclusion?

10. What recommendations (if any) are made?

B. RESEARCH DESIGN

The second task listed at the beginning of the chapter was the identification of a data base which would represent school district evaluation in order that the exploration of how program evaluation is done in practice could proceed. This data base was defined as the formal written evaluation reports produced in school districts. The rationale for using reports as data was given in Chapter I. If these reports could be obtained from school districts then it would be possible to provide a description of how program evaluation is reported and, by inference, how it is carried out in practice. Contained within this part of the chapter is a description of the steps taken to obtain recent evaluation reports. The description also indicates the number of school districts represented and the total number of written reports included in the data base.

The research design for the study is described in the following pages under the headings: The Method of Content Analysis, Data Collection and Content Classification. The final section sets forth the delimitations of the study.

1. The Method of Content Analysis

This section provides a general overview of content analysis and addresses the reliability and validity of the use of this method.

a. Content Analysis

Content analysis is a method used by researchers to analyze a wide range of communications. These communications can take a variety of forms, not the least of which is the written word. Holsti describes content analysis as a "multipurpose research method developed specifically for investigating any problem in which the

content of communication serves as the basis of inference" (1969:2). The method is often used as a way of producing descriptive information, cross-validating research findings or testing hypotheses (Borg and Gall, 1983). It is the first of these that is pertinent in pursuit of the second part of the purpose of this study, to describe evaluation practices in British Columbia school districts. The development of a framework (if one is not readily available), is a vital stage in the implementation of the method of content analysis. It is the identification of coding categories for reporting content which facilitates the analysis of the content of the evaluation documents.

Content analysis provides a means by which a body of written material can be conceptualized, described and analyzed. Questions are asked of the text in order to facilitate both the identification of categories and the process of "unitizing," i.e., identifying the boundaries of the units recorded within that category (Holsti, 1969). Content analysis draws much of its strength from allowing a preliminary identification of categories based on one set of questions to suggest other useful categories. Researchers using content analysis as a research method would be ill-advised to specify in precise terms every possible advance organizer, because premature specificity could result in overlooking important information. As a result of this process of identifying units of content and categories pertinent to each of the guiding questions, detail about each category of information can be explored. Thus, in content analysis, there should be a balance between the identification of questions and concepts which serve to focus the research, and the identification of categories and specific units of content as they emerge from the examination of the data.

The identification of categories and of their component content units allows for the

results to be quantifiable. The quantifiable nature of the results is an aspect of content analysis which is emphasized by those who have written about the process. For example:

Content analysis is a way of asking a fixed set of questions unfalteringly of all of a predetermined body of writings, in such a way as to produce countable results (Carney, 1972:6).

In general, content analysis applies empirical and statistical methods to textual material. Content analysis particularly consists of a division of the text into units of meaning and a quantification of these units according to certain rules (Lindkvist, 1981:34).

Thus, once the content has been categorized, it can be subjected to statistical analysis which provides such descriptors as absolute and relative frequency counts, and, according to the purpose of the research, can also indicate relationships among content variables by means of techniques such as cross-tabulation.

Not only is content analysis a method which facilitates the critical examination of documents by enabling researchers to identify a whole range of attributes which might otherwise go unnoticed, but also it is appropriate for use when the researcher is unable to, or chooses not to collect data from primary sources such as interviews or participant observation. As the researcher is not present in the particular situation reported in the document and as variables in this situation are not controlled in any way, the use of content analysis does not have any effect on the research situation or on the respondents within that situation.

b. Reliability and Validity

Important in the consideration of any research method are its implications for reliability and validity. If valid inferences are to be drawn from the documents used in a content analysis of written materials, then the categories and units of content identified must be reliable, i.e., there must be consistency in that different people with an understanding of the content (program evaluation) and of the common context (school districts) would use the coding categories in a similar manner. In terms of validity, the units of content and categories must represent what it is intended that they represent. If there is semantic ambiguity in understanding what the authors of the reports mean, or if the units of content and categories are not clearly defined, then it is likely that reliability and validity will be reduced.

In his discussion of validity, Weber (1985) points out that validity in content analysis commonly refers to the correspondence between categories (and their units of content) and the concepts they represent, i.e., the researcher defines a concept and then the category that represents it. This is face validity, a form of validity which is perceived as being weak. This weakness is intensified in those situations where face validity is the only form of validity, i.e., no external validity is established. Establishing external validity requires making a comparison with an external criterion. For example, a study has construct validity when its results correlate with other measures of the same construct. It has predictive validity when its results and inferences based on these results can be generalized to other contexts. Content analysis, however, rarely has construct or predictive validity.

The problem of reliability and validity in content analysis has been addressed (Auman,

1987; Babbie, 1986). Auman notes that Babbie distinguishes between validity and specificity rather than between validity and reliability. The concept of "specificity" as an alternative to reliability is a useful one. Content that is specific, clear and unmistakable (manifest content) can be coded with ease and its coding is reliable. Content that is not clearly specified and has to be inferred from the text (latent content) may be crucial in identifying meaning of the communication and thus is pertinent in the consideration of validity. Babbie suggests that communications can be coded for their manifest and their latent content. Identification of manifest content may raise reliability but, if it is the only content used, may lower the validity of the findings. The identification of latent content may raise validity but have lower claims to reliability. Thus, in order to increase the likelihood that the results of a study will be both reliable and valid, it is necessary to code both manifest and latent content. To this end, the coding developed in this study and described in the next chapter makes provision for coding content which is both manifest (stated) and latent (implied).

In his discussion of reliability, Weber (1985) refers to the three types of reliability identified by Krippendorff (1980). They are: accuracy, stability and reproducibility. In order to be accurate, coding of content must correspond to a fixed standard. In order to be stable, coding of the same content done by the same researcher must be consistent over time. In order to be reproducible, coding of the same content done by different researchers must be consistent, i.e., there must be inter-coder reliability.

In this study, these forms of reliability apply. Accuracy was achieved by developing decision rules for coding. Details of these decision rules are contained in Chapter IV. To ensure stability, the researcher did two things: first, after an interval of four

months, she recoded a random sample of reports for both manifest and latent content; second, she examined the differences in percentage frequencies of the occurrence of units of content in two groups of reports: those submitted by two school districts (n=34 reports) and the remaining districts (n=51 reports). This was intended to show if the presence of comparatively large numbers of reports submitted by two school districts affected the characteristics of the data base. Neither analysis revealed an appreciable difference between groups and it was therefore concluded that coding was reliable with respect to stability. The reproducibility of the content was ensured by the use of two volunteer coders. They were given instruction in category definition and were asked to code a random sample of twelve reports. Initial intercoder agreement was 92% for the categories of definition, and function; 75% for the category of intents and 85% for the categorization of recommendations. Following this initial coding, discussion about disagreements was undertaken in order to arrive at a final agreed decision. The final agreements confirmed the researcher's own coding and made clear that full reproducibility is dependent not only on a knowledge of the rules, but also on an intensive familiarity with the nature of the documents and which parts of them may provide evidence for the assigning of content to particular categories.

It is possible, however, that questions of reliability and validity are less appropriate to the kind of content analysis described here, than they would be in other research contexts (Miles and Huberman, 1984; Hutchinson, Hopkins and Howard, 1988). Although there are accepted ways of ensuring that appropriate standards of qualitative analysis are maintained, such as the audit trail proposed by Guba and Lincoln (1981), Miles and Huberman (1984) point out that researchers who interpret qualitative data do not have a generally accepted arsenal of clearly defined methods such as that

available to researchers who interpret data amenable to statistical analysis. These authors suggest that it is necessary to make explicit the processes that are used in qualitative data analysis. They identify three interactive components of qualitative data analysis: data reduction, data display and conclusion-drawing and verification, and they offer suggestions as to how researchers can ensure that these activities are carried out effectively. Thus, good data analysis can occur, not only by using psychometrics and statistics to ensure reliable and valid results, but also by engaging in practices such as ensuring that comprehensive records are kept of the activities that actually take place during data reduction, data display and conclusion-drawing and verification. Risks of error are countered by the researcher making explicit both the process involved in expanding the framework into coding categories, and by making explicit the decision rules that guided the coding of the content of the evaluation reports. Thus, the kinds of safeguards suggested by Miles and Huberman (1984) were used in this study and are described as appropriate in the remainder of this chapter and in Chapter IV.

2. Data Collection

Data were collected during an eight month period. The process began in October 1985 and was completed in July 1986. On October 31st 1985 letters of inquiry and a one-page response sheet were mailed to the Superintendents of the seventy-five school districts in British Columbia. (See Appendix 1.)

Superintendents were asked

1. whether any formal program evaluation activities had taken place in their districts during the past five years¹⁰

¹⁰ Five years was chosen as the designated time frame for the evaluation data collected. The imposition of a time frame such as this provided one way of

2. if so, whether any of these activities had resulted in the production of a written report
3. whether copies of the evaluation report(s) could be made available
4. whether a contact person was available to answer questions about the report(s) and if so, who this person was.

Stamped, addressed envelopes were provided for returns. This first mailout, together with a follow-up mailout (sent one month later to the nineteen school districts which had not responded to the initial communication) resulted in returns from sixty-four of the seventy-five school districts surveyed. Of the sixty-four responses received from superintendents or their designates, staff from three school districts chose not to participate and did not complete the response sheet. The nature of the remaining sixty-one responses is summarized in the following paragraph.

Responses from fourteen school districts indicated that they had not engaged in formal program evaluation. Responses from forty-seven school districts indicated that they had engaged in formal program evaluation. Five of the districts that had engaged in program evaluation reported that their evaluations did not result in written reports. Responses from forty-two districts indicated that program evaluation had taken place and had resulted in the production of written reports. One district was unable to provide copies of reports. In total, forty-one districts indicated that written reports were, or might be, available to the researcher.

Two of these forty-one districts responded to the initial mailout not only by completing the response sheets but also by submitting evaluation reports. In March 1986, letters

¹⁰ (cont'd) bounding the study and ensuring the acquisition of relatively recent materials.

requesting copies of reports, ensuring confidentiality of the materials received, and answering queries raised by respondents were sent to those thirty-nine school districts which had not already submitted reports but had indicated that reports were, or might be, available.¹¹ Soon after these requests were mailed, reports were received from three districts. Follow up telephone calls requesting reports were made to the remaining thirty-six districts in April 1986. The telephone requests resulted in submissions from twenty-one districts, of which submissions from nineteen were received. (Staff at Canada Post were unable to locate the packages mailed by the other two districts.) Visits were made to six lower mainland districts, where, if the reports could not be borrowed, extensive notes on the reports were taken on-site. In total, reports from thirty school districts were examined from the available data base.

These thirty districts submitted a total of 110 reports. On arrival, each report was assigned a three digit code number from #001 to #110. The assignment of these numbers was to ensure anonymity and ease of referencing. The number of reports submitted by each district ranged from one to as many as twenty. Of the 110 reports submitted, ninety-seven were retained for the first stage of analysis. The thirteen reports rejected consisted of seven elementary self-assessments, one secondary accreditation, one duplicate report, three reports conducted prior to the designated time-frame, and one report which was submitted by a district other than that in which it was produced.¹²

¹¹ It is because of the undertakings of confidentiality that no list of reports received is included in this dissertation.

¹² Approaches to evaluative studies such as elementary school self-assessments, secondary school accreditations, or those based on student achievement and diagnostic tests are, in the main, prescribed by Ministry personnel, i.e., not conceived, planned and carried out under the auspices of the school district. Because the present study was designed to examine the practice in school districts, such studies were not included in the data collected and analyzed.

During the initial perusal of the ninety-seven reports it became apparent that some might not be suitable for inclusion in the analysis. Reasons for this included the following:

- a. the absence of an identifiable object of evaluation;
- b. incomplete reports, i.e., with pages or sections missing, or which were illegible in parts, and of which replacement copies could not be obtained;
- c. evaluation sponsorship and implementation by Ministry personnel or by evaluators contracted by the Ministry of Education (even though data were collected in specific school districts);
- d. a focus on the process of test development or standard setting for the purpose of individual student assessment;
- e. the absence of reports on all stages of a multi-stage assessment (i.e., several written reports describing these stages were produced at different times, but not all of them were submitted or could be obtained), and
- f. provision of information on evaluative activities which, although planned, had not yet taken place at the time of the submission.

A list of the twelve reports excluded on the basis of the above, together with a brief description of each, is given in Appendix 2. These exclusions, in the case of two school districts, removed from the data base all the documents submitted by those districts. Thus, the eighty-five reports retained for the analysis were submitted not by thirty, but by twenty-eight school districts.

3. The Recording of Content

The recording of content in such a way as to provide systematically comparable answers to the ten questions of the framework, was undertaken in three stages. First, detailed notes focussing on similarities and differences were made on twenty randomly selected reports. These notes confirmed the viability of the ten questions asked, identified units of content, and suggested tentative categories for the recording of this content. This stage resulted in the production of an information collection sheet. In the

second stage, this information collection sheet was used to record information from another twenty randomly selected reports in order to verify that the units of content and the categories were applicable to more than merely the first twenty reports. Again detailed notes were made on the similarities and differences among reports in terms of the list of questions and in terms of the units of content and the categories derived from the reports. The information collection sheet was adapted and used as the basis for the development of a coding instrument, intended to facilitate the third stage of content recording, namely the re-recording of information from the first forty reports and the recording of information from the forty-five remaining reports. A copy of the completed coding instrument is presented in Appendix 3. Thus, in total, forty reports were used in the development and detailed fleshing out of a system of recording, coding and categorizing the content. By the time these reports had been examined in this way, no new kinds of content and no new categories were being identified. This suggested that the system of coding could be usefully applied to all the reports retained for final analysis. A detailed description of the way the content was coded is given in Chapter IV.

4. Delimitations of the Study

There are five delimitations of this study which should be noted. First, the present study deals only with those British Columbia school districts which submitted written evaluation reports to the researcher. Formal evaluation reports were received from a comparatively small number of districts (40% of the seventy-five districts in the province). This does not mean that program evaluation does not take place in the other districts, although it may mean that program evaluation in these districts takes a different form. Alternative ways of approaching program evaluation in school districts

are not addressed in this study. The findings may not be applicable to those school districts which did not respond to the researcher's request for information.

Second, the study is delimited to consideration of evaluation as reflected in written reports. The written reports submitted are the only source of data about program evaluation in those twenty-eight districts which produced the reports examined for this study. Further to this point, it is possible that some of these reports did not wholly reflect the actual evaluation process. When reports are intended for use within the district, those involved are usually cognizant of the evaluation situations and of the normal processes for report dissemination and use. Thus, it may not have been necessary for these reports to include as much information as would be useful in cases where the readers of the reports were not familiar with school district routines.

Third, the study excludes consideration of activities subsequent to the production of a report. No information on the dissemination or utilization of the evaluation results was obtained. Thus, it was not possible to determine the impact of the report; whether, for example, the recommendations were or were not implemented.

Fourth, "program evaluation" was delimited to include only those evaluation studies initiated from within the school districts (and which resulted in the production of a written report). Other kinds of evaluation such as secondary school accreditations or elementary self-assessments, or studies initiated by Ministry of Education personnel were excluded in the design of the study. This was because the focus of the study was school district practices, thus what was of concern here was how school districts evaluate programs in the absence of specific prescriptions from the Ministry about what

or how to evaluate.

Fifth, the study covers only the period 1980-1986. There is no attempt to comment on changes in practice which may have occurred since then. Not only are the data restricted to those years, but also the framework for analysis was developed on the basis of the literature up to and including this time period.

C. SUMMARY

The first part of this chapter traced the development of the framework used for the study. First, four basic questions about evaluation were identified. Second, the list of questions, suggested by Nevo (1983) as useful for conceptualizing evaluation was modified by the removal of the normative element and by the exclusion, addition or reframing of some items. Third, the resulting set of ten questions clustered under the four general questions identified earlier was presented as the completed framework to be used in the examination of the content of the evaluation reports.

The second part of the chapter described the research design for the study. The method of content analysis, the process of data collection, and the process whereby content recording was designed, were discussed. Finally, the delimitations of the study were noted.

In essence, the study followed what Miles and Huberman (1984) have called qualitative data analysis. The methodology was inductive and was guided by the issue-oriented framework described in the first part of this chapter. The discerning of categories into which the various aspects of evaluation could be sorted, occurred primarily as a result

of examining the content of the evaluation reports themselves. The next chapter describes in detail the rules for coding and categorizing which were developed as the exploration proceeded.

CHAPTER IV.

THE DEVELOPMENT OF RULES FOR THE CODING OF CONTENT

The purpose of this chapter is to describe the rules developed for the coding and recording of content. The chapter is divided into five major sections. The first four correspond to the four clusters of questions: Evaluation — to what end? Evaluation — by what means? Evaluation — for whom and by whom? Evaluation — with what conclusion? The fifth provides general information about the reports.

In some cases the coding was straightforward. For example, information concerning the year a document was produced, the data collecting techniques used or the numbers and positions of evaluators was usually easily identifiable. In other cases, however, the coding was more difficult and did necessitate the creation of specific decision rules. For example, describing a report in terms of the definition of evaluation portrayed therein was often problematic because an explicit statement about "definition" was rarely found in the documents and it was necessary to ensure the same rules were applied in every case. This chapter provides explanations for all coding decisions. Quotations taken from the reports for illustrative purposes are cited by assigned numbers.¹³ The coding instrument which resulted from the procedures described here is shown in full in Appendix 3.

¹³ For the purpose of maintaining confidentiality, in those cases in which the quotations specify information which could identify the reports, this information has been replaced by "XX."

A. EVALUATION – TO WHAT END?

As explained in Chapter III, the first cluster of questions is concerned with the definition of evaluation, its intents, the reason the evaluation was undertaken, and the object of evaluation. Accordingly, the four questions are:

How was evaluation defined?

What were the intents of the evaluation?

Why was the evaluation undertaken?

What was the object evaluated?

1. How was Evaluation Defined?

Chapter III drew attention to the distinction between evaluation as judgement and evaluation as the provision of information. Evaluators can function, therefore, as judges or information brokers. In order to code the content appropriately, it was necessary to decide whether or not there was evidence of the evaluator, i.e., the author of the report, making judgements as to the value of the whole program or parts of the program under discussion. If there was not, then the evaluator was deemed to be acting only as information broker. A number of reports did not contain any explicit evaluative comments made by the evaluator. These were coded as, "provision of information (no judgement)." Even though no report in the first forty was found which contained the reverse situation (judgements without supporting information), the existence of such a document was considered possible. Thus, provision for coding "judgement (no supporting information)" was made.

Between these two extremes, it was found that those documents which did contain explicit evaluator judgements varied in the amount of supporting information provided.

Hence two additional categories “judgement (some supporting information)” and “judgement (much supporting information)” were included. However, there were still a number of reports which could not be coded. These were reports which were primarily concerned with the provision of information but which contained the occasional evaluative comment or word which indicated that the evaluator had indeed placed some value on the description or interpretation of observations or results. This led to the inclusion of a final category “provision of information (some judgement).” These five categories of content are listed in Table 4.1 and decision rules for coding were as follows.

TABLE 4.1
Coding Categories for Content Pertaining to the Definition of Evaluation

Judgement (No Supporting Information)
Judgement (Some Supporting Information)
Judgement (Much Supporting Information)
Information (Some Judgement)
Information (No Judgement)

a. Judgement [No Supporting Information]

This category, as described above, is for those documents which judge a program without providing some degree of supporting information.

None of the reports examined fell into this category.

b. Judgement [Some Supporting Information]

Reports coded here are those in which the evaluative purpose of the report is clear but the results and conclusions are not fully documented.

There is insufficient information for the reader to draw conclusions and compare them with those stated in the report. The reader is required to accept the ability of the evaluator to interpret the available information, i.e., there is reliance on the evaluator as expert. Some authors do provide copies of the instruments in the appendices (without the responses) and some provide illustrative quotations. In the main, however, data are summarized and recommendations, where they occur, are based on these summaries rather than on a comprehensive, all-inclusive presentation of the data. For example, in document #035, the author provides a narrative summary of staff responses to questionnaire items and bases his recommendations on this summary. Document #036 consists of a list of recommendations, each preceded by a limited number of explanatory descriptive sentences. In neither example is there sufficient information for the reader to draw his or her own conclusions.

c. Judgement [Much Supporting Information]

Documents included here are those in which the evaluative purpose of the report is made clear and the results and conclusions are fully documented in such a way that it is possible for the reader to interpret the information and arrive at similar conclusions.

The data collected by the authors of the report directly inform the evaluative purpose of the report or the specific evaluation questions identified at the outset. These reports are usually structured in such a way that the purpose or evaluation questions are given first, followed by an explication of the process used, the results, conclusions, and when appropriate, recommendations. The instruments are referenced, and if they are not included in the body of the report, they are in the appendices or there is a

statement indicating where copies may be obtained. There is a complete breakdown of results which may be in tabular form or indicated on copies of the instruments themselves. With the inclusion of all relevant information, it is possible for the reader to follow the arguments of the writers and understand why particular conclusions and recommendations are given.

d. Provision of Information [Some Judgement]

Documents coded here contain evidence of evaluator judgement, even though the making of judgements may not have been the original intent of the report.

There are two main types of study found in this category. The first is the research-oriented study in which variables are identified and their effects measured; and the second is the year-end report. For example, in their report on a quasi-experimental study, the authors of document #008 conclude that program XX "has a positive impact on student skills . . . and is, in itself an effective method . . ." (#008:11). The evaluators not only have reported the results, but they have also interpreted them and based their conclusions as to the value of the program on this interpretation.

Year-end reports, although primarily concerned with the provision of information, often contain some evaluative comments. For example, document #033 includes a single evaluative comment among many pages of description, information, and statistics:

A significant amount of progress has been made. Most of the objectives have been accomplished; or are in the process of being accomplished. The task ahead is to maintain the program now in place and to set new objectives in light of the progress attained (#033:1).

In document #046, evaluator judgement is demonstrated when activities perceived as

highlights are described. In choosing to emphasize the most successful activities, the evaluator has passed judgement on aspects of the program, labelling some as more successful than others.

e. Provision of Information [No Judgement]

Reports coded as belonging to this category are those which contain no evaluator comment as to the value of the object evaluated.

In some cases, the writers made explicit their role as information providers rather than evaluators. In reference to their conclusions, the authors of document #058 write:

These conclusions are summaries of documented facts and are not intended to be judgemental. Judgement will be applied by School Trustees and administrators (#058:48).

In other cases, the writers of the report were not so explicit about their role. Nevertheless, careful attention was paid to ensure that reports coded here contained no explicit judgements or value-attributions based on the information collected or presented.

2. What were the Intents of the Evaluation?

As noted in Chapter III, the notion of intents was given two interpretations. It was used to refer to the familiar distinction between "formative" and "summative" functions of evaluation, and it was also taken to refer to the purposes for which the evaluation was used. The term "function" was used for the first of these meanings, the term "purpose" for the second.

The following discussion deals first with the way in which "Purpose" was coded; second, with difficulties experienced in this coding of "Purpose;" and third, with the way in which "Function" was identified.

a. Purpose

An explanation of two levels of purpose which may be stated or implied in the reports and of the relationship between them is necessary for an understanding of the coding. First-level purposes are often described by **what** the evaluator is doing in the evaluation, such as providing information, investigating a program's impact, or determining stakeholder opinion. Second-level purposes are discerned by noting **why** the evaluator is doing such things as providing information, i.e., they are the purposes of first-level purposes. The following example of first and second-level purposes is taken from a report in which statements of purpose are explicit. At the first level, the purpose of the evaluation is

to observe what happens in Special Education Programs, speak with the personnel responsible for the delivery of service and meet with the community groups and organizations interested in the district's programs (#022:1).

Evaluators fulfil this first-level purpose in order that the information gathered will be used "as a basis for developing Team observations and recommendations" (p.1) which will be of assistance "in improving special education services" (p.12). In other words, the ultimate intent of gathering information is the improvement of services.

It is this second level of purpose that provides the basis for the categorization of purpose statements. In total, nine different second-level purposes were identified. These are shown in Table 4.2. Reports may be coded as having more than one purpose. Each second-level purpose is described below. Included within the explanations are the decision rules for coding in those cases in which the authors of the reports do not make their second-level purposes explicit.

TABLE 4.2

Coding Categories for Content Identifying the Functions and Purposes of Evaluation

<i>A. Purpose</i>
Improvement and Change Decision Making Development and Planning Awareness and Knowledge Accountability Requirements of Funding Agency Informing the Policy Process Provision of Comparative Data Other Not Stated
<i>B. Function</i>
Formative (Specified) Formative (Implied) Summative (Specified) Summative (Implied) Formative (Specified) and Summative (Implied) Formative and Summative (Both Specified) Formative and Summative (Both Implied) No Statement Made, No Inference Possible

■ **Improvement and Change**

Reports with Improvement and Change as a second-level purpose answer the implicit question "What can be done better?" Reports coded under this heading are those in which most, but not necessarily all, of the following conditions apply:

- termination of the program is not suggested*
- the unit of analysis is the program component (rather than the program in its entirety)*
- improvement or change is given as an explicit intent*
- the evaluation's formative function is ascribed by the author or is evident from the text (e.g., a statement occurs as to the evaluation's part in the ongoing review process)*

- *questions and comments related to improvement can be identified*
- *recommendations which suggest change are suggested.*

The intention is to provide feedback to program or to school district personnel so that modification of the components takes place. Typically, reports which describe evaluations of programs at the end of their first year of operation were coded here. In these cases the information collected is intended to be used formatively for the improvement of the program in its second year of implementation.

■ **Decision Making**

In contrast to reports coded as "Improvement and Change," those coded here deal with the entire program. Reports grouped under this heading are those in which most, but not necessarily all, of the following conditions apply:

- *there is a summary statement concerning the value of the program as a whole*
- *continuation of the program is recommended*
- *the unit of analysis is the program as a whole*
- *the decision to be made is clearly evident, i.e., is stated or can be inferred from the text*
- *the evaluation's summative function is ascribed by the author or is evident from the text*
- *questions and comments related to the overall effectiveness of the program or its success in meeting its goals and objectives can be identified.*

Summative information about whether or not the program is successful is provided and used to inform decisions. These decisions affect the fate of the program in its entirety: they determine whether or not the program should be continued, expanded or introduced into other school district sites, or instead, if alternatives to the present offerings should be considered. Program pilots and reports which describe evaluations of programs at the end of their second year of operation are frequently found here.

■ Development and Planning

Reports with second-level purpose statements which are intended to result in development and planning and have an orientation towards the long-term future (generally more than one year in advance) are coded here.

The major difference between this purpose and others is time orientation. Needs assessments, studies intended to determine priorities for future action as well as those concerned with major program restructuring are included. In addition, there are those considering the effects of projected population trends, or those which project the consequences of policies and projected population growth on services in the school district. Other purpose statements categorized here include such activities as providing direction for the school district, recommending a long-range plan, and identifying areas requiring attention in and beyond the next school year.

■ Awareness and Knowledge

Reports coded here are those in which

- *raising awareness and increasing knowledge are specified in purpose statements and are clearly not only first-level purposes*
- *the purpose statements may also refer to the provision of descriptive information from which raised awareness and increased knowledge and understanding will result.*

Examples are: the examination of a particular situation, the description of an object, its portrayal from a number of perspectives, or the documentation of program processes.

■ Accountability

Reports coded here are those which make specific mention of accountability or which state that the results of the evaluation will provide the public with outcome measures such as student achievement scores.

Although it may be argued that, by definition, the production of evaluation reports and their submission to elected public representatives (trustees) is a demonstration of accountability to the public, this category includes only those reports in which this intention is made explicit.

■ **Requirements of Funding Agency**

Reports coded here are those which make specific mention of the requirements of the funding agency.

It is accepted practice for funding agencies to receive reports on the programs to which they have contributed, and reports may allude to the funding available from the Secretary of State for French programs, or to the Provincial government funds for special programs. Although they rarely state that the purpose of the report is to meet funding requirements, the reports do sometimes mention it in the introductory paragraphs as part of the rationale or mandate for the evaluation.

■ **Informing the Policy Process**

What distinguishes these reports from those with development and planning as a second-level purpose, is that specific mention is made of policy and of the implications for policy contained in the documents.

In some cases, the consequences of present Board policies may be anticipated and in others, the information collected may be intended for use in future policy development. Some reports coded here determine whether the program is in compliance with Board policy or whether Board policy needs to be reviewed or modified in any way. Others determine whether the program is in compliance with not only Board policy, but also with Ministry of Education guidelines and Ministry of Human Resources guidelines.

■ **Provision of Comparative Data**

Reports are coded here if they contain a statement as to their utility by providing "benchmark" or "base-line" data which can be used for comparative purposes at a later date.

■ **Other/Not Stated**

The category "Other" was created in case second-level purposes not noted above emerged.

Cases in which it is not possible to identify the first or second-level purpose of the report are coded as "Not Stated."

The only other purpose to emerge was "Validation" which referred to the validation of

a program or aspects of a program, or the validation of criteria for the evaluation of a program.

b. Difficulties in Coding Purpose Statements

The coding of purpose statements may be hampered by lack of information. For example, a report may be intended to inform the policy process or meet funding requirements even when these second-level purposes cannot be discerned from the text. For the present study, however, coding was based on sections of text where purposes, if not explicit, were clearly identifiable. Not only may purposes vary, but so also may their format and number. Some reports have their purpose statements embedded in introductory paragraphs, while others list several discrete purposes.

The following example illustrates how several purposes at one or both levels can be identified in any given statement. For example, the purpose of document #084 may be discerned from the final sentence of the second paragraph:

At the direction of the Board, an appraisal of the XX Program was to be conducted in the spring of 19XX to determine the future direction of the program (#084:1).

The purpose of the evaluation is to appraise the program (first level) in order to determine future direction (second-level purpose). The third paragraph reads:

Apart from the basic decision to continue or discontinue the XX Program at XX School, the evaluation team determined to make a number of recommendations for the consideration of the staff in its endeavours to improve the content of the existing program (#084:1).

Here, the reader learns that within the determination of future direction, there are two second-level purposes: one, to continue or to discontinue the program (decision-making),

and two, to make recommendations to improve the program (improvement).

Some reports, such as document #004, have multiple statements of purpose:

- (a) to describe the operation of the XX Program and the services provided,
- (b) to evaluate the effectiveness and the viability of the XX Program in serving the community,
- (c) to investigate XX Programs in other school districts and attempt to determine their trends and directions,
- (d) to make recommendations where appropriate for improvements in the XX Program (#004:8).

The first-level purposes are as stated. The second-level purpose of the first purpose would appear to be "awareness and knowledge," although this is not mentioned in the text. The second-level purpose of the remaining three purposes is "improvement," as apparent from the focus of the recommendations in the third purpose statement.

c. Function

The categories used for coding "Function" are shown in Table 4.2 (p.68). The formative function is one in which the evaluation is intended to answer the general question, "What can the program do better?" while the summative function is intended to answer the question, "How well has the program done?" When either or both functions were made explicit by the evaluators, the functions were coded as "formative (specified)," "summative (specified)," or "formative and summative (both specified)." In the following example the evaluators have made explicit both the formative and the summative functions.

- The evaluation of the XX centre was intended to be both (a) formative and (b) summative in that:
- (a) the XX staff would receive continual feedback on the services that they have provided to assist in planning, and
 - (b) information on the services of the XX centre in meeting its goals would be summarized in two reports to the Board of School Trustees (#098:4).

However, formative and summative functions were not always made explicit. In those cases in which these functions were implied it was necessary to develop decision rules as follows:

■ **Summative (implied)**

Reports coded here fulfil most of the following conditions:

- *they make a judgement about overall program success*
- *purpose statements include such phrases as "to determine the effectiveness of," "to assess the success of the project," or "to determine if the program provided the services to the school that it was intended to provide"*
- *emphasis is placed on outcomes (or products)*
- *recommendations for program continuation or termination are included.*

■ **Formative (implied)**

Reports coded here meet most of the following conditions:

- *their purpose is improvement*
- *their formative function is evident from the text, i.e., there are comments as to the utility of the report in terms of sharing information with program personnel (often during the course of the evaluation itself, i.e., prior to the submission of the report) for use in the ongoing improvement of the program*
- *emphasis is placed on process*
- *recommendations for improvement of process are included.*

In the following example, the formative and summative functions are implicit.

The purpose of this report is to convey to you our analysis of the XX delivery model in XX with attention to the appropriateness of its goals and the degree to which it meets these goals, together with suggestions to help improve your programs (#034:1).

Determining the degree to which a program meets its goals is a summative function, while offering suggestions for improvement is a formative one.

In those reports in which formative or summative attributions are not possible to infer from the text, the designation "no statement made, no inference possible" is given.

In some instances, classifying the reports as summative, formative, or both, was

problematic. This was primarily because the authors of the reports used the terms in different ways. For example, it appeared that some evaluators believed that summative evaluation was a necessary precursor to formative evaluation. Thus, if the intent of the evaluator was to identify the aspects of the program which could be improved (formative), it was necessary to determine how well that program was functioning (summative) first, in order to identify both successful activities and less successful activities which were in need of improvement. In other cases, formative and summative functions appeared to be discrete. For example, some reports provided information which was intended to provide program personnel with useful feedback on ways that they could improve their program but contained no statements which could be interpreted as summative in nature. Those reports to which a summative function could be attributed contained some kind of statement attesting to the value of the whole program and no suggestions for improvement. One further interpretation of summative was also identified. In this case the evaluator had used the word "summative" in the sense of summarizing, i.e., descriptive information was presented but no summative statement, in terms of assessing value, accompanied the description. As the decision rules (listed above) were used for coding, these difficulties did not affect the way in which reports were assigned to categories.

3. Why was the Evaluation Undertaken?

This question is concerned with the circumstances surrounding the initiation of each evaluation. The content of the reports showed that evaluations could be required by policy or practice, or undertaken in response to a particular request, or both. (See Table 4.3.) Decision rules for coding are as follows:

■ **Required**

Coded here are those reports which were required either by school district policy or established practice, or by the design of the program itself or as part of a school district short term objective.

■ **Requested**

Those reports done in response to a request from one or more of a variety of stakeholders or stakeholding groups are coded here.

TABLE 4.3

Coding Categories Showing the Reasons for Evaluation

<i>Reason Identified</i>	<i>Broader Category</i>
Required by: SD Policy Established Practice Program Design Short-term SD Objective	Required
Requested by: Program Sponsor School Staff Trustees Senior SD Administrators Representative/Evaluation Committee Other	Requested
Not Stated	

4. What was the Object Evaluated?

The initial reading of the reports indicated ways of answering this question. First, specific programs, services, facilities and the like were designated as the objects evaluated. Second, however, these objects could also be grouped according to other characteristics of the objects evaluated which were identified in the reports. Accordingly, two kinds of answers were recorded: first, the object itself, second, certain additional characteristics. The following paragraphs deal with each of these in turn.

a. Type of Object Evaluated

Table 4.4 shows the types of objects categorized. The search for the object evaluated resulted in a long list, the majority of which were programs or program-related practices. The remainder were categorized as "Organizational Unit" and "Facilities." The objects coded within the broad category of "Program or Program-related Practice" were subdivided into four categories: "Curriculum and Instruction," "Testing," "Special Services," and "Organization for Instruction." The objects are listed in Table 4.4 and decision rules for each of the categories pertaining to type of object follow.

■ Program or Program-related Practice

▼ *Curriculum and Instruction*

Those reports which evaluate some aspect of the regular curriculum are coded here.

▼ *Testing*

Those reports evaluating the results of district and provincial examinations within districts are coded here and are listed by subject area.

TABLE 4.4

**Coding Categories for Content Identifying the Type of Object
Evaluated**

<i>Reported Object</i>	<i>Broader Categories</i>	
French, Counselling, Computers, Personal Safety, Career Preparation, Science, Music, Library, Learning Assistance, Reading, Physical Education, International Baccalaureate, Police Liaison, Consumer Education, Art/Science, Continuing Education, Instructional Innovation, Grading Practices, Other	Curriculum & Instruction	
Writing, Math, Reading, Chemistry, Other	Testing	Program or Program-related Practice
Special Programs (Comp.), Native Programs (Comp.), Native Program, Alternate/Rehabilitation, Gifted/Talented, Learning Disabled, Behaviour Disturbed, Hospital, Residential, Multiple Handicaps, Hearing Impaired, TMH, Other	Special Services	
Split Grades, Computer Aids, Other	Organization for Instruction	
School, Centre, Other		Organizational Unit
Portables, Space, Other		Facilities

▼ *Special Services*

Reports which evaluate a program which is accessible only to certain groups of students are coded here.

▼ *Organization for Instruction*

Reports which evaluate ways of organizing groups of students and of approaching the teaching task, rather than programs, are coded in this category.

■ **Organizational Unit**

Reports which evaluate whole schools and centres (centres are places where students go for various kinds of assessment and assistance), rather than individual programs, are included here.

■ **Facilities**

Reports in which the object evaluated is some sort of physical facility, or use of the space within a facility, are coded here.

b. *Additional Characteristics of Objects Evaluated*

During the straightforward process of identifying the type of object evaluated, a number of other characteristics of the objects became clearly discernible. These are shown in Table 4.5 and are discussed below.

■ **Permanence of Object Evaluated**

The objects evaluated were either continuing (in that they included established programs, organizational units or permanent facilities), or temporary (in that they were set up for a limited time). Decision rules are as follows.

▼ *Continuing*

Reports which evaluate programs that are offered on a regular basis (usually those which are prescribed or recommended by the Ministry of Education) are coded here.

▼ *Temporary*

Reports which evaluate programs that are offered for a limited time period in order to perform a specific educational task are coded here.

Examples include: "projects" (activities designed to perform an educational function, but not expected to continue for an extended period of time), "pilots" (projects used to test,

TABLE 4.5
**Coding Categories for Content Referring to Additional
 Characteristics of Objects Evaluated**

<i>Example of Content</i>	<i>Characteristic</i>
Continuing, Pilot, Project, Display, Equipment, Other	Permanence
General, Service Delivery, Impact, Needs Assessment, Planning, Student Life, Consistency, Reading, Other	Aspect Evaluated
School, SD, Joint Sponsorship, Non-SD Agency	Object Sponsor
School or Centre, SD, School & SD, Non-SD Facilities	Object Base
Primary, Intermediate, Elementary, Elementary/Junior Secondary, Junior Secondary, Senior Secondary, Secondary, Elementary/Secondary, Adult	Grade level

on a small scale, how programs will work if or when implemented), and “displays” (exhibits of content related materials intended to fulfil educational purposes).

■ **Aspect of Object Evaluated**

Evaluators may provide evaluations of objects in their entirety or they may focus on a particular aspect. Thus, an evaluation may be “General” or “Specific.”

▼ **General**

Reports are classified here when the evaluation provides a general, global view of a number of different facets of the object.

▼ **Specific**

Reports are coded here when they show that the evaluators are explicitly concerned with one or more particular aspects of the object under consideration.

Examples of specific aspects include: delivery of services, impact, needs assessment,

forward planning, student life, consistency of grading practices, and reading as part of a language program.

■ Object Sponsor

All reports were coded as to who sponsored the object of the evaluation. Four kinds of sponsor were identified: a school or centre, the school district, a combination of sponsors or an outside agency.

▼ *School or Centre*

Included in this category are reports in which the object of evaluation is available only at specific schools or centres.

▼ *School District*

Included within this category are reports in which the object of evaluation is one of a wide variety of general instructional programs or special education programs available throughout the district.

▼ *Joint Sponsors*

Included in this category are reports in which the object of evaluation is sponsored by more than one group.

For example, objects may be sponsored by two school districts or by a school district and an agency outside the school district, such as the federal government.

▼ *Non-district Agency*

Included in this category are reports in which the evaluated object is sponsored by an outside agency, such as a private foundation.

■ Object Base

This refers to the physical site of the objects of evaluation; four sites or object bases were identified.

▼ **School-based**

Reports which focus on objects which are school-based are coded here.

To be coded as "school-based," an object must:

- ▼ *be sponsored by a school and be evaluated within the context of that school,*
- ▼ *have space either in school buildings or on the grounds of a school, or*
- ▼ *be self-contained in a building which, although not within the buildings or grounds of a school, is still known as a school (alternate schools fall into this group).*

Pilot studies, projects and a variety of both general instructional programs and special programs are coded as school-based.

▼ **District-based**

Reports coded here are those which describe evaluations of objects which are in all schools or are available to all schools across the district.

Thus, District-based objects are those which may:

- ▼ *occur in all schools serving a particular age-range, e.g., all elementary or all secondary age students (reports on district-wide testing are subsumed here),*
- ▼ *be available to all schools in that schools or individual teachers may volunteer to participate in them,*
- ▼ *be staffed by itinerant personnel, or*
- ▼ *take place in non-traditional sites, i.e., store-front or hospital programs, which are, nevertheless, supported by the school district.¹⁴*

▼ **School and District Based**

Reports coded here are those in which the objects of the evaluations are present in selected school sites and in schools across the district.

This occurs when the object being evaluated is a composite of a number of different programs.

¹⁴ When hospital and homebound programs have been included as part of a composite report, the object base ascription is "school and district."

▼ *Non-district Facilities*

Reports coded here are those in which the object of the evaluation is located in premises which are not owned or leased by the school district.

For example, a report evaluating a summer work experience, whereby students worked in business and industry, was coded here; as was a report on a residential program.

■ **Grade Level**

In all the reports, reference was made to the grade level (broadly, as elementary or secondary, or specifically to particular age-groups or year-groups) of the students involved in, or affected by the object of the evaluation. Thus the category "Grade-level" was developed.

Reports were coded according to the grade levels specified by the evaluators.

Where age-groups, year-groups or numerically ascribed grade levels (rather than the grade level descriptors listed) were mentioned, the reports were coded by the researcher at the appropriate level. For example, a report specifying grades 7-9 was coded as Elementary/Junior Secondary.

B. EVALUATION – BY WHAT MEANS?

The second cluster of questions is concerned with the kinds of information collected, the criteria used for evaluation and the methods of inquiry used by the evaluators:

What kinds of information regarding each object were collected?¹⁵

What criteria were used to judge merit and worth?

What methods of inquiry were used in the evaluation?

¹⁵ As this study concerns the analysis of written materials, it seemed more appropriate to examine the kinds of information reported to have been collected, i.e., contained within the documents. For this reason the decision was made to replace "collected" with "reported" in this question.

1. What Kinds of Information Regarding Each Object were Reported?

During the initial reading of the documents, it became apparent that to speak of “kinds of information” could be to speak of either the source of the information collected or its nature. Accordingly, these two aspects were treated separately. The coding categories for the “Source” of the information reported are displayed in Table 4.6. The coding categories for the “Nature” of this information are displayed in Table 4.7.

a. Source of the Information

From the documents it was possible to identify three kinds of source of information upon which the evaluations were based: people (other than the evaluator), existing written materials, and the evaluator him or herself.

■ **People other than the Evaluator**

All the reports were coded according to the respondents (i.e., people who provided information to the evaluators). Three such categories were identified: “Stakeholders (school district employees),” “Stakeholders (non-school district employees),” and “Externals.” “Stakeholders” are those individuals who are affected in some way by the object to be evaluated, the evaluation process, or by the outcomes of that evaluation. Each individual, therefore, has a “stake” in the evaluation. “Externals” are those who are not connected with the school district and have no stake in the specific program under consideration. The coding instrument shows who was allocated to which category.

■ **Existing Written Materials**

Reports were coded according to two kinds of written information collected:

TABLE 4.6

**Coding Categories for the Identification of the Source of
Information Reported**

<i>Source</i>	<i>Broader Categories</i>	
Senior SD Administrators, SD Office Staff, Program-operating Personnel, Program-using Personnel, Uninvolved Personnel, School-based Administrators, Other, Not Stated	Stakeholders (SD Employees)	
Trustees, Parents, Members of Local Organizations, Community Members (at large), Faculty (University/College), Provincial Ministry Employees, Student Program Participants, Non-program Participants (Students), Post-program Participants (Students), Other, Not Stated	Stakeholders (Non-SD Employees)	People Other than Evaluator
SD Staff from other SDs, Faculty/Graduate Students (University/College), Private Consultants, Representatives from Provincial Ministries, Outside Experts (Program Area), Other, Not Stated	Externals	
Documents & Records, Literature Reviews, Not Stated		Written Materials
Evaluator Observations, Evaluator Assisted Tasks, Other, Not Stated		Evaluator

▼ *Specific*

Specific documents and records about the object to be evaluated, and

▼ *General*

Published literature on the general area pertinent to the object.

Reports coded as "Specific" include such written materials as student records, and program and policy documents at school, district and Ministry levels; reports coded as "General" include reference to reviews of the literature, or to annotated bibliographies.

■ **Evaluator**

Reports were coded in terms of how the evaluators, themselves, provided information. This information was of two kinds:

▼ *Observations of the program in action,*

in which the evaluators collect information first-hand from observing programs in progress; and

▼ *Evaluator assigned tasks or tests,*

whereby evaluators collect information by assigning particular tasks to program personnel and program participants.

b. Nature of the Information

The reports varied considerably in the generality and specificity of their information. When information was specific, it was of two kinds, information about people's opinions or more objective descriptive information. Within both these kinds of specific information, a wide range of topics was treated. It was possible to group these topics under the headings of "Process," "Outcomes," "Participants" and "Similar Objects in Other Sites." Table 4.7 shows these various groupings and rules for coding follow.

TABLE 4.7

**Coding Categories for Content Pertaining to the Nature of
Information Reported**

<i>General</i>	<i>Opinion</i>	<i>Specific</i>	<i>Descriptive</i>
Exclusively General Predominantly General	Re: Process Outcomes Participants Similar Objects in Other Sites Other	Re: Process Outcomes Participants Similar Objects in Other Sites Other	

■ **General**

Those reports categorized as "General" are those in which the evaluators include insufficient specific information from primary sources to warrant its categorization in topic areas or as opinion or description. Although the evaluator may have based his or her evaluation on such specific information, it is not included in the report.

Two further divisions ("Exclusively General" and "Predominantly General") were found to be warranted.

▼ ***Exclusively General***

Reports coded here are those in which much that is reported comes from secondary sources, and information which is obtained from primary sources is summarized.

Here, there may be no indication of who provided the information or what questions were asked.

▼ ***Predominantly General***

Reports coded here are similar to those identified as “exclusively general” except that, in addition, they do include some specific information obtained from identified primary sources.

Thus, an evaluator may single out specific examples of activities or responses for comment; for example, in document #110, the evaluator describes her general impressions of the program and then focusses on certain aspects illustrating them with quotations from respondents.

■ **Specific Information (Opinion)**

Reports which provide information about how stakeholders regard the object of the evaluation are included in this category.

This includes their opinions of its strengths and weaknesses, and their suggestions as to how it could be improved.

■ **Specific Information (Description)**

Reports which provide descriptive information about certain aspects of an object, such as how a program worked, who its clients were or student scores on measurable program outcomes, are coded in this category.

The information contained here is less subjective than in the “Opinion” category.

■ **Topics about which Opinion and Description were Reported**

In the reports, both opinion and descriptive information were often provided about the object of evaluation in terms of how it operated and the effects it had. Thus, both opinion information and descriptive information were given about the following topics: “Process,” “Outcomes,” “Participants” and “Similar Objects in Other Sites.” Decision rules for topics follow.

▼ **Process**

Reports coded here include information pertaining to program background, program mandate, and program operations.

Opinion information on process includes stakeholder opinion of such items as which program goals and objectives should take priority; the relevance of program content to further post-secondary training; or stakeholder perceptions of the adequacy of facilities and equipment. Descriptive information on process includes lists of program goals and objectives; information on course content, program organization and program procedures; or on program enrollment figures.

▼ **Outcomes**

Reports coded here include information on what has occurred as a result of the program.

Opinion information on outcomes includes, for example, whether or not staff think that there have been shifts in the behavioural characteristics of students because of their experiences in the program; whether or not stakeholders perceive that participation in the program prepares students adequately for entrance into post-secondary training; or whether or not stakeholders hold positive views of the impact of the program on staff and students. Descriptive information on outcomes includes records of student academic achievement; records of numbers of students completing the program together with their post-secondary activities; or scores on tests of basic academic skills, responses to work attitude surveys and life-skills checklists.

▼ **Participants**

Reports coded here include information on individual students and staff involved in a program.

Opinion information on participants is rarely given; however, the occasional example can be found, as in one document which provides staff opinion on

individual student suitability for entry into a program. Descriptive information includes staff and student profiles which record staff qualifications and teaching experience or individual student abilities and interests.

▼ *Similar Objects in Other Sites*

Reports coded here include information on similar programs in other locations.

Program personnel from other sites can often provide useful information about their programs which can be used for comparative purposes. Opinion information includes the views of staff members working at these sites and descriptive information includes information on program processes.

2. What Criteria were Used to Judge Merit and Worth?

Examination of the reports resulted in two general observations about criteria: first, that criteria for evaluation in school districts are more frequently implicit than explicit, and second, that many evaluation criteria may be applied to any particular object. This made the identification of criteria difficult. It was necessary, therefore, to find ways of identifying the criteria which were implicit in the texts of the reports. When evaluators engage in evaluation they focus on some areas more than others. Clearly, the areas on which they choose to focus are of importance to them, and these areas of focus were taken for purposes of this study as implicit criteria. Areas of focus were found throughout the reports. A first operational question, therefore, was where to look in the reports for indications of criteria. As answers to this question began to emerge, they were recorded and coded, thus yielding a category labelled "Location of Criteria." A second category resulted from the observation that a number of reports explicitly or implicitly indicated where the criteria for evaluation had originated. This observation led to the creation of the broad category "Source of Criteria." Finally, it was found that

the criteria identified in the reports focussed on different aspects of the object of the evaluation. Thus, the broad category "Nature of Criteria" was created.

a. Location of Criteria

Coding categories for content identifying the location in the document of criteria for judgement are listed in Table 4.8. Decision rules follow Table 4.8.

TABLE 4.8
Coding Categories for Content Identifying the Location in the Document of Criteria for Judgement

Identified Criteria
Checklists
Goals/Objectives
Evaluation Questions
Evaluation Introduction/Abstract
Interview/Questionnaire Items
Recommendations
Evaluation Summary
Measurement Data
Other

■ **Criteria identified as such**

Reports in which criteria were explicitly identified are coded here.

■ **Checklists**

Those reports which incorporate checklists or those in which in-district evaluators follow Ministry of Education guidelines are coded here.

These reports frequently contain charts with lists of incomplete statements (grouped under headings such as staff, materials and equipment, facilities, program organization, program implementation and evaluation) which refer to various aspects of the object.

These statements are arranged opposite columns headed "highly satisfactory," "satisfactory," "unsatisfactory," and "non-applicable." The evaluator places check marks in the columns which, in his or her opinion, most appropriately completes the statement.

■ **Goals and Objectives**

Reports in which criteria can be identified in lists of goals and objectives are coded here.

For example, if a program objective is to improve the academic and social skills of students, pertinent criteria would be academic achievement and social skills.

■ **Evaluation Questions**

Reports in which criteria can be inferred from questions are coded here.

For example, in questions such as "Is there a change in academic performance of students during their stay in the program?" or "Is there a behaviour or attitude change in students where behaviour and attitude have been identified as a problem?" criteria may be identified as student achievement, student behaviour and student attitude respectively.

■ **Introductory Statements/Abstract**

Reports coded here are those in which criteria may be inferred from the evaluator's abstract or introductory statements.

At the beginning of report #038, for example, the evaluators list three areas for investigation: (a) awareness and familiarity with program personnel, facilities and procedures; (b) validity, appropriateness and utilization of services offered to students, parents and teachers; (c) the performance and standard of service offered by program personnel.

■ **Interview/Questionnaire Items**

Reports coded here are those in which criteria can be inferred from interview protocols or questionnaire items.

For example, from an item such as “Does your child speak French at home?” it may be inferred that the evaluator considers the activity of speaking French at home to be a criterion for making a judgement about the child’s French program.

■ **Recommendations**

Reports coded here are those in which the criteria are implicit in the recommendations.

In the following recommendation, the implicit criterion is equality of access, i.e., all children have the right to participate in all areas of the school curriculum.

. . . that¹⁶ consideration be given to the restructuring of school organization so that XX children miss such curricular activities as physical education and music less often (#022:7).

■ **Summary Statements**

Reports coded here are those in which key criteria are included in the summary statements.

For example,

In summary, there is clear evidence that the XX program is functioning well in this district. It is meeting the expectations of the XX parents and it is producing very good levels of student achievement (#110:15).

Two criteria for judgement are indicated in this quotation, parental expectations and student achievement.

¹⁶ Evaluators vary in the way they choose to present recommendations. “. . . that” is used throughout in order to achieve consistency in the documenting of recommendations provided for illustration in this chapter.

- **Measurement Data**

Reports coded here are those in which implicit criteria can be identified in measurement data such as achievement scores and ratings on attitude scales.

Here, the pertinent criteria are student achievement and student attitude.

- **Other**

Reports coded here are those in which criteria can be identified in locations other than those listed above.

b. Source of Criteria

The coding categories for content identifying the source of criteria for judgement are listed in Table 4.9. Decision rules follow.

TABLE 4.9

Coding Categories for the Content Identifying the Source of Criteria for Information

SD/Ministry Guidelines Goals/Objectives Terms of Reference Identified by Evaluator Alternate Objects Other

- **Written School District or Ministry Guidelines**

Reports in which criteria are derived from Ministry or school district documents produced by experts in the educational field are coded here.

- **Goals and Objectives**

Reports in which criteria are derived from the stated goals and objectives of a particular program are coded here.

- **Terms of Reference**

Reports in which criteria can be traced to the particular values of the clients or other stakeholders are coded here.

For example, in those cases in which the mandate for the evaluation is given to the evaluators by the clients or in which the terms of reference are arrived at jointly by evaluators and clients, the criteria can be traced to what they perceive to be important and thus wish to have examined in the ensuing evaluation.

- **Identified by Evaluator(s)**

Reports in which no source of evaluative criteria is indicated other than the values and expertise of the evaluator are coded here.

Evaluators are often considered as experts and are given the authority to choose pertinent criteria.

- **Alternate Objects**

Reports are coded here if the criteria for judgement originate from alternate, but similar objects.

When, for example, an exemplary program in another district is visited by personnel operating a similar program in their own school district, points of comparison may be identified.

- **Other**

Reports coded here are those in which it is not possible to identify the source of the criteria.

c. *Nature of Criteria*

Coding categories for content identifying the nature of criteria for judgement are listed in Table 4.10. The criteria identified in the reports focussed on the process involved in an evaluation and on the outcomes of the evaluation, hence, two broad categories "Process" and "Outcomes" were identified. Decision rules follow.

TABLE 4.10

**Coding Categories for the Content Identifying the Nature of
Criteria for Judgement**

<i>Observed Criteria</i>	<i>Broader Category</i>
Adherence to Guidelines, Philosophy, Program/Procedures, Policy/Administrative Practices, Personnel, Instructional Practices, Professional Development, Buildings/Facilities, Materials/Equipment, Evaluation/Research, Community Relations, Other	Process
Student Achievement, Student Behaviour/Attitudes, Changed State, Stakeholder Satisfaction, Other	Outcomes

■ **Process**

Reports in which the object is evaluated by means of criteria concerned with its operation are included here.

Sometimes criteria were identified in relation to adherence to guidelines, sometimes in relation to particular areas of focus. The decision rules follow.

▼ **Adherence to Guidelines**

This category includes reports in which a criterion can be identified indicating the extent to which written guidelines (either from the school district office or the Ministry) are followed.

Here, the congruency between a guideline from a level outside the program, and action on the level of the program becomes important. Thus, evaluators can base their perceptions of program success on an assessment of whether program staff have or have not followed these guidelines.

▼ *Various Areas of Focus*

In addition to the assigning of reports in terms of the extent to which programs adhere to guidelines, reports could also be categorized according to the particular areas of focus that evaluators chose to address. Eleven areas of focus were identified and are listed in Table 4.10. These are not, strictly speaking, criteria, but factors which evaluators deemed important to examine in providing information for evaluating programs. These categories are fairly straightforward, but a full account of them is given in the final section of this chapter under "Recommendations" where more explanation is warranted. In order to avoid redundancy, they are not discussed here.

■ **Outcomes**

Reports in which the object is evaluated by means of criteria concerned with its outcomes are included here.

The kinds of criteria are described below.

▼ *Student Achievement*

Reports coded here include those in which student achievement is used as a criterion for judging the success of a program.

▼ *Student Behaviour and Attitudes*

Reports coded here include those in which student behaviour and attitudes are used as criteria for the evaluation of a program.

▼ *Indications of a Change in State*

Reports coded here include those in which indicators of a change in state, usually an improved or desirable one, are used as criteria for the evaluation of the object. To be coded here, it is necessary for a base-line standard to be identified.

Examples include a decrease in the drop-out rate or in the numbers of students referred for disciplinary action; the development of procedures when none existed before; or an increase in the number of community members visiting a school.

▼ *Stakeholder Satisfaction*

Reports coded here include those in which data on aspects of a program's worth as perceived by its stakeholders are collected and used as determinants of a program's success.

▼ *Other*

Reports which refer to outcome criteria not included above are subsumed here.

3. What Methods of Inquiry were Used in the Evaluation?

In the reports, this question was answered in two ways. First, the authors of some reports claim to have followed specific approaches, including those associated with a particular writer noted in the evaluation literature. Second, the evaluators used a number of different methods of collecting data. Thus, two kinds of answer were coded separately, as shown in Table 4.11. The number of data collection techniques used was also recorded.

a. *Approach Ascribed by Evaluators*

The approaches made explicit by the evaluators are described below.

■ **On-site Visit**

Reports coded here are those in which the evaluators write that the evaluation was carried out by means of one or more visits to the program site.

The observations of the program in action and the interaction with program personnel and participants which takes place at this time form the basis for the evaluation.

TABLE 4.11

Coding Categories for Content Pertaining to Methods of Inquiry

<i>A. Approach</i>
On-site Visit Survey Description Experimental/Quasi-experimental Situational Interpretation Judgement Matrix Other Not Stated
<i>B. Data Collection Techniques</i>
Questionnaires Interviews/Meetings Documents/Records On-site Observation Off-site Inquiries Checklists/Rating Scales Attitude Scales/Inventories Achievement Measures Other

- **Survey**

Reports coded here are those in which the evaluator states that the evaluation consists of a survey of one or more groups of people.

- **Description**

Reports coded here are those in which the evaluator identifies the approach as descriptive in nature.

- **Experimental/Quasi-experimental**

Reports included here are those which are identified by the evaluator as being experimental or quasi-experimental.

- **Situational Interpretation**

The one report included in this category made specific mention of this approach.

“Situational Interpretation” is intended to uncover the relevance and meaning of a program for those involved. In this way the perceptions of groups such as students, parents, teachers and administrators are made explicit.

- **Judgement Matrix**

The one report included here described how Stake’s Judgement Matrix with its four categories of intents, transactions, standards and judgement was modified for use in a specific evaluation situation.

- **Other**

Reports coded here are those which make explicit some approach which does not fit the other categories.

- **Not Stated**

Those reports which do not make explicit any approach are coded here.

b. *Types of Data Collection Techniques*

As shown in Table 4.11, reports differed as to the kinds of techniques used for collecting data. Decision rules follow.

- **Questionnaires**

Reports coded here are those in which respondents are asked for their opinions on a variety of aspects of the object under consideration and may be given the opportunity to make general comments.

The kind of data collected might include: profiles of respondents; their descriptions of aspects of the program; their perceptions of the relative importance of program objectives, or of the strengths and weaknesses of the program; their general impressions or their specific responses to particular procedures.

■ Interviews and Meetings

Reports coded here include those in which face-to-face meetings are held between the evaluator and individuals or groups in order to elicit in-depth information about the object under investigation, about its operation and about how it is perceived by those involved.

These can be highly structured and relatively formal, or can take the form of informal discussion sessions in which the agenda is determined more by the respondents than by the evaluators. Concerns and issues may be identified and suggestions given for change.

■ Documents and Records

Reports coded here are those in which information is obtained from written sources. Documents provide information about recent developments in research; about similar activities in other districts; about guidelines for implementation within the district; and about the normative operation of programs on-site. Records provide information about actual program processes and outcomes.

Examples of the former include published reports, Ministry directives and information brochures designed to inform parents about school district activities. Examples of the latter include data on attendance, referrals for district assessment, report cards, Individual Educational Plans (IEP's) and unsolicited letters from parents.

■ On-site Observation

Reports are coded as on-site observation only when observations of the program in action are specifically mentioned as a data collecting technique by the evaluators.

Although some evaluators produced in-depth descriptions of particular activities observed within a specified time-frame, in the main, evaluators tended to make on-site visits both to obtain a general idea of how the program was put into practice and to meet with program personnel.

■ Off-site Visits and Inquiries

Reports are coded here if there is an indication that the evaluators have collected data by visiting similar programs in different sites; by obtaining information about other programs by telephone; or by attempting to contact students for the purpose of post-program follow-up.

■ Checklists, Rating Scales, Attitude Scales and Inventories

Reports which document the use of checklists, rating scales, attitude scales and inventories are included here.

Checklists and rating scales include teacher ratings of such items as students' life-skills, social skills or study habits, or behaviour profiles of particular program participants. Checklists and rating scales can also provide student self-esteem or self-concept ratings, while attitude scales and inventories can be designed to elicit stakeholder opinions about certain aspects of the object of the evaluation.

■ Achievement Measures

Reports which include outcome measures related to student achievement are coded here.

Three specific kinds of outcome measure were identified in the reports. The first, and most frequently used, is the achievement test. Such tests include standardized tests, modified standardized tests and teacher-made tests. Examples include: the Canadian Tests of Basic Skills, the Gates-MacGinitie Reading Test, the Arlin Test of Formal Reasoning, the Designing Innovative Projects Test, the Stanford Task Test and a number of teacher made language measures including sound discrimination, reading comprehension and listening comprehension tests. The second, the written assignment, occurs less frequently, and is used primarily to ascertain the knowledge level of students enrolled in the program being evaluated. Evaluators may require the completion of a specific written assignment or they may request completed examples of students' written work. The third, (which occurred only once) is controlled observation. In this case, students were provided with a particular stimulus and their reactions were

observed.

■ **Other**

Reports coded here are those which make explicit a type of data collection technique which does not fit the other categories.

c. Number of Data Collection Techniques

As evaluators commonly use more than one technique for collecting data, it was also thought useful to record the number of different data collection techniques used.

C. EVALUATION – FOR WHOM AND BY WHOM?

The third cluster of questions is concerned with identifying both the recipients of the reports and the evaluators. The two pertinent questions are:

To whom was the report submitted? and
Who were the designated evaluators?

1. To Whom was the Report Submitted?

It was not always possible to identify the recipient of a particular report from the information provided therein. However, in those reports in which the identity of the recipient or recipients was made explicit, five different types of recipients were identified. As it was possible for a recipient to be someone other than those listed, a category "Other" was added. As it was also possible for the identity of the recipient not to be reported, a category "Not Stated" was included. The coding categories are listed in Table 4.12.

TABLE 4.12

**Coding Categories for Content Identifying the Recipients of
Evaluation Reports**

Trustees
Superintendent/Senior SD Administrators
Representative/Evaluation Committee
School Staff
Funding Agency
Other
Not Stated

2. Who were the Designated Evaluators?

The reports indicated that evaluators differed in four major respects. They differed, first, in where they came from, i.e., whether or not the evaluators were from within the district which sponsored the program to be evaluated. Second, they differed in their position, i.e., their job classification or their association with the school district, or both. Third, different numbers of evaluators were involved in the different studies. Finally, they differed in the advisory structure available to them; some evaluators, for example, were assisted by formalized advisory groups or committees. Although evaluators were found to differ in these four respects, the first two are complementary and for this reason are considered together. Thus, the reports indicated that it would be useful to explore three aspects of the question about designated evaluators. First, who the evaluators were; second, the number of evaluators involved in each study; and third, whether there was an advisory structure in place to assist the evaluators. Coding categories for identifying the designated evaluators are given in Table 4.13.

TABLE 4.13

Coding Categories for Content Identifying the Designated Evaluators

<i>Designated Evaluators</i>	<i>Broader Category</i>
Senior SD Administrators, SD Office Staff, Program-operating Personnel, Program-using Personnel, Uninvolved Personnel, School-based Administrators, Other, Not Stated	Stakeholders (SD Employees)
Trustees, Parents, Members of Local Organizations, Community Members (at large), Faculty (University/College), Provincial Ministry Employees, Student Program Participants, Non-program Participants (Students), Post-program Participants (Students), Other, Not Stated	Stakeholders (Non-SD Employees)
SD Staff on loan from other SDs, Faculty/Graduate Students (University/College), Private Consultants, Representatives from Provincial Ministries, Outside Experts (Program Area), Other, Not Stated	Externals

a. Designated Evaluators

Relevant details of the evaluators in terms of whether they were stakeholders (school district employees or non-district employees) or externals (from outside the district) are shown in Table 4.13. Reports in which it was not possible to identify the evaluators are coded as "Not Stated."

b. Number of Evaluators

The number of evaluators involved in an evaluation can range from one to at least ten, when, for example, evaluators may form committees or groups with specific mandates.

c. Advisory Structure

In some reports, mention was made of evaluators being assisted by advisory groups. Here, reports are coded on the basis of the kinds of support structures available to the evaluators. For example, sometimes evaluators were assisted by advisory (or steering) committees with mandates such as:

- (a) to identify and collect relevant documents;
- (b) to assist in identifying the focus of the study;
- (c) to assist in identifying questions to be answered during the course of the study;
- (d) to assist in the preparation of data collecting instruments;
- (e) to receive and comment on the draft or interim report; and
- (f) to review the final report for bias and completeness.

An analysis of the composition of such groups showed that they could be described in the same way as the designated evaluators, i.e., the members included representatives from stakeholding groups (both SD and non-SD stakeholders) and externals.

D. EVALUATION – WITH WHAT CONCLUSION?

This final general question is concerned with recommendations made by the designated evaluators. The final specific question follows.

1. What Recommendations (if any) were Made?

Of the eighty-five documents used in the final analysis, fifty-one (60%) included a section containing evaluator recommendations.¹⁷ The recommendations contained in the reports differed in terms of the kinds of actions they suggested, the areas targeted for these actions and in the number of recommendations included. For each report the number of recommendations was recorded within each of four **action** categories (Continuation, Modification, Innovation and Termination), and within each of eleven categories which indicate the **target area** of the recommendations. The coding categories are shown in Table 4.14 which displays actions and target areas as the two dimensions of an 11×4 matrix.

Decision rules for coding action and target categories respectively are given in the two following sections, and a final section draws attention to some of the difficulties that arise when recommendations are assigned to categories.

a. Assigning Recommendations to Action Categories

Recommendations for future action were found to be of four basic kinds. They could suggest continuation, modification, innovation or termination. Decision rules for coding recommendations follow.

■ Continuation

Recommendations coded in this category are those which make use of the words, or derivatives of the words 'continue' and 'maintain'.

The implication here is that the activity should be maintained "as is" or at the "present level" or that a process which has already started, e.g., of investigation, of

¹⁷ Appended to two documents were lists of recommendations produced, not by the evaluators, but by senior school district officials. These recommendations were not included in the analysis of recommendations.

TABLE 4.14

Coding Categories for the Analysis of Recommendations

<i>Target Area of Recommendations</i>	<i>Action Recommended</i>			
	Continuation	Modification	Innovation	Termination
Philosophy				
Program and Procedures				
Policy and Administrative Practices				
Personnel				
Instructional Practices				
Professional Development				
Buildings and Facilities				
Materials and Equipment				
Evaluation and Research				
Community Relations				
Other				

Note: In any given cell, the number of actions recommended in each target area was recorded on the coding sheets for each report.

review, or of improvement, should be continued. For example:

. . . that the practice of giving teachers a wide choice of reading material be continued (#030:2).

■ **Modification**

In order to be 'modified', the activity must already be occurring. Recommendations are assigned to this category if they suggest such actions as doing more of something, less of something, putting greater emphasis on, focussing on a particular aspect of, decreasing, reducing, adjusting, strengthening, expanding, clarifying, defining, changing the emphasis of, or coordinating the various parts.

For example:

. . . that placement and exit processes ensure that District policies and M.H.R. mandates are followed (#055:44).

■ **Innovation**

In order to be coded as 'innovative', the recommendation must suggest something that has not, as far as is evident from the recommendation and the information in the report, been part of the program before.

For example:

. . . that when the goal for the student in the XX program is eventual reintegration into the neighborhood school, the XX program staff establish an outreach program to facilitate this reintegration (#051:2).

■ **Termination**

Recommendations in this category suggest the cessation of a particular activity.

For example,

. . . that the practice of assigning students to counsellors by sex be eliminated (#096:11).

b. Assigning Recommendations to Target Areas

Initial examination of the reports indicated that there were ten areas in which evaluators recommended that action be taken. An eleventh area "Other" was added in the event that recommendations were found in areas other than those identified in the initial analysis. The target area categories are described below.

■ **Philosophy**

The philosophy of a program is its foundation. A stated philosophy makes explicit the values and beliefs which determine the direction of a program. Stated goals and objectives operationalize this philosophy and provide program personnel with a number of desirable and specific outcomes.

To be assigned to this category, a recommendation has to be primarily concerned with philosophy or with program goals and objectives.

Examples are as follows:

. . . that the XX department work with representatives of the elementary and secondary administrators and teachers to review the present philosophy and to develop definitive goals which are related to identified needs (#042:3).

. . . that each XX community develop a detailed set of educational goals and objectives for their community. Without clearly established goals, efforts will be directionless and insufficient (#045:68).

■ **Program and Procedures**

"Program and Procedures" has two major areas of focus. The first is the program in its entirety.

If a recommendation includes reference to the whole program, whether in terms of its continuation, its general improvement, or its long-term function within the school district, it is included in this category.

For example:

The XX programs continue to provide an effective alternative educational

format for a certain segment of the secondary school student body and *should be continued* (#005:23). (Emphasis in the original.)

The second area of focus views the program as a collection of a number of different activities, some of which are identified in recommendations. These may concern program planning and design, curriculum, or organizational strategies and procedures.

If a recommendation is directed towards program planning or curriculum design, or organizational strategies and procedures it is included here.

Examples of these recommendations include:

The addition of XX is an added feature of the program this year and one that has been very successful. It should continue to be part of the program (#031:7).

. . . that a district K-12 XX curriculum be developed and a plan to implement that curriculum be initiated (#064:20).

. . . that exit procedures should be revised to assure for post assessment and the receipt of current data at exit sites to allow for adequate planning there (#053:40).

■ Policy and Administrative Practices

This category also has two areas of focus which are conceptually different from the two areas of focus described under "Program and Procedures." The first, "Policy," includes recommendations which concern the establishment of new programs, or which concern the development or implementation of policy. The second, "Administrative Practices," is concerned with the relationship between the program and the organizational context in which the program is situated. This context may be a program's school-base or the school district office or both. The administrative practices in this area differ from those in the "Policy" section because, rather than addressing intra-program organization, they are concerned with the relationship between a program and its context.

To be included in this category, recommendations must refer to policy development or implementation (including the establishment of new

programs); or to administrative practices which provide a link between the program and its organizational context.

Examples of Policy recommendations concerning the establishment of new programs follow:

. . . that kindergarten classes be established at XX and XX. It is further recommended that the kindergarten classes be located on-reserve if possible, and that they be Native-controlled in close cooperation with the public schools (#045:72).

. . . that the district consider the possibility of a short-term pull-out program that provides intense concentrated instruction for severely disabled children (#022:9).

The following are examples of policy recommendations which include the development of policy and the processes by means of which such policies may be implemented:

. . . that the Board and appropriate staff begin the process of establishing a district policy for XX services. Through this process it is important that certain priorities be established, i.e., the amount of time spent teaching guidance versus being available for individual or small group counselling, the amount of time spent teaching regular classes, the amount of time spent on quasi-administrative services, etc. (#096:11).

. . . that the school district develop a policy and a set of procedures for dealing with accusations of racial or ethnic discrimination. Furthermore, a plan for reducing prejudice and discrimination must be developed (#045:75).

Examples of recommendations concerning administrative practices are:

. . . that the district encourage schools wishing to embark on XX programs to:

- set goals early in September
- allow teachers to “buy-in”
- decide upon the learning outcome or goal for the program
- look at establishing a school-based resource person who can act as coach and coordinator for the school’s XX program (#047:5).

. . . that elementary counselling time be increased (#096:11).

■ Personnel

If the recommendation refers to the complement of program personnel, the characteristics of these personnel or their roles, then the recommendation is assigned to this category.

Examples follow.

To provide sustained expert counsel, the district should employ a XX specialist (teaching or non-teaching) whose role would be to provide routine maintenance and to advise schools and the purchasing department (#072:12).

Individuals considered for teaching and other positions in XX schools should have the following characteristics:

- patience
- enthusiasm
- positive and realistic expectations
- physical energy
- appropriate competencies in the areas of diagnosis and remediation of learning difficulties
- awareness of the powerful influence of their role as models for appropriate behavior
- ability to relate positively and quickly with the students involved (#005:24).

. . . that the role of the XX teacher be more clearly defined and understood by those involved at all levels in the provision of service (#022:7).

■ Instructional Practices

Here, recommendations must pertain to instructional strategies for different grade and ability levels, or to assessing the progress of individual students.

For example:

Encourage, by planning with students, the development of greater independence for senior students (#084:13).

. . . that flexibility in the evaluation of exceptional students be reviewed so the reporting system has continuity in the district and so alternative assessment techniques can be developed, i.e., oral examinations, extended time limits, etc. (#022:8).

■ **Professional Development (In-service)**

Although the recommendations in this area always affect personnel, they differ from those in the "personnel" category in that they are concerned with the opportunities for staff development available to personnel, rather than with the number of staff members, or with the particular functions performed by these staff members within their particular programs.

To be assigned to this category, recommendations must refer only to professional development activities.

These include recommendations for increasing in-service opportunities for teachers and administrators; recommendations for making long-range, professional development plans; and recommendations for making use of available consulting services. For example:

. . . that In-service opportunities should be made available locally for all staff dealing with the areas of counselling and academic programming (#005:24).

. . . that the school district develop an in-service model that will consider such things as need, resources, teacher turnover, and district priorities, and that an in-service plan be prepared on a long-range basis (#042:3).

. . . that Ministry of Education XX staff be invited to the district to discuss programs and to give a provincial perspective to staff (#022:4).

■ **Buildings and Facilities**

Recommendations included here are concerned with the standards of the buildings themselves, with their locations or with the facilities inside those buildings.

For example:

. . . that consideration should be given to either attaching modulars to main buildings or supplying them with water (#082:12).

. . . that consideration should be given to the feasibility of locating XX Alternate on the grounds of XX School (#005:24).

Facilities inside buildings may pertain to the use of space or with what the space contains. Examples include:

. . . that space facilities for XX within the school should be expanded and, if possible, centralized to facilitate use of equipment and materials (#086:9).

. . . that the library will be painted before the end of the 1984-85 school year. The colours chosen will take into consideration the northern exposure and the colours of the curtains and the carpeted area (#036:10).

■ Materials and Equipment

In order to be assigned to this category, recommendations must be concerned with the acquisition, replacement or removal of materials and equipment or with the provision of funding for them (so that standards can be maintained or improved); or they must be concerned with the organization of available materials and equipment.

Examples include:

. . . that picture sets will be inventoried and obsolete material discarded (#036:3).

. . . that the material should be reorganized and catalogued. [. . . that] a list for teachers of what is available should also be prepared (#068:94).

■ Evaluation and Research

There are two major areas of focus subsumed here. The first is evaluation.

To be included as evaluation, the recommendation must concern either subsequent evaluations of the whole program or the specific monitoring devices employed on a continual basis while the program is in progress.

For example:

. . . that an internal program evaluation model be developed that reviews programs individually or collectively. [That] such criteria as testing, I.E.P.'s, student, teacher and parent surveys be utilized (#022:10).

. . . that once the direction of the XX program has been determined there be a planned external review by XX staff (#042:11).

The second focus is research.

To be included here, the recommendation must concern the investigation of particular problems or the exploration of particular questions identified in connection with a program.

For example:

. . . that there should be some exploration done to establish why math teachers feel that calculators should not be used (#067:77).

. . . that a committee of grade 10 Mathematics teachers be struck to investigate whether the high frequency of "failure" at grade ten lies with the exam or with student performance (#070:32).

■ Community Relations

For inclusion here, recommendations must concern the relationship between the program, school or school district and the community at large.

For example:

. . . that the native parents be informed of the XX through the Home/School Co-ordinator, and through newsletters from the schools and the Friendship Centre, explaining the services of the XX (#048:7).

. . . that the district and schools continue to utilize the expertise of the different parent groups (#022:9).

■ Other

Included within this category are recommendations which could not be assigned to the categories listed above.

As it was possible to assign each recommendation to an identified category, the designation "Other" was left blank in this study.

c. Difficulties in Coding Recommendations

Two types of difficulties were encountered in coding recommendations. The first was one of identification and the second, of interpretation. The first type of difficulty took three forms. First, identification problems occurred when recommendations were embedded in text and not clearly flagged by such phrases as "some suggestions for improvement are" or "it is recommended that." Careful reading of the text, however, ensured that recommendations, wherever they occurred, were noted. Second, identification difficulty arose when the identical recommendation appeared twice. In such cases, the recommendation was coded only once. Third, identification problems occurred when more than one recommendation was embedded in a single statement. For example:

. . . that we, as staff, promote organized outdoor games that include individual and dual activities. For this to occur, lines will be painted on the tarmac and tetherball poles erected (#029:8).

The promotion of organized outdoor games is one recommendation (action recommended: modification; target area: program and procedures), and the painting of lines and erecting of poles is another (action recommended: innovation; target area: buildings and facilities). In such a case, each embedded recommendation was considered separately.

The second type of difficulty was in interpreting the kind of action recommended. Sometimes it was difficult to decide whether a recommendation was suggesting that some existing aspect of the program be modified or whether a new aspect was being introduced. Here, contextual clues played an important part. For example:

. . . that XX programs should be available throughout the academic year.

. . . that some consideration be given to a suitable program of an

enrichment nature for those pupils who would benefit from same (#035:6).

In the first example, if the programs were already available throughout the year then the recommendation would be categorized as continuation; if the programs were available throughout part, but not all, of the year, then the descriptor would be modification; and, if the programs were not available at all, the action recommended would be innovation. It is only from additional information contained in the text that it was ascertained that the programs were available for part of the year and that the evaluators were recommending that they be available throughout the year; hence here the correct designation is "modification."

In the second example, if a program already existed but was considered unsuitable, then the designation would be "modification," as the recommendation would be suggesting that consideration be given to a "suitable" program. If, as was the case, the text elsewhere indicates that there is no enrichment program, then the correct designation would be "innovation."

Interpretational difficulties also occurred in assigning recommendations to target areas.

For example:

. . . that standardization and a schedule for facilities changes should be established (#072:12).

. . . that goals and objectives for the XX program should be presented to the Principals' meeting early in September of the 1985-86 school year (#047:5).

Although on initial perusal, the former appears to be about facilities and the latter about goals and objectives, they were both categorized as "policy and administrative

practices.” This is because both deal with organization and scheduling. Some recommendations are ambiguous. For instance, in the example above it is not clear to what “standardization” refers. When recommendations are ambiguous, the part of the recommendation that is understood, together with the cues available in the text, are used as the basis for categorization.

E. GENERAL INFORMATION

Information provided in this section indicates the year the report was produced and places the report in school district context. Five kinds of information were coded:

■ Document Identification Number (ID)

To facilitate identification, each report was assigned an ID. As 110 reports were received in total, each was given a three digit number from 001 to 110.

■ Year

Copies of evaluation reports produced between 1980 and 1986 were requested. Thus, the reports are coded according to the year in which a report was completed as indicated by its authors.

■ School District Number (SD No)

Each school district in British Columbia has a code number assigned by the Ministry of Education. Although there are seventy-five school districts in the province, their identification numbers range from 1 to 92. These numbers formed the basis for coding.

■ School District Size (SD Size)

A category of school district size was created when it became apparent that school district size could be a factor affecting the number of reports produced in each school district.¹⁸ The sizes of the school districts in British Columbia are described as follows:

¹⁸ The possible significance of school district size was mentioned several times during the course of follow-up telephone calls to small and very small districts.

<u>No. of Students</u>	<u>Size of School District</u>
1-1000	Very Small
1001-3000	Small
3001-8000	Medium
8001-15000	Large
15000-up	Very Large

This classification was developed from one used by R.R. Rayborn (1986). The allocation of school district to size value is based on FTE figures from September 30, 1985.

■ **Number of Documents submitted by each School District**

As the number of documents submitted by each school district varied considerably, this category was included. Reports were coded according to the total number of reports submitted by the school district in which the report in question was produced.

F. SUMMARY

This chapter has described the development of rules for the coding of program evaluation reports according to their content. These were discussed in five sections: four clusters of questions from the framework, and a final section containing general information. The next chapter presents the results of coding the contents of the eighty-five reports.

¹⁸ (cont'd) Respondents indicated that small districts lacked the personnel and resources necessary for evaluative activities.

CHAPTER V.

THE CONTENT OF THE EVALUATION REPORTS

The purpose of this chapter is to report the results of the content analysis of the program evaluation reports. It is divided into four major sections. The first contains general information about the reports. The second presents a description of the frequency of various features of the reports, discussed under the four general questions of the framework. The third reports the results of an examination of associations among various features of the reports. A summary of the content of the reports is provided in the final section. Displays of the frequency with which different elements of content were found in the data are provided either in tables in this chapter or in Appendix 4.

A. GENERAL INFORMATION

Although school district personnel had been asked to submit documents containing reports of evaluations of school district programs, a review of the titles of the reports indicated that a variety of process descriptors other than "evaluation" were commonly employed. These descriptors included assessments, reviews, reports, pilots, surveys, discussion papers and those in which the name of the program was used as the title of the report. The process descriptor "evaluation" was used in forty-six cases (54%), "report" in fourteen cases (16%), "assessment" in eleven (13%), and the name of the program in six cases (7%). The remaining descriptors each occurred in only one or two cases.

Not only did the descriptors vary, but also the physical characteristics of the reports themselves differed. They ranged in length from three to almost three hundred pages. Some took the form of memos or letters sent from one school district employee to another, while others were bound and presented as formal submissions to identified clients. Some contained photographs of student activities and others contained examples of student work. Some had no appendix while others appended a variety of pertinent information ranging from copies of student timetables or relevant correspondence, to copies of the instruments used to collect the data. The variation in the amount of information included is a topic which is taken up again in the subsection on definition.

The length of time taken to do the evaluation and produce the report also varied. One report took almost three years to complete while others took only a few days. In other cases it was not possible to ascertain the duration of the evaluation and report preparation. All of the evaluation reports, however, were dated. This made it possible to identify the year in which they had been completed. Of the eighty-five evaluation documents used as the data base for this study, thirty-one (37%) were dated between 1980 and 1983, forty-seven (55%) were produced in 1984 and 1985, and seven (8%) were completed in 1986.

The evaluation documents analyzed came from twenty-eight school districts. Table 5.1 provides a breakdown of these school districts according to size and gives the number of reports submitted by them. School district sizes are listed down the left side of the table. The number of usable (in that they were included in the data base used for the analysis of content) documents submitted by districts varied considerably. These numbers are shown across the top of the table and the columns show the number of

school districts in each size category which submitted various numbers of evaluation documents.

TABLE 5.1

**Number of School Districts in each of Five Size Categories
and Six Categories of Number of Reports Submitted**

<i>SD Size</i>	<i>No. of Reports Submitted</i>						<i>No. of SDs</i>
	1	2	3	4	15	19	
Very Small	1	—	—	—	—	—	1
Small	7	—	—	—	—	—	7
Medium	4	1	3	3	—	—	11
Large	1	—	2	—	1	—	4
Very Large	1	1	2	—	—	1	5
Total No. of Reports	14	4	21	12	15	19	28

Apart from one large and one very large school district which submitted fifteen and nineteen usable reports respectively, the maximum number of reports submitted by any other district was four. One very small district submitted a single report; seven small districts submitted one report each; eleven medium-sized districts submitted between one and four reports; one large district submitted one report, while two large districts each submitted three and one submitted fifteen reports. Four very large school districts submitted either one, two or three reports, while one submitted nineteen. It is noteworthy that the small and very small school districts which submitted reports each

made a single submission; in contrast, more than one report was submitted by fourteen of the twenty medium, large and very large districts with reports in the data base.

B. REPORTING THE CONTENT: FREQUENCIES

The following pages are arranged in four sections corresponding to the sections of the framework. These sections are: Evaluation — to what end? Evaluation — by what means? Evaluation — for whom and by whom? and Evaluation — with what conclusion? The results of the analysis for each section are presented.

1. Evaluation — To What End?

This section has four component questions. The questions address evaluation definition, evaluation intent, reason for the evaluation and the object evaluated. Each is described in turn.

a. How was Evaluation Defined?

Table 5.2 shows the definition of evaluation inferred from analyzing the evaluation reports. This definition is based on whether the designated evaluators were responsible for making judgements as to the value of the object evaluated or whether their task was to provide information on which others could base evaluative decisions. The columns show the number and percentage of documents in each group. The implied definition of evaluation in sixty-four of the eighty-five reports (75%) was judgemental. The implied definition in the remaining twenty-one (25%) was provision of information. All the reports in which the underlying definition of evaluation was judgement, supplied information in support of the evaluative judgements made. Quantity of information did, however, differ. In almost two-thirds of the sixty-four reports which were judgemental,

TABLE 5.2
Definition of Evaluation

<i>Definition</i>	<i>No. of Reports</i>
<i>Judgement</i>	
No supporting information	0 (0%)
Some supporting information	23 (27%)
Much supporting information	41 (48%)
TOTAL	64 (75%)
<i>Provision of Information</i>	
Some judgement	7 (8%)
No judgement	14 (17%)
TOTAL	21 (25%)
<i>Total Number of Reports</i>	85 (100%)

much supporting information was provided, while in the remaining one third, less supporting information was reported. In the twenty-one cases in which the provision of information was identified as the underlying definition of evaluation, two-thirds contained no evidence of judgement by the designated evaluator, while there was some evidence of judgement in the remainder.¹⁹ Of particular note is that the majority of the reports were judgemental in nature (75%), while far fewer (25%) were intended to provide information rather than to make judgements. In addition, almost half the reports examined (48%) were not only judgemental, but also provided a great deal of

¹⁹ Reports included in this group were those in which evaluators may have intentionally, or unintentionally, included value-loaded words or phrases in their descriptions, but in which the underlying intent was clearly one of information provision.

supporting information.

b. What were the Intents of the Evaluation?

Evaluation intents were examined as revealed by the purposes of the evaluations and by whether their function was predominantly formative or summative. Table 5.3 provides a breakdown of the purposes of the evaluations, while Table 5.4 shows their function.

Part A of Table 5.3 shows the types of purposes which were identified. As some reports contained more than one statement of purpose, the total number of purpose statements identified exceeded the number of reports. Part B expands on part A in that it shows the number of reports with single statements of purpose and the number of reports with multiple purpose statements. Part C shows, for the three most commonly found purposes, the frequency with which each occurred alone or in combination with the other two. Part A of Table 5.3 shows that the most common purpose identified was "Improvement and Change." This purpose was found in fifty-eight of the eighty-five reports. The second and third most commonly identified purposes were "Decision Making" and "Awareness and Knowledge." Decision making was identified in thirty-two reports, while awareness and knowledge occurred in twenty-four reports. The remaining purposes occurred in two to nine reports. Included within the "Other" descriptor was one report in which uncertainty as to the purpose of the evaluation was expressed by the evaluator; and two reports in which it was possible to identify the purpose as one of validation. In one of these reports, the purpose was to validate criteria for evaluation; in the other, the purpose was to validate the existing program.

TABLE 5.3
Purposes of Evaluation

<i>A. Types of Purpose</i>	<i>No. of Reports</i>
Improvement and Change	58
Decision Making	32
Awareness and Knowledge	24
Development and Planning	9
Accountability	5
Informing the Policy Process	4
Meeting Funding Requirements	3
Comparison	3
Other	2
Total Number of Purpose Statements is 140	
<i>B. Number of Purposes</i>	
1	41 (48%) Single
2	34
3	9 (52%) Multiple
4	1
Total Number of Reports	<hr/> 85 (100%)
<i>C. Frequency of Occurrence of Commonly Found Purposes</i>	
Improvement and Change Only	20 (23.5%)
Decision Making Only	8 (9%)
Awareness and Knowledge Only	10 (12%)
Improvement and Change and Decision Making	15 (18%)
Improvement and Change and Awareness and Knowledge	7 (8%)
Decision Making and Awareness and Knowledge	3 (3.5%)
Reports Containing Only Frequently Occurring Purposes	<hr/> 63 (74%)
Reports Including Additional Purposes	22 (26%)
Total Number of Reports	<hr/> 85 (100%)

Part B of Table 5.3 shows that slightly fewer than half the reports (48%) contained single statements of purpose, while slightly more than half of them (52%) contained multiple purpose statements. Of those reports with multiple purpose statements, it was possible to identify two purposes in thirty-four reports and three purposes in nine reports. One report contained four statements of purpose.

Part C of Table 5.3 shows the frequency of occurrence, singly and in combinations of often occurring purposes. "Improvement and Change" as a single purpose statement was found most often, while "Decision Making" and "Awareness and Knowledge" as single purposes also occurred relatively frequently, and the three in different combinations accounted for 74% of all the purposes identified.

Table 5.4 shows the numbers of documents with a formative function (either stated in the report or implied therein); the number of documents with a summative function (either stated or implied), and the number of documents with a combination of both formative and summative functions. In eight cases, it was not possible to infer whether the document was intended to have a formative or summative function. Function was specified in comparatively few cases, i.e., only twelve of the eighty-five reports (14%) stated whether they were intended to fulfil a formative or summative function or both. However, in most cases in which the formative or summative nature of the evaluation was not made explicit, it was possible to make inferences and classify the reports accordingly. A purely formative function was either stated or implied in fifteen cases, while a purely summative function was implied in seventeen other cases. In forty-five (53%) of the cases, there was evidence of the reports fulfilling both types of function. Two observations may be made. The first is that in none of the reports was a

TABLE 5.4
Formative and Summative Functions of Evaluation

<i>Function</i>	<i>No. of Reports</i>
Formative (F)	Stated 2 (2%)
	Implied 13 (15%)
	TOTAL 15 (18%)
Formative and Summative	F stated, S implied 1 (1%)
	F implied, S stated 1 (1%)
	F and S both stated 8 (9%)
	F and S both implied 35 (41%)
	TOTAL 45 (53%)
Summative (S)	Stated 0 (0%)
	Implied 17 (20%)
	TOTAL 17 (20%)
No Statement Made, No Inference Possible	8 (9%)
	TOTAL 8 (9%)
Total Number of Reports	85 (100%)

solely²⁰. summative function stated, although in seventeen cases this was the only function implied. The second is that more than half the reports fulfilled both formative and summative functions.

²⁰ with no accompanying stated or implied formative function.

c. *Why was the Evaluation Undertaken?*

Table 5.5 shows the kinds of reason given for each evaluation and the number of documents which contained them. Two types of reason were identified. The first was that evaluations were required and the second was that evaluations were requested. In some documents more than one reason was identified. For example, an evaluator may have indicated both the required nature of the evaluation and the source of the request which initiated the evaluation process. For this reason, Table 5.5 and the following text show numbers which total more than the number of reports.

Reasons for the evaluations were identified in sixty-two reports (73%) and were not identified in the remainder (27%).²¹ The reason for evaluation given in thirty-six reports was that the evaluations were required by school district policy, established practice or by program design or short-term, school district objectives such as those found in annual school district goal statements. The reason given in thirty reports was that the evaluations were carried out in response to particular requests. Seven reports gave reasons relating to particular sets of circumstances. These circumstances encompass those reasons listed as "other." They include changes in pupil enrolment numbers, length of elapsed time since the previous evaluation, changes in personnel, and changes in response to feedback from stakeholders. In one case, this feedback took the form of concerns expressed in the community and, in the other, feedback was obtained from a questionnaire which had been circulated in the previous year.

²¹ The absence of a reason in these twenty-three reports may indicate that their authors simply chose not to comment on the background of the evaluations with regard to the reason for their initiation.

TABLE 5.5
Reason for Evaluation

<i>Reason</i>	<i>No. of Reports</i>
<i>Required by SD Policy or Established Practice</i>	14
<i>Required by Program Design or Short-term SD Objective</i>	22
TOTAL	36
<i>Requested by</i>	
Program Sponsors (6)	
School Staffs (2)	
Trustees (8)	30
Senior SD Administrators (11)	
Representative or Evaluation Committee (3)	
<i>Other</i>	7
<i>No Statement Made, No Inference Possible</i>	23
<i>No. of Reports Containing One Reason</i>	51 (60%)
<i>No. of Reports Containing More than One Reason</i>	11 (13%)
<i>No. of Reports with No Reason Stated or Implied</i>	23 (27%)
Total Number of Reports	85 (100%)

d. What was the Object Evaluated?

The data bearing on this question are summarized in four tables, showing the type of object evaluated; the permanence of the object evaluated; the aspect of the object evaluated; and the object sponsor, the object base, and the grade level pertaining to these objects. Table 5.6 shows the three divisions of objects evaluated. These are "Program or Program-related Practice," "Organizational Unit," and "Facilities." Within

TABLE 5.6
Object of the Evaluation

<i>Object</i>	<i>No. of Reports</i>
<i>Program or Program-related Practice</i>	
Curriculum and Instruction	44 (52%)
Special Services	33 (39%)
Organization for Instruction	2 (2%)
TOTAL	79 (93%)
<i>Organizational Unit</i>	4 (5%)
<i>Facilities</i>	2 (2%)
<i>Total Number of Reports</i>	85 (100%)

“Program or Program-related Practice,” reports concerned with the evaluation of instructional programs were of three kinds: “Curriculum and Instruction,” “Special Services,” and “Organization for Instruction.” The types of program in each category were listed in Chapter IV (Table 4.4) and are also listed in the coding instrument (Appendix 3).

Seventy-nine (93%) of the reports had as their objects programs or program-related practices. Forty-four (52%) formed part of the regular school program, while thirty-three (39%) were classified as “Special Services” and two (2%) were classified as “Organization for Instruction.” Of interest is the very low number of reports submitted concerning core content courses (except when test results form the focus) in areas such as Mathematics, English and Social Studies.

The objects of the remaining reports were the whole organizational unit, i.e., the school or centre as the unit of analysis; and school facilities. Four reports (5%) were classified as the former and two (2%) as the latter.

TABLE 5.7
Permanence of Object Evaluated

<i>Permanence</i>	<i>No. of Reports</i>
<i>Continuing</i>	67 (79%)
<i>Temporary</i>	
Pilot (9)	
Project (3)	
Display (1)	
Equipment (1)	
Other (3)	
	17 (20%)
<i>No Inference Possible</i>	1 (1%)
<i>Total Number of Reports</i>	85 (100%)

Table 5.7 shows the permanence of the programs evaluated. Sixty-seven (79%), were intended as continuing programs (although it may be assumed that an unsatisfactory evaluation could result in termination of some of them). Seventeen (20%) were temporary and included a variety of programs implemented in the short-term. Pilot studies were the most common of these.

Table 5.8 shows whether the object was evaluated in its entirety, i.e., with emphasis

TABLE 5.8
Aspect of Object Evaluated

<i>Aspect</i>	<i>No. of Reports</i>
<i>General</i>	73 (86%)
<i>Specific</i>	
Impact (5)	
Service Delivery (1)	
Needs Assessment (1)	
Long-term Planning (1)	
Student Climate (1)	
Consistency (1)	
Reading (1)	
Test Data (1)	
	12 (14%)
<i>Total Number of Reports</i>	85 (100%)

on a multiplicity of aspects, or whether one specific aspect formed the focus. General aspects were considered in seventy-three cases (86%), while one specific aspect was examined in twelve cases (14%). The most frequently evaluated of these specific aspects was "Impact," which occurred in five cases.

Table 5.9 shows the three organizational aspects of the object: "Object Sponsor," "Object Base," and "Grade Level." In 84% of the reports the program sponsor was the school district and in 58% of the reports the object base was the school or centre. In terms of grade level, the breakdown falls approximately into thirds. Twenty-nine percent of the reports described evaluations done of programs intended for grades Kindergarten to Grade 7. (This group includes two reports of programs which also

TABLE 5.9
Organizational Aspects of Object Evaluated

<i>Object Sponsor</i>	<i>No. of Reports</i>
School or Centre	8 (9%)
School District	71 (84%)
Joint Sponsorship	5 (6%)
Non-district Agency	1 (1%)
TOTAL	85 (100%)
<i>Object Base</i>	
School or Centre	49 (58%)
School District	28 (33%)
School and School District	5 (6%)
Non-district Facilities	3 (4%)
TOTAL	85 (100%)
<i>Grade Level</i>	
Primary	2 (2%)
Intermediate	4 (5%)
Elementary	17 (20%)
Elementary/Junior Secondary	2 (2%)
Junior Secondary	8 (9%)
Senior Secondary	6 (7%)
Secondary	16 (19%)
Elementary/Secondary (District)	27 (32%)
Adult	1 (1%)
Not Stated	2 (2%)
TOTAL	85 (100%)

involved students from the junior-secondary age range.) Thirty-five percent of the reports described evaluations of programs intended for secondary school students and 32% considered the whole elementary-secondary range. As the evaluations were distributed across grade levels, it appears that no programs pertaining to any particular grade level were evaluated more than any other.

2. Evaluation – By What Means?

This section has three component questions. They address the kinds of information collected about the objects, the criteria which are used to judge these objects and the methods of inquiry used by the evaluators.

a. What Kinds of Information Regarding Each Object were Reported?

The following five tables illustrate the kinds of information collected and reported in the evaluation documents. Tables 5.10 and 5.11 show the sources of information used by evaluators, while Tables 5.12, 5.13 and 5.14 illustrate the nature of this information.

The information reportedly collected about each object evaluated came from one, two or three major sources: from respondents, from written materials or from the evaluator him or herself. Table 5.10 lists the specific sources within each major source and gives the number of reports in which information was obtained from these sources. The first major source is "People Other than Evaluator," the respondents. The categories are, "SD Employees," "Non-SD Employees," and "Externals." Within the SD employees, program operating personnel were mentioned as having provided information in seventy of the eighty-five reports (82%).²² School-based administrators provided information in

²² It might be expected that program operating personnel would be required to provide information in every case; however, this was not possible to ascertain in

TABLE 5.10

Sources of Information Reported about Each Object*

		<i>Respondents</i>				<i>Written Materials</i>		<i>Evaluator</i>	
SD Employees	No. of Reports	Non-SD Employees	No. of Reports	Externals	No. of Reports		No. of Reports		No. of Reports
Program-operating Personnel	70 (82%)	Student Program Participants	57 (67%)	District Staff from other SDs	5 (6%)	Documents and Records	58 (68%)	Observations of Program in Action	25 (29%)
School-based Administrators	45 (53%)	Parents	42 (49%)	Representatives from Provincial Ministries	5 (6%)				
Program-using Personnel	30 (35%)	Post-program Participants	12 (14%)	Outside Experts in Program Area	1 (1%)	Review of Literature	10 (12%)	Evaluator Assigned Tasks	75 (88%)
District Office Staff	20 (23%)	Students Not Enrolled in Program	12 (14%)	Research/Evaluation Consultants	0 (0%)				
Senior SD Administrators	16 (19%)	Provincial Ministry Employees	10 (12%)	University/College Faculty	0 (0%)	Not Stated	2 (2%)	Other	1 (1%)
Uninvolved Personnel	11 (13%)	Members of Local Organizations	8 (9%)	Other	1 (1%)			Not Stated	2 (2%)
Other	2 (2%)	Community Members	8 (9%)	Not Stated	2 (2%)				
Not Stated	2 (2%)	Trustees	4 (5%)						
		University/College Faculty	1 (1%)						
		Not Stated	4 (5%)						

*Because there were multiple sources in the program evaluation reports, the total number of reports is >85.

forty-five cases (53%). The designation "Other" was used twice. In these cases it was possible to determine that the information was provided by school-based teams whose membership included SD employees, but it was not possible to determine the particular positions held by these employees.

Of the non-SD employees, student program participants provided information in fifty-seven cases (67%), while parents provided information in forty-two cases (49%). By comparison, very few externals were asked to provide information. Those who were asked were primarily district staff from other districts (five cases), and representatives from Provincial Ministries (also five cases).

Within the written source materials, documents and records were reported to have been accessed in fifty-eight cases (68%) and reviews of the literature in ten cases (12%). The table also shows that the evaluator was an information source in the majority of cases. In seventy-five of these (88%), the evaluator requested that some action be taken specifically to inform the evaluative process (e.g., that meetings would take place or records be submitted), while in twenty-five cases (29%) the evaluator made observations of program activities. One case received the designation "Other" because the evaluators themselves formed a planning group and were responsible for consolidating the information each had available.

Table 5.11 provides more detail about the sources used for the provision of information. Part A focusses on the respondents and shows the number of reports in which

²² (cont'd) all cases. It may be that some authors did not make the sources of information explicit or it may be that contributions from program operating personnel were not pertinent to the questions asked or to the purposes fulfilled.

TABLE 5.11

Sources Used for the Provision of Information about Each Object

<i>A. Respondents</i>	<i>No. of Reports</i>
SD Employees	12 (14%)
Non-SD Employees	7 (8%)
Externals	0 (0%)
SD and Non-SD Employees	52 (61%)
SD Employees, Non-SD Employees and Externals	5 (6%)
SD Employees and Externals	4 (5%)
Non-SD Employees and Externals	0 (0%)
No Information from Respondents	5 (6%)
TOTAL	85 (100%)
<i>B. Respondents, Written Materials and Evaluator</i>	
<i>Respondents</i>	80 (94%)
SD Employees	73
Non-SD Employees	64
Externals	9
<i>Written Materials</i>	58 (68%)
<i>Evaluator</i>	78 (92%)
<i>C. Sources used Singly or in Combination</i>	
Evaluator Only	0 (0%)
Respondents Only	1 (1%)
Writings Only	2 (2%)
Writings and Evaluator	1 (1%)
Respondents and Writings	2 (2%)
Respondents and Evaluator	24 (28%)
Respondents, Writings and Evaluator	53 (62%)
Not Stated	2 (2%)
TOTAL	85 (100%)

information was obtained from single groups of respondents, or from these groups in combination. Part B shows the number of reports in which respondents were accessed (eighty), in which written materials were accessed (fifty-eight), and which used the evaluator as a source of information (seventy-eight). Part C provides a breakdown of the number of reports in which information was collected from particular single sources or from these sources in combination. Part A indicates that, in those cases in which respondents provided information, stakeholders formed the particular source group polled most often, i.e., a combination of SD and non-SD employees. Part B shows that, whereas the combination of SD employees, non-SD employees and externals was polled in eighty cases (94%), documents and records and reviews of the literature were used to provide information in fifty-eight cases (68%) and the evaluator was classified as an information source in seventy-eight cases (92%). More specifically, among the three groups of respondents, the group polled most often was that consisting of SD employees. This group was polled in seventy-three cases (86%). Non-SD employees were involved in sixty-four cases (75%) and externals in only nine (11%). Part C indicates the number of evaluations which were based on information obtained from a single source (4%) or a combination of sources (94%). In fifty-three cases (62%), information was obtained from respondents, writings and the evaluator, while in twenty-four cases (28%), respondents and the evaluator were responsible for information provision.

From an examination of Table 5.10 and Table 5.11, it may be concluded that, although the evaluator was classified as a major information source (in that actions initiated by the evaluator facilitated the provision of information), stakeholders (SD and non-SD employees) in single or combined groups were primarily responsible for supplying the

desired information. More particularly, the key subgroups among the SD employees were program operating personnel and school-based administrators; and the key subgroups among the non-SD employees were the students participating in the programs and their parents. In addition, use was made of written source materials in more than half of the documents examined.

Tables 5.12 and 5.13 show the nature of the information collected and reported regarding each object. Table 5.12 provides a broad breakdown of the nature of the information collected and reported, while Table 5.13 indicates the kinds of general and specific (opinion and descriptive) information collected.

TABLE 5.12
Nature of Information Reported about Each Object

<i>Nature</i>	<i>No. of Reports</i>
<i>General</i>	14 (16%)
<i>Specific</i>	71 (84%)
Opinion and Description	55
Description Only	9
Opinion Only	7
<i>Total Number of Reports</i>	85 (100%)

Table 5.12 shows that in the majority of cases (84%), specific rather than general information was provided. This specific information was of two types: opinion and

description. In most cases, fifty-five, both opinion and descriptive information were reported. However, nine documents contained purely descriptive information, and seven documents contained only opinion information.

Table 5.13 provides a breakdown of the nature of the general and of the specific information which was collected and reported. The table shows that both opinion data and information, and descriptive data and information were collected. Both kinds of information were collected about program process, program outcomes, program participants and about similar objects in other sites, and "Other" foci. The table shows that thirteen documents contained exclusively general information, while the information contained in one was predominantly general. Table 5.13 also shows that the remaining documents, containing opinion information, descriptive information or a combination of both, were further subdivided depending on the focus of that information. Fifty-six reports provided opinion data and information on program process, fifty-three on outcomes and three on participants. Fifty-two reports provided descriptive data on process, fifty-three on outcomes and twenty-five on participants.

Within the opinion information, the designation "Other" was given to two reports. In one, stakeholder opinion on areas not addressed by the programming was given prominence; and in the other, stakeholder opinion of the facilities was sought. Within the descriptive information, the designation "Other" was given to six reports because they provided descriptions of objects other than the object of the evaluation or other than descriptions of similar objects in other sites. For example, information was provided on provincial test scores, general district enrollment data, population trends, ideal program conditions and reasons for enrolling students in particular programs.

TABLE 5.13

Types of General and Specific Information Reported about Each Object

<i>General</i>	No. of Reports	<i>Specific</i>			
		Opinion Information	No. of Reports	Descriptive Information	No. of Reports
Exclusively General	13 (15%)	Process	56 (66%)	Process	52 (61%)
Predominantly General	1 (1%)	Outcomes	53 (62%)	Outcomes	53 (62%)
		Participants	3 (4%)	Participants	25 (29%)
		Similar Objects in Other Sites	0 (0%)	Similar Objects in Other Sites	6 (7%)
		Other	2 (2%)	Other	6 (7%)

It is notable that process and outcomes received approximately equal attention. There was, however, slightly more emphasis on opinions about process rather than descriptions of process; and on descriptions of participants and similar objects in other sites rather than on opinions about them.

TABLE 5.14

Areas of Focus of Reports Containing Both Opinion and Descriptive Information Reported about Each Object

<i>Focus of Information</i>	<i>No. of Reports</i>
Process and Outcomes Only	19 (22%)
Process and Outcomes with Description of Participants	15 (18%)
Process and Outcomes with Information on Participants, Similar Objects or Other	6 (7%)
TOTAL	40 (47%)
Reports with Other Areas of Descriptive and Opinion Focus	15 (18%)
<i>Reports Containing Both Opinion and Description</i>	55 (65%)
<i>Reports Not Containing Both Opinion and Description</i>	30 (35%)
<i>Total Number of Reports</i>	85 (100%)

Table 5.14 has been included to show the focus of information gathering when both opinion and descriptive information are obtained. (It is possible to focus on process, outcomes, participants, similar objects and other, or on any of these in combination.) The table provides a breakdown of the reports which contained both opinion and descriptive information. Of the fifty-five reports which contained opinion and descriptive

information, forty contained information on process and outcomes and, in some cases, on other foci as well.

b. What Criteria were Used to Judge the Merit and Worth of the Object?

In identifying the criteria used to judge the merit and worth of the objects evaluated, those reports which contained no judgements could not be used. This analysis, therefore, was based on the sixty-four reports in which the definition of evaluation was judgement rather than the provision of information.

However, before describing the findings relating to criteria, it is important to note that in only seven cases were these criteria made explicit. In the majority of cases (fifty-seven of the sixty-four reports or 89%), criteria were not made explicit and were therefore difficult to identify. This difficulty was resolved by noting what the evaluator chose to focus on and identifying these areas as indicative of implicit criteria.²³ Indications of evaluator focus occurred in a variety of locations throughout the reports. Criteria were evident in checklists, in goals and objectives, in introductory and summary paragraphs, in recommendations and in interview protocols and survey instruments. It is because the criteria were implicit and references to them were scattered throughout the documents, that location has not been included in Table 5.15.

The first column of Table 5.15 shows that the prime source of criteria was the evaluator. Criteria identified by the evaluator were evident in forty-nine of the

²³ Areas of focus were identified also in the twenty-one reports produced by authors whose task was to provide information on which others could base judgements. Areas of focus and numbers of reports within which these areas could be identified were proportionally similar when compared with the reports in which the evaluators made judgements.

TABLE 5.15

Source and Nature of Criteria Used to Judge the Merit and Worth of the Object Evaluated (No. of Reports: 64)

<i>Source</i>		<i>Nature</i>			
	No. of Reports	Process	No. of Reports	Outcomes	No. of Reports
Identified by Evaluator	49 (77%)	<i>External</i>			
Program Goals and Objectives	22 (34%)	Adherence to Guidelines	8 (13%)	Stakeholder Satisfaction	51 (80%)
		<i>Internal</i>			
		• Program and Procedures	59 (92%)	Student Behaviour and Attitudes	32 (50%)
		• Policy and Administrative Practices	55 (86%)		
		• Instructional Practices	43 (67%)		
		• Materials and Equipment	33 (52%)		
Written Ministry or SD guidelines	14 (22%)	• Personnel	31 (48%)	Student Achievement	31 (48%)
		• Facilities	30 (47%)		
Terms of Reference for the Evaluation	12 (19%)	• Professional Development	29 (45%)	Indications of a Change of State	22 (34%)
		• Community Relations	29 (45%)		
Alternate Objects	4 (6%)	• Philosophy	23 (36%)		
		• Evaluation and Research	16 (25%)		
Other	5 (8%)	• Other	15 (23%)	Other	4 (6%)

sixty-four cases (77%). Criteria were also obtained from the goals and objectives of the program (34%), and from written program, school district or Ministry guidelines (22%). Criteria were identified in the terms of reference for the evaluation whenever terms of reference were drawn up by both client and evaluator and presented in the document. However, this occurred in twelve cases only (19%). Alternate objects of a similar type were sources of criteria in four cases (6%). The designation "Other" was used in five cases in which test scores (at school, district or province levels) or previous evaluation or planning documents relating to the program were explicitly used as a source of information for comparative purposes.

The remaining two-thirds of the table shows the nature of the criteria. Process criteria and outcome criteria were identified. Process criteria were of two kinds: external and internal. In only eight cases (13%) were criteria identified as external. These focussed on the extent to which the program was implemented in accordance with specified, written guidelines at school district or at Ministry level. The remaining process criteria were internal and focussed on a variety of program components combined in various ways in order to provide a picture of, and make comments on, a number of different aspects. Internal program procedures and the program in school and district context were important areas of focus (92% and 86% respectively). The remaining areas of focus are given in decreasing order in Table 5.15. Process criteria designated as "Other" include involvement in extra-curricular activities not included as part of the program itself; program atmosphere; and economic considerations in terms of the budget for the program and the redistribution of these resources (there was, however, no mention of the cost-effectiveness of a program as an area of focus or as an outcome criterion).

Listed in the third column are outcome criteria. The outcome criterion used most frequently (in 80% of cases) was stakeholder satisfaction. In most cases (see respondents as source of information in Table 5.10 and opinion data as nature of information in Table 5.13) a number of SD employees (primarily including program operating personnel and school-based administrators) and non-SD employees (primarily including student program participants and their parents) were polled. Student behaviour and attitudes and student achievement were used as outcome criteria in approximately half the evaluations (50% and 48% respectively), while an indication of a change of state was an outcome criterion in 34% of cases. The "Other" designation includes the impact of the program on school personnel themselves and the effect of the program on student access to further study and to employment opportunities.

c. What Methods of Inquiry were Used in the Evaluation?

Table 5.16 shows the approaches identified by the evaluators. Of interest here is that in the majority of cases (61%) no particular approach was made explicit. Of those reports which made method explicit, the method most frequently mentioned was the on-site visit (twelve cases). In only two cases was mention made of approaches from the literature. These were "situational interpretation" (Werner, 1979), and the "judgement matrix" (Stake, 1967). The methods included as "Other" are the external review and the use of an "interpretation panel" for the interpretation of provincial test scores prior to making comparisons between these and district scores. Tables 5.17 and 5.18 show the types and number of data collection techniques used in the evaluations. Not only were a variety of techniques incorporated into the studies, but also, in most cases, evaluators used more than one technique for gathering data. Table 5.17 shows that documents and records, questionnaires and interviews and meetings were used in

TABLE 5.16

Methods of Inquiry Made Explicit by Evaluators

<i>Method</i>	<i>No. of Reports</i>
<i>Stated</i>	33 (39%)
On-site Visit (12)	
Survey (9)	
Description (8)	
Experimental/Quasi-experimental (6)	
Situational Interpretation (1)	
Judgement Matrix (1)	
Other (2)	
<i>Not Stated</i>	52 (61%)
<i>Total No. of Reports</i>	85 (100%)

TABLE 5.17

Data Collection Techniques Used in the Evaluations

<i>Technique</i>	<i>No. of Reports</i>
Documents and Records	56 (66%)
Questionnaires	55 (65%)
Interviews and Meetings	55 (65%)
On-site Observation	27 (32%)
Checklists, Rating Scales, Attitude Scales and Inventories	19 (22%)
Achievement Measures	18 (21%)
Off-site Visits and Inquiries	4 (5%)
Other	3 (4%)

TABLE 5.18

Number of Data Collection Techniques Used in the Evaluations

<i>Number of Techniques</i>	<i>No. of Reports</i>
1	19 (22.5%)
2	17 (20%)
3	25 (29.5%)
4	13 (15%)
5	7 (8%)
6	4 (5%)
<i>Total Number of Reports</i>	85 (100%)

the majority of cases (66%, 65%, and 65% respectively). On-site observation, behaviour and attitude scales and achievement measures were used less frequently (32%, 22%, and 21% respectively), while in very few cases (5%) were inquiries made to destinations outside the school district. Reports with data collection techniques identified as "Other" included in-depth case studies of participants and observations of particular behaviours under controlled conditions. Table 5.18 shows the number of data collection techniques used by the evaluators. The number of techniques employed ranged from one to six. Three techniques were used in twenty-five cases, while more than three were used in twenty-four cases and fewer than three were used in thirty-six cases.

3. Evaluation – For Whom and By Whom?

This section has two component questions. Data collected in answer to the question “Evaluation – for whom?” are contained in Table 5.19 which identifies the recipients of the evaluation reports. Data collected in answer to the question “Evaluation – by whom?” are presented in a series of tables concerning the source and number of evaluators as well as the advisory groups convened in some cases.

a. To Whom was the Report Submitted?

It was possible to identify the recipients in only forty cases (47%). These are shown in Table 5.19.

TABLE 5.19
Recipients of Evaluation Reports

<i>Recipient</i>	<i>No. of Reports</i>
<i>Stated</i>	40 (47%)
Superintendent or Senior SD Administrator (17)	
Board of School Trustees (15)	
School Staff (4)	
Funding Agency (3)	
Representative or Evaluation Committee (3)	
Other (3)	
<i>Not Stated</i>	45 (53%)
<i>Total Number of Reports</i>	85 (100%)

The most commonly cited recipients were the Superintendent or other senior school

district administrator or the Board of School Trustees. (The figures in parenthesis in the table add up to more than forty because some reports were submitted to more than one recipient.)

b. Who were the Designated Evaluators?

There are four tables in this series. Table 5.20 shows the positions held by the evaluators and whether they are SD employees, non-SD employees, or externals. Table 5.21 shows the number of evaluations done by members of these three groups, either singly or in combination. Table 5.22 provides information on the numbers of evaluators involved in each study. In Table 5.23, the composition of those groups functioning in an advisory capacity to the evaluators is illustrated. In eleven cases no source group or number of evaluators was identified and, as a consequence, Tables 5.20, 5.21 and 5.22 provide a breakdown for seventy-four, rather than eighty-five, documents.

Table 5.20 shows that senior district administrators, within the district itself, took on an evaluative function in twenty-six cases (35%), while other district office staff were involved in twenty-four cases (32%). In four of these cases, it was not possible to determine the positions of the evaluators, even though their status as SD-employees was evident. Non-SD employees (i.e., those with a stake in the program) featured comparatively infrequently as evaluators. Trustees were involved in an evaluative capacity in two cases, while parents, college or university faculty and provincial Ministry employees were each involved in a single evaluation. District staff on loan from other districts featured as external evaluators in ten cases (13.5%), whereas faculty or graduate students from colleges or universities were involved in seven cases

TABLE 5.20

Position-based Source of Designated Evaluators

<i>Evaluators</i>	<i>No. of Reports*</i>
<i>SD Employees</i>	
Senior SD Administrators	26 (35%)
District Office Staff	24 (32%)
Program-operating Personnel	13 (18%)
School-based Administrators	11 (15%)
Program-using Personnel	4 (5.5%)
Uninvolved Personnel	0 (0%)
SD Employees (position unclear)	4 (5.5%)
<i>Non-SD Employees</i>	
Trustees	2 (3%)
Parents	1 (1.5%)
Faculty and Graduate Students from Colleges and Universities	1 (1.5%)
Provincial Ministry Employees	1 (1.5%)
Members of Local Organizations	0 (0%)
Community Members at Large	0 (0%)
Other	1 (1.5%)
<i>Externals</i>	
District Staff on Loan from Other School Districts	10 (13.5%)
Faculty and Graduate Students from Colleges and Universities	7 (9.5%)
Private Research and Evaluation Consultants	6 (8%)
Representatives from Provincial Ministries	1 (1.5%)
Outside Experts in Program Area	0 (0%)

*In eleven cases it was not possible to discern the position or source of the designated evaluators. Therefore, only seventy-four reports are included in this table.

(9.5%). Private research or evaluation consultants were external evaluators in six cases (8%) and in one case only was there a representative from a provincial ministry.

TABLE 5.21

Source Groups of Designated Evaluators

<i>A. Source of Evaluators</i>	<i>No. of Reports*</i>
SD Employees Only	48 (65%)
Externals Only	17 (23%)
Non-SD Employees Only	0 (0%)
SD Employees and Externals	4 (5.5%)
SD Employees and Non-SD Employees	3 (4%)
Non-SD Employees and Externals	0 (0%)
SD Employees, Non-SD Employees and Externals	2 (3%)
<i>B. Involvement of Evaluators from Major Source Groups</i>	
SD Employees	57 (77%)
Externals	23 (31%)
Non-SD Employees	5 (7%)

*In eleven cases it was not possible to discern the position or source of the designated evaluators. Therefore, only seventy-four reports are included in this table.

Table 5.21 shows the number of evaluations carried out by particular groups of evaluators classified according to their source, i.e., whether they are SD employees, non-SD employees or externals. Part A shows that forty-eight (65%) of the evaluations were done by SD employees only, while seventeen (23%) evaluations were completed by externals only. The remaining nine evaluations (12%) were carried out by evaluators from combinations of source groups. In more general terms, Part B shows that SD

employees were involved in fifty-seven evaluations (77%). External evaluators were involved in twenty-three cases (31%), while non-SD employees were involved in five (7%).

TABLE 5.22
Number of Designated Evaluators

<i>Number of Evaluators</i>	<i>No. of Reports</i>
<i>Stated</i>	
1	34 (46%)
2	15 (20%)
3	5 (7%)
4	9 (12%)
5	3 (4%)
6	4 (5.5%)
7	1 (1.5%)
8	1 (1.5%)
9	1 (1.5%)
12	1 (1.5%)
TOTAL	74 (100%)
<i>Stated</i>	74 (87%)
<i>Not Stated</i>	11 (13%)
<i>Total Number of Reports</i>	85 (100%)

Table 5.22 provides a breakdown of the number of evaluators involved in each of the seventy-four reports included in this part of the analysis. Single evaluators were involved in thirty-four evaluations. In fifteen evaluations, the services of two evaluators were used. Teams of varying sizes were responsible for evaluating the

remaining programs. In general terms, single evaluators or two person teams were responsible for two-thirds (66%) of the evaluations, while teams of three or more were involved in 29% of the evaluations examined.

Ten evaluations (12%) were completed with the aid of an advisory group. Table 5.23 provides a breakdown of the number of members of each group and of the composition of these groups. Numbers of members ranged from six to ten while the groups were composed of representatives of the SD employees, the non-SD employees and the externals. Senior school district administrators, district office staff and school-based administrators formed part of seven of the ten groups, while program-operating personnel were members of six of the groups. Among the non-SD employees, provincial Ministry employees were represented on two advisory groups, while one parent and one student also functioned in an advisory capacity. Among the externals were representatives from provincial ministries who served in three of the ten advisory groups, while district staff from other districts served in two groups and one faculty member and one research and evaluation consultant also served.

TABLE 5.23

Number of Members and Composition of Advisory Groups

<i>Number of Members</i>	<i>No. of Groups*</i>
6	3
7	1
8	1
9	1
10	1
Not Stated	3
<i>Composition of Advisory Groups</i>	
<i>SD Employees</i>	
Senior SD Administrators	7
District Office Staff	7
School-based Administrators	7
Program Operating Personnel	6
Uninvolved Personnel	1
Program Using Personnel	0
Not Stated	2
<i>Non-SD Employees</i>	
Provincial Ministry Employees	2
Parents	1
Students	1
Trustees	0
Members of Local Organizations	0
Community Members at Large	0
Faculty Members and Graduate Students	0
Other	1
Not Stated	2
<i>Externals</i>	
Representatives from Provincial Ministries	3
District Staff on Loan from Other School Districts	2
Faculty Members and Graduate Students	1
Research or Evaluation Consultants	1
Not Stated	0

*Ten evaluations (12%) were completed with the aid of an advisory group.

4. Evaluation – With What Conclusion?

This section corresponds to the fourth and final section of the framework. The section had one question which concerned recommendations.

a. *What Recommendations (if any) were Made?*

Recommendations were included as part of fifty-one of the eighty-five reports or 60% of the total number of reports examined. From these fifty-one reports a total of 738 recommendations were identified. The number of recommendations emerging from each evaluation ranged from one to sixty-four.

TABLE 5.24

Actions Recommended

<i>Action Recommended</i>	<i>No. of Recommendations</i>
Continuation	49 (7%)
Modification	546 (74%)
Innovation	142 (19%)
Termination	1 (0.13%)
TOTAL	738 (100%)

Table 5.24 shows the recommendations categorized in terms of their intended action as identified in the original coding scheme. Once the data had been examined, a decision was made to modify this scheme in two ways. These modifications are illustrated in

Table 5.25 and Table 5.26.

Table 5.24 provides a breakdown of the numbers of recommendations in each of the categories of recommended action. Notable here is that only one (0.13%) of the 738 recommendations involves termination, in this case, the termination of the practice of assigning counsellors to students according to gender. Also of interest is the finding that a comparatively small percentage (7%) of the recommendations suggest continuation of an existing activity. The majority of recommendations (74%) suggest modification of existing activities, while 142 recommendations (19%) suggest innovations. Making clear distinctions, however, between recommendations for modification and those for innovation was difficult because varying amounts of supporting information were provided. Consequently, the creation of these two categories may be spurious and the decision to collapse them was made. Thus, the categories of modification and innovation were combined and labelled "Change."

Table 5.25 shows, for each of the three actions recommended, the areas at which the recommendations were targeted. Of note here is that 93% of recommendations suggest change, while 7% suggest continuation. In terms of target area, the two categories containing more recommendations than any other were program and procedures (20%) and policy and administrative practices (26%). The rest of the categories each accounted for between 3% and 11% of the remaining target areas. The target area categories of "Program and Procedures" and of "Policy and Administrative Practices" both have an administrative focus. The former is concerned with the smooth running of the program itself and the latter is concerned with the administration of the program in its school and district context. Because administrative procedures feature highly in

TABLE 5.25

Combined Action Recommended and Target Area of Recommendations

<i>Action Recommended</i>	<i>Target Area</i>											<i>Total</i>
	Philosophy	Program & Procedures	Policy & Administrative Practices	Personnel Practices	Instructional Practices	Professional Development	Facilities	Material & Equipment	Evaluation & Research	Community Relations	Other	
Continuation	1	16	12	2	3	5	—	2	6	2	—	49 (7%)
Change	21	132	183	43	71	55	46	76	38	23	—	688 (93%)
Termination	—	1	—	—	—	—	—	—	—	—	—	1 (0.13%)
Totals	22	149	195	45	74	60	46	78	44	25	—	738

TABLE 5.26

Combined Action Recommended and Combined Target Area of Recommendations

<i>Action Recommended</i>	<i>Target Area</i>										<i>Total</i>
	Philosophy	Admini- stration of Program in Context	Personnel Practices	Instructional Practices	Professional Develop- ment	Facilities	Material & Equipment	Evaluation & Research	Community Relations	Other	
Continuation	1	28	2	3	5	—	2	6	2	—	49 (7%)
Change	21	315	43	71	55	46	76	38	23	—	688 (93%)
Termination	—	1	—	—	—	—	—	—	—	—	1 (0.13%)
Totals	22 (3%)	344 (47%)	45 (6%)	74 (10%)	60 (8%)	46 (6%)	78 (11%)	44 (6%)	25 (3%)	—	738

both, the decision to collapse the two categories was made. The new category became "Administration of Program in Context." Table 5.26 provides a breakdown of the three actions recommended and the modified list of target areas. It is noteworthy that the newly formed category of "Administration of Program in Context" accounts for almost half (47%) of all the recommendations. In terms of the intention of the recommendations in the new category, 315 of the 344 of them (92%) are recommendations for change. One (0.13%) recommends termination and twenty-eight (8%) recommend continuation. Clearly, administrative activities feature highly in the suggestions for the improvement of programs, while there is comparatively little attention paid to the other target areas identified.

C. REPORTING THE CONTENT: CROSS-TABULATIONS

The analysis described to this point has been in terms of frequency distributions of twenty-five different features of the content of the reports. It was also possible to examine (by means of cross-tabulations) whether or not there were associations among any of these features. To run a cross-tabulation of each feature with all the other features would necessitate three hundred cross-tabulations. Not all features, however, lend themselves to analysis. In some cases, the frequencies showed one category to be so dominant that there seemed little point in further exploration.²⁴ In other cases, the presence of small cell sizes in some categories would make further analysis relatively meaningless. In still other cases, insufficient information was given in the reports.²⁵

²⁴ For example, in the list of additional characteristics of the objects evaluated, 79% were continuing programs, 86% were general in that they did not focus on specific program aspects, and 84% of the programs evaluated were sponsored by the school district. In "Nature of Information," 84% of reports fell in the category of specific information, and most of these were found to contain both opinion and descriptive information.

²⁵ Two categories were excluded from the analysis for this reason. These were "Approach Ascribed by Evaluator(s)," and "Recipient of Evaluation Report." In the

Finally, sometimes it appeared relevant to look for associations in relation to particular features, but not all. For example, the number of documents submitted by a school district seemed usefully cross-tabulated with the size of the school district, but not with other variables.

If the number of possible cross-tabulations is reduced for the reasons given above, ninety-one remain to be explored. When these ninety-one were performed, three kinds of result emerged. First, were those which showed no obvious pattern of association. Second, there were those which showed clearly discernible but not unexpected patterns. Third were those which showed clearly discernible patterns which might not have been expected. Most results fell into the first of these categories. The following paragraphs discuss those in the second and third.

1. Expected Associations

The discussion in Chapter III of the validity of the procedures used in content analysis noted that because the coding of "latent" content was more subjective than that of "manifest" content, it was likely to contribute more to the validity of the analysis. One confirmation of such validity among categories of content, given their logic, is the extent to which one finds associations that might be expected. Thus, for example, given that reports may have as their purpose, "decision making" and "awareness and knowledge," one would expect more of the former to have a judgemental definition, and more of the latter to have a provision of information definition.

²⁵ (cont'd) former, no specification of approach was given in fifty-two reports (61%), and in the latter, it was not possible to infer from forty-five (53%) of the reports the identity of their recipients.

Such expected associations were found between “Definition of Evaluation” and “Purpose”, “Definition of Evaluation” and “Recommendations”, “Definition of Evaluation” and “Function”, “Purpose” and “Function”, and “Purpose” and “Recommendations.” The presence of these expected associations also served to validate the content analysis; they are illustrated in Tables 5.27 – 5.31 below.

Table 5.27 shows the breakdown of purpose statements containing the three most commonly identified purposes (“Improvement and Change,” “Decision Making” and “Awareness and Knowledge”) according to the definition of evaluation of the report in which these purpose statements were found. The expected association between purpose and definition is clearly seen.

TABLE 5.27

Definition of Evaluation and Number of Purpose Statements in which Improvement and Change, Decision Making, and Awareness and Knowledge were identified as Purposes

<i>Definition</i>	<i>Purpose Statements</i>			<i>TOTAL</i>
	Improvement & Change	Decision Making	Awareness & Knowledge	
Judgement	53	27	9	89
Provision of Information	5	5	15	25
TOTALS	58	32	24	114

A disproportionate number of the reports whose definition was "Provision of Information" had as their purpose "Awareness and Knowledge." Similarly, the majority whose definition was judgement had as their purpose "Improvement and Change." These findings were not unexpected. Authors of reports that provide information (on which others base judgements) are not responsible for making judgements themselves or for identifying areas for improvement. Rather, they make the client group more knowledgeable. When evaluators are requested to make judgements, it is likely that actions based on these judgements will lead to improvement and change.²⁶

Table 5.28 shows the number of reports with and without recommendations coded according to definition of evaluation. The table illustrates that evaluations in which the evaluators made judgements contained recommendations while those which were intended to provide information did not contain recommendations.²⁷ Clearly this finding was not unexpected as recommendations for action are usually suggested once judgements have been made. Table 5.29 shows the breakdown of evaluation functions according to definition of evaluation. Among those reports with provision of information as their definition, there were disproportionately fewer evaluations with identified functions, while in the reports with a judgemental definition, there was a disproportionately greater number of reports with formative and summative functions. This finding was not unexpected. For the most part, it is the person (or persons) given

²⁶ Purpose statements are not mutually exclusive. It is possible for any one report to contain purpose statements intended to promote awareness and knowledge as well as to result in improvement and change.

²⁷ In one case (#027) a recommendation was embedded in the text although the mandate of the evaluation report was the provision of information and not judgement.

TABLE 5.28**Definition of Evaluation and Recommendations**

<i>Definition</i>	<i>Recommendations</i>		<i>TOTAL</i>
	Present	Absent	
Judgement	50	14	64
Provision of Information	1	20	21
TOTALS	51	34	85

TABLE 5.29**Definition and Function of Evaluation**

<i>Definition</i>	<i>Function</i>				<i>TOTAL</i>
	Formative	Formative and Summative	Summative	Not Stated or Implied	
Judgement	12	41	10	1	64
Provision of Information	3	4	7	7	21
TOTALS	15	45	17	8	85

the responsibility of judging the program who decides whether the information provided is to be given a formative or summative function. Sometimes this can be discerned from the purpose statements but in those cases where the intents of the reports are not clear and where the authors of the reports are not responsible for making judgements,

it is not always possible to identify the formative or summative nature of the reports.

TABLE 5.30
Purpose and Function of Evaluation

<i>Purpose</i>	<i>Function</i>			<i>TOTAL</i>
	Formative	Formative and Summative	Summative	
Improvement and Change	15	43	0	58
Decision Making	1	18	13	32
Awareness and Knowledge	4	9	6	19*
TOTALS	20	70	19	109

*Although twenty-four "Awareness and Knowledge" purpose statements were identified, it was not possible to discern the function of evaluation in five of these cases.

Table 5.30 displays type of evaluation function with the three most common evaluation purposes found in the reports. Not unexpected here is the disproportionate number of reports with a summative function only and "improvement and change" and "decision making" purposes. No report with an exclusively summative function contained a purpose statement intended to lead to improvement and change, whereas there were thirteen reports with an exclusively summative function which had decision making as a purpose. Clearly this association was not unexpected.

Table 5.31 shows the association between the three most common evaluation purposes and the presence or absence of recommendations. This association was also expected. Of the reports with "Improvement and Change" as a purpose, it is interesting to note that nine did not contain any recommendations.

TABLE 5.31
Purpose of Evaluation and Recommendations

<i>Purpose</i>	<i>Recommendations</i>		<i>TOTAL</i>
	Yes	No	
Improvement and Change	49	9	58
Decision Making	14	18	32
Awareness and Knowledge	10	14	24
TOTALS	73	41	114

2. Unexpected Associations

Unexpected associations were found between "Type of Object" and "Function" and, when evaluation was judgement rather than provision of information, between "Type of Object" and "Designated Evaluators." An unexpected association was also found between "Designated Evaluators" and "Size of School District." These associations are illustrated in Tables 5.32–5.34 below. Table 5.32 shows an unexpected association between "Type of Object" and "Evaluation Function." When the evaluations were classified as having a function which was summative only, there were

TABLE 5.32

Type of Object Evaluated and Function of Evaluation

<i>Function</i>	<i>Type of Object</i>			<i>TOTAL</i>
	Curriculum & Instruction	Special Services	Other	
Formative	8	7	0	15
Formative and Summative	18	23	4	45
Summative	14	2	1	17
TOTALS	40	32	5	77*

*It was not possible to identify the function of the evaluation in eight of the eighty-five reports.

disproportionately more objects classified within "Curriculum and Instruction" than there were within "Special Services." However, when the objects were classified within "Special Services" there were disproportionately more reports with a "Formative and Summative" function than there were in the "Curriculum and Instruction" category. It is possible that evaluators are aware that, because of the financial and emotional costs involved in setting up new special service programs, a purely summative evaluation could result in the loss of much needed services. Thus, school district personnel might be reluctant to evaluate special services exclusively summatively. Table 5.33 displays "Type of Object" and "Designated Evaluators" in those reports where evaluators were responsible for making judgements. Reports in which both stakeholders and externals were designated evaluators were excluded. Thus, two groups of evaluators, stakeholders and externals, were used in this analysis. The table shows that there

TABLE 5.33

**Type of Object Evaluated and Designated Evaluators when the
Definition of Evaluation is Judgement**

<i>Type of Object</i>	<i>Designated Evaluators</i>		<i>TOTAL*</i>
	Stakeholders	Externals	
Curriculum & Instruction	22	4	26
Special Services	12	9	21
Other	3	0	3
TOTALS	37	13	50

*Those reports in which stakeholders and externals both served as designated evaluators have been excluded.

were disproportionately more external evaluators evaluating programs in Special Service areas than there were in other areas. This might be accounted for in that people with a specialized knowledge are required to determine the merit of Special Services. The final unexpected association was found between designated evaluators and size of school district. The findings are displayed in Table 5.34. In small school districts the evaluations were carried out by external personnel, while in large school districts, fewer externals were involved than might have been expected. This could be explained by the fact that many small school districts do not have the resources for departments of evaluation and research, and thus tend to request outside assistance when program evaluations are required.

TABLE 5.34**Designated Evaluators and School District (SD) Size**

<i>SD Size</i>	<i>Designated Evaluators</i>		<i>TOTAL*</i>
	<i>Stakeholders</i>	<i>Externals</i>	
Small	0	5	5
Medium	13	7	20
Large	38	5	43
TOTALS	51	17	68

*Those reports in which both stakeholders and externals served as designated evaluators have been excluded.

D. SUMMARY OF THE CONTENT OF THE REPORTS

This section contains a point form summary of the results of the content analysis presented in this chapter. These points are arranged under the general headings used in the body of the chapter. A point form summary is useful because it serves to display the array of findings in a succinct form so as to facilitate the drawing of conclusions (Miles and Huberman, 1984).

General Information

1. The formal evaluation reports submitted varied considerably in form and in aspects addressed.
2. Small and very small school districts submitted fewer formal written evaluation reports than medium, large and very large districts.

Evaluation -- To What End?

3. a. Two major kinds of evaluation definition were evident, i.e., evaluator as judge and evaluator as information broker.
b. Although there was variation in the amount of supporting information provided, there was no evidence of unsubstantiated judgement.
4. a. Reports were intended to fulfil one or more purposes.
b. The three most common purposes were improvement and change, decision making, and awareness and knowledge.
5. a. The formative nature or summative nature of the evaluations was not usually made explicit.
b. In those cases in which the formative or summative nature of an evaluation was stated, the formative nature of the evaluation was made explicit more frequently than was the summative.
c. Approximately half the evaluations were both formative and summative.
d. The majority of reports included a formative focus.
6. Two chief reasons for the occurrence of the evaluations emerged. They took place as a matter of course (i.e., evaluations were required by school district policy, established practice, the design of the program itself or a school district short-term objective), or programs were evaluated in response to an in-house request.
7. a. The majority of reports submitted were for instructional programs which form part of the curriculum available to students enrolled in the regular program; or which form part of the specialized programming available to members of specific target groups.

- b. Programs forming part of the core curriculum were the object of comparatively few reports.
8. Most of the programs evaluated formed part of the regular school program, i.e., they were designed to be continuing rather than temporary.
9. More than one aspect of a program was examined by the evaluators in the majority of cases.
10.
 - a. The majority of programs evaluated were sponsored by the school district.
 - b. Most programs evaluated were school-based.
 - c. Programs at all grade levels were evaluated; no one level appeared to be targeted more than any other.
11.
 - a. Those evaluations in which evaluators provided information and not judgement were more likely to have "Awareness and Knowledge" as a purpose, and to have fewer identified formative and summative functions; and were less likely to contain recommendations, than were those in which the evaluators made judgements.
 - b. Those evaluations in which the evaluators made judgements were more likely to have "Improvement and Change," or "Decision Making" as a purpose, and to have more identified formative and summative functions, and to contain recommendations, than were those in which the evaluators provided information only.
 - c. Objects within "Curriculum and Instruction" were more likely to be evaluated summatively, than were those classified in "Special Services"; while those in "Special Services" were more likely to have both a formative and a summative designation.

Evaluation – By What Means?

12. a. Although the evaluator was classified as a major information source in that actions initiated by the evaluator facilitated the provision of information, stakeholders in single or combined groups were primarily responsible for supplying the desired information.
- b. The key subgroups among the SD employees were program operating personnel and school-based administrators.
- c. The key subgroups among the non-SD employees were the students participating in the programs and their parents.
- d. Use was made of written source materials in more than half of the documents examined.
- e. In most cases, information was obtained from more than one source.
13. The majority of documents contained opinion and descriptive information on both program processes and program outcomes.
14. a. The evaluator was primarily responsible for the identification of criteria for evaluation.
- b. Process criteria focussed primarily on the administrative workings of the program itself and on the administrative procedures linking the program with its administrative school and district context.
- c. Stakeholder satisfaction was the key outcome criterion used in the majority of evaluations.
- d. No report made explicit mention of cost effectiveness as an outcome criterion or as a specific area of focus.
15. Evaluators rarely referred explicitly to particular methods of inquiry reported in the evaluation literature.

16. Evaluators incorporated a variety of techniques into their studies and also tended to use more than one technique for gathering data.

Evaluation – For Whom and By Whom?

17. Evaluation reports were submitted most frequently to senior school district administrators or to Boards of School Trustees.
18.
 - a. School district employees were most frequently involved as evaluators.
 - b. District staff from other districts were frequently involved as external evaluators.
 - c. Objects classified within “Special Services” were more likely to be evaluated by an external evaluator than were those in “Curriculum and Instruction.”
 - d. Programs in small districts were more likely to be evaluated by external evaluators, than were programs in large school districts.
 - e. Programs in large districts were more likely to be evaluated by internal evaluators than external evaluators.
 - f. Most evaluations were carried out by one or two evaluators.
 - g. Evaluation teams (of three or more) often contained a number of people representing different stakeholding groups.
 - h. In some cases, an advisory committee composed of representatives from stakeholding groups was struck.

Evaluation – With What Conclusion?

19. a. Not all of the evaluation reports resulted in recommendations.
- b. Almost half of the recommendations were concerned with the administration of the program in context.
- c. The intention of a large majority of the recommendations was modification; comparatively few recommended the continuation of activities and recommendations for termination were almost non-existent.

CHAPTER VI.

SALIENT ASPECTS OF PROGRAM EVALUATION IN BRITISH COLUMBIA SCHOOL DISTRICTS

The detailed reporting of findings in Chapter V has supplied a picture of evaluation as practiced, or at least as reported, in British Columbia which has not before been available. By focussing on each detailed level of analysis, however, the description has not to this point allowed any overall picture to emerge. The present chapter attempts to provide the missing overview of evaluation as reportedly practiced in school districts in the province.

The approach taken is to draw on the quantitative results described in Chapter V, but to do so in a way which focusses on the more striking findings of that analysis. Initially, the results found for each of the four basic questions were reviewed for the degree to which they could be clustered on the basis of similarities, and the degree to which particular features could be subsumed under a descriptor at a higher level of abstraction.²⁸ The results were also reviewed for the degree to which they seemed noteworthy or even unexpected in terms of the recent literature. From this consideration, four aspects of evaluation emerged as salient in the present data.

The remainder of this chapter is in three sections. First, the identification of the salient aspects is described. Second, each of the four aspects is discussed; and third, in

²⁸ "Clustering" and "Subsuming Particulars into the General" are specific tactics for drawing meaning from a particular configuration of data (Miles and Huberman, 1984).

the concluding section, a generalization about program evaluation practices in British Columbia is proposed.

A. THE IDENTIFICATION OF SALIENT FINDINGS

Of the separately listed findings at the end of Chapter V, twelve were considered worthy of further consideration in "building a logical chain of evidence" (Miles and Huberman, 1984:229) for drawing conclusions. Four of these findings arise in connection with the first of the general questions (Evaluation — to what end?); three arise in connection with the second; three from the third, and two arise from the fourth.

In response to the question, Evaluation — to what end?, four findings seem to stand out. First, evaluators could take a judgemental role, as they did in 75% of the reports, or a role as information broker, as they did in 25% of cases. Second, "Awareness and Knowledge" was found with "Improvement and Change" and "Decision Making" as one of the three most common purposes of the evaluations examined. Third, over 70% of the reports took, either implicitly or explicitly, a formative focus. No reports were explicitly and exclusively summative in nature. Fourth, evaluations regularly took place as a result of school district policy or accepted school district practices, or in response to an in-house request.

In response to the question, Evaluation — by what means?, three findings were particularly notable. First, most of the information about the programs was supplied by within-district stakeholders, particularly program-operating personnel, school-based administrators, students and their parents. Second, process criteria focussed on the

administrative workings of program and on the administrative links between the program and the school and school district. Third, the most commonly used outcome criterion was stakeholder satisfaction.

In response to the question, Evaluation — for whom and by whom?, three findings emerged. First, it was noteworthy that evaluation reports were submitted most frequently to senior school district administrators or to school trustees. Second, school district employees were involved as evaluators in majority of evaluations (77%). Third, advisory groups consisting of members of stakeholding groups were struck in several cases.

The two notable findings identified in response to the fourth general question, Evaluation — with what conclusion?, were first, that the intention of the vast majority of recommendations was the modification of aspects of the program; and second, that the target area for almost half of the total number of recommendations was the administration of the program rather than the program itself.

To consider these twelve findings together is to see that they are classifiable in more ways than simply by the four questions which gave rise to them. In one way or another, the set of findings may be seen to illuminate four different aspects of evaluation in school districts. These are (1) the participation of the stakeholders in the evaluation process; (2) the role of the evaluators themselves; (3) the purposes of evaluation, and (4) the identification of criteria. These aspects are discussed in the following section.

B. FOUR SALIENT ASPECTS OF THE FINDINGS OF THE STUDY

Although the four aspects identified above emerged from the present data, it is also true that they feature as important in the literature. For this reason, the discussion of each which follows begins with a brief consideration of its treatment in the literature.

1. Stakeholder Participation

Stakeholder²⁹ participation in the conduct of evaluations has been debated in the literature of the 1980s (Bryk, 1983; House, 1986; Greene, 1988). Emphasis on the participation of stakeholders in the actual practice of evaluation can be traced back to the late 1970s when the National Institute of Education (NIE) developed the stakeholder approach to evaluation. This approach challenged the traditional authority relationship between evaluators and stakeholders as stakeholders became directly involved in making decisions about how the evaluations were to be conducted. Stakeholder evaluation was based on the assumption that stakeholders who had been instrumental in the evaluation process would be committed to the evaluation outcomes and thus, more likely to use the results of these evaluations.

The approach was implemented in the evaluations of two high-profile programs in the United States, both activated in politically charged environments. One program, entitled "Cities-in-Schools," was supported by the Carter administration, and the other, "PUSH/Excel," was advocated by Jesse Jackson. Case studies of the two programs

²⁹ Key stakeholder subgroups were program operating personnel, school-based administrators, students participating in the program and their parents. Trustees or SD administrators are not identified as stakeholders in these discussions even though they are a special subgroup of stakeholders who are key to the evaluation process. In this study they have been referred to as clients, (a term used frequently in the literature) to distinguish them from other stakeholders. Evaluators are responsible to clients and it is to them that the final report is submitted.

indicated that both remained controversial after the evaluations and that in neither case was the approach implemented as intended (Farrar and House, 1983; Bryk and Raudenbush, 1983). These examples notwithstanding, Greene (1988), provides evidence that stakeholder participation can be a viable approach to evaluation when stakeholders are genuine collaborators in that they, as well as the evaluators, are responsible for determining the direction of the evaluation.

In the present study, only ten cases were identified in which advisory groups of stakeholders were involved in determining the direction of the evaluations (i.e., stakeholder participation in the sense introduced by the NIE). Of these ten advisory groups, the group affiliations or positions of the members were made explicit in seven. The seven groups all contained senior school district administrators, district office staff and school-based administrators. Parents and students were represented in only one of these groups. Thus, although it may be said that stakeholder evaluation, in the sense introduced by the NIE, took place in those ten evaluations carried out in conjunction with representative groups of stakeholders, the stakeholder groups most represented were those consisting of school district personnel.

The findings of the present study show that stakeholders are of vital importance in that they provide evaluators with information about the objects of the evaluations. Stakeholders (especially non-SD employees) are less important, however, in the actual design and implementation of the evaluation. Thus, although stakeholders did feature prominently in the evaluations, the evidence is clear that their involvement (other than as designated evaluators or clients) was limited to the provision of information, primarily by means of responding to predetermined questions such as those found in

questionnaires or in interview protocols. In very few cases were they instrumental in determining the direction of the evaluation. The implication here is that the evaluators retain control over the majority of school district evaluations, and that it is the concerns and preferences of the evaluator and clients that determine the direction of the evaluation rather than the concerns or preferences of the stakeholders.

2. The Role of the Evaluator

Most of the evaluation literature assumes that evaluations are conducted by external evaluators working on a contract basis (House, 1986). They may be evaluators by profession or may come from universities or other school districts. They are perceived as experts either in evaluation or in the content area of the program being evaluated. Internal evaluators may be attached to a school district evaluation or research office or they may become involved in evaluation as part of their regular activities. Although the evaluation reports examined in the present study were produced by both external and internal evaluators, about two-thirds of them (65%) were conducted by internal evaluators only, specifically school district employees.³⁰ External evaluators alone were responsible for 23% of the evaluations, while combinations of internal and external evaluators were responsible for 7% of the evaluations. The observation that so many evaluations were carried out either by school district employees alone or in conjunction with others (77%), provides support for House's (1986) contention that evaluators are increasingly from inside the organizations sponsoring the programs to be evaluated.

The role of the evaluator implicit in much of the evaluation literature is one of

³⁰ Internal evaluators provide a cost-saving alternative to contracting out evaluations to external evaluators, especially in times of restraint (a pertinent condition in B.C. in the early eighties).

judgement. This is because evaluation, broadly conceived, is reflected in the literature as a process designed to determine the merit and worth of an object (House, 1980; Joint Committee, 1981; Guba and Lincoln, 1981). There is, however, a contrasting view in which evaluation is conceived as a non-judgemental process whose success is determined by what others learn (Cronbach *et al* 1980, Cronbach 1982). This distinction between non-judgemental and judgemental evaluation has clear implications for the role of the evaluator.

The roles of internal evaluators are influenced by the organization (Kennedy, 1983) in that these evaluators are subject to the administrative structures, organizational norms and authority relationships within the school district organization. In addition, with regard to the evaluator as judge and the evaluator as information broker, the potential for conflict of interest and resulting bias may be greater when the evaluator is employed by the organization sponsoring the object of the evaluation. For example, it might be difficult and potentially precarious in terms of an internal evaluator's survival within the organization for him or her to act as judge and question underlying assumptions or challenge accepted practices. If, however, the evaluator is charged with providing useful, but relatively neutral data (in that there is no evidence of any value judgements having been made),³¹ he or she is in a much less vulnerable position and conflict of interest is much less likely than when the evaluator takes a judgemental role.

In the present study, most of the evaluations were conducted by internal evaluators

³¹ Feldman and March (1981) suggest that making value judgements appears to be inconsistent with the kind of requirements that organizations have, i.e., for data which serve to reduce uncertainty about the context within which the organization operates.

and some, a sizeable minority of 25%, contained no evidence of judgement. In addition, the reports indicated that in numerous school districts information was collected on a regular basis and evaluations were carried out as a result of school district policy or accepted practice. Moreover, the recipients of the reports were generally senior school district officials or school trustees. Even in those cases in which identification of the recipient was not possible,³² it can be assumed that the reports were intended for internal use, i.e., the ultimate destination of the reports was clear to the evaluators, and presumably also to the recipients, and as such it was not necessary to indicate this on the document. Thus, these evaluations were conducted by in-house evaluators for in-house consumption only.

In the analysis of the recommendations it became evident that the recommendations made were primarily concerned with making changes to aspects of existing programs. There was only one recommendation which suggested terminating anything, in this case a particular program procedure. There appeared to be an emphasis on the administrative mechanics of keeping the programs functioning. Of those recommendations which concerned the fate of the whole program, all recommended continuation. Thus, the changes suggested were, for the most part, incremental in nature. Incremental changes are usually developmental, intended for continuous improvement and are small enough so that corrections can be made without risk of major failure (Stufflebeam, 1971).

³² This was the case in some of the reports produced by internal evaluators, and in a few cases in which it was not possible to identify the evaluators. It was, however, always possible to identify the report recipient in those reports produced by identified external evaluators.

3. Purposes of Evaluation

Stufflebeam and Webster suggest that although evaluation is intended to "assist some audience to judge and improve the worth of some educational object" (1980:6), many studies done in the name of educational evaluation have very different purposes which do not directly involve the value of the program or its improvement. One such purpose, "awareness and knowledge," was found frequently in the evaluation reports examined. Evaluations with awareness and knowledge as a purpose are intended to provide descriptive information about programs in order to raise the awareness of the stakeholders and increase their level of knowledge and understanding of those programs. Nevo (1986) places "increasing awareness" among those purposes to which he ascribes the descriptor "socio-political." He suggests that socio-political purposes include the motivation of personnel and the promotion of public relations and suggests that none of these socio-political purposes has received much attention in the literature.

The formative and summative nature of the evaluations were rarely made explicit, although when it was possible to infer a formative or summative function, it was found that just over half the reports (53%) fulfilled both functions. Thus, even when value judgements about the programs were made by evaluators about how well the programs had done (summative), the evaluators still either alluded to how the programs could be improved or suggested specific courses of action directed to this end (formative). In none of the reports in which the evaluator was judge did the evaluator give an overall rating which was anything other than positive. The point here is that it may be possible to infer that not only do the evaluations record the positive nature of the programs, but also by doing so appear to provide some degree of reassurance to school district constituents, i.e., those who pay taxes to support the school system, thus

inspiring in those who have dealings with the district, confidence that school district programs are indeed meeting the needs of students.

4. The Identification of Criteria

In the early days of program evaluation in education, it was usual to collect information on the outcomes of a program and to use only this for evaluative purposes. More recently, the collection of information on which judgement was to be based has been extended to include information on a wide variety of program aspects. This broadening of scope is reflected in the evaluation literature (Stake, 1967; Stufflebeam *et al*, 1971; Guba and Lincoln, 1981). The reports examined in this study varied in the amount of information collected. In the main, however, a variety of sources were accessed for each evaluation; both opinion and descriptive information were obtained, and both program processes and program outcomes were focussed upon. Overall, the reports are in accordance with the prescriptions in the literature for the collection of a variety of kinds of information depending on the particular intents of the evaluations undertaken.

Choice of criteria for the judgement of merit and worth depends to a considerable extent on the object of an evaluation and the purpose of the evaluation; thus, this question is not generally discussed in the literature in specific terms. However, the literature does discuss a number of sources of these criteria. One such source is the professional group who set and support particular and standards (Eisner, 1979a; Guba and Lincoln, 1981). Some criteria are developed through comparison with programs of a similar type in other settings (House, 1980). At a broader level, the demands and the needs of stakeholders provide criteria for evaluation (Joint Committee, 1981), and

at a broader level still are the norms and values accepted by the society at large (House, 1980).

In 89% of the sixty-four reports in which the evaluators made judgements about the programs, criteria for this judgement were not made explicit. It is possible that those writing the reports assumed a shared understanding of what is of value in the educative process, and thus found it unnecessary to make their criteria explicit. Thus, evaluation in school districts may be built on the premise that people who work in the field and have a broad experiential base, can be trusted to make correct judgements about educational activities in school districts. As they are trusted to interpret appropriately the information collected, they are not explicitly required to justify their actions or the reasons for them.

A second point with regard to the sixty-four reports in which the role of the evaluator was judge rather than information broker is that process criteria were the ones predominantly inferred, with the major emphasis being placed on program and procedures and on policy and administrative practices, rather than on more fundamental questions concerning the justifiability of the program itself, the instructional procedures or the outcomes. What is surprising is that the major outcome criterion was stakeholder satisfaction, while substantive outcome criteria such as student achievement or behaviours were employed much less frequently. In cases in which evaluation is perceived as determining value, then one might assume that basic questions should be addressed before focussing on procedural details. For example, relevant questions might include: Does the program achieve its intents? Are the intents themselves justifiable? Is the program needs-based? Do the students learn what they

should be learning? Is what they are learning worth learning? Is the program effective? Is the cost-benefit ratio acceptable? Is the learning experience a quality one? Although these kinds of questions are addressed in some reports, the more common focus on process and stakeholder satisfaction prevailed in the majority.

C. CONCLUSION: ON NOT ROCKING THE BOAT

The previous section has shown that when the findings reported in Chapter V are considered in a general sense, rather than at the level of detail reported in that chapter, certain aspects of the practice of program evaluation emerge which might otherwise escape notice. When considered individually, these four aspects do not necessarily seem to imply anything in particular about the state of program evaluation in school districts. When considered together, however, as various facets of a whole picture, they suggest that program evaluation might serve to avoid rocking the boat (i.e., to maintain and reinforce, rather than to challenge, the status quo).

The discussion of the first aspect pointed out that senior school district administrators, trustees and evaluators maintain control over the direction that the evaluation takes and that the stakeholders are given a relatively passive role. Although their opinions about a program are sought, rarely do stakeholders have control over the direction the evaluation takes. Thus, there appears to be a basis for the suggestion that stakeholders are accorded a role in evaluation which makes their input less likely to challenge the status quo than to maintain it.

The second aspect was concerned with the role of the evaluator. The discussion focussed on the role of the internal evaluator who is charged with the provision of

information or who is responsible for judgement. The point was made that internal evaluators, relative to external evaluators, are subject to organizational norms and as such may find that their roles are circumscribed to the extent that they are not in a position to challenge existing organizational norms and values. This is particularly true for those evaluators who are responsible only for providing information.

The discussion of the third aspect drew attention to the evaluation purpose which was intended to raise awareness and increase knowledge. People who are aware and who have knowledge and understand a program may be more accepting of that program. Providing information to school district employees and to the community is good public relations. Collecting and ascribing value to the opinions of these stakeholders is often even better. In these cases, not only is the school district indicating that the opinions of stakeholders are valued, but also there is evidence that these opinions matter in that they are included in the report. When judgement is provided it is usually positive, and when suggestions for change are made, they suggest incremental change and focus mainly on program processes and administrative procedures, i.e., on making what exists function more effectively. Evaluation which serves to fulfil socio-political purposes such as raising awareness may serve to perpetuate existing norms, as may those evaluations which suggest program modifications and focus mainly on making program processes and administrative procedures more effective.

The discussion of the fourth aspect, concerned as it was with the identification of criteria, suggested that the source of these criteria was a group of educators. If it is correct that those selected to conduct evaluations tend to reinforce the values of those requesting the evaluations (in that most of the evaluators were educators, and the

majority of these were from the sponsoring school district) then major shifts in the status quo would seem unlikely.

It might be suggested that written evaluation reports, destined in some cases for wide, if not public, circulation, are unlikely to be written in such a way as to flag major problems or disturbances. This suggestion has some appeal, particularly during the stressful restraint period in British Columbia (1982-1986), and it may be that this explanation can account in part for the conservatism found here. It is unlikely, however, to be the whole explanation (after all, many reports are destined to go no further than the Board of School Trustees). Program evaluation in school districts as seen in this study has the effect of maintaining and reinforcing the school district status quo. Although no single finding leads to this conclusion, several findings taken together suggest strongly that program evaluation practice in British Columbia school districts is not an activity which is used to bring about fundamental change.

CHAPTER VII.

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

The practice of evaluating educational programs appears to be widespread, although relatively little has been written about the process as it takes place in school districts. What kinds of programs are evaluated? What are the purposes of these evaluations? Who conducts them? What methods are used? What are the findings? The present study has attempted to contribute to an understanding of program evaluation in practice. This final chapter has three main sections. The first provides a summary of the thesis. The second presents conclusions concerning the usefulness of the framework and coding categories for the study of the practice of program evaluation in school districts, and concerning the picture of program evaluation in British Columbia which emerges from the content analysis of the reports examined. The third section contains implications of the study for evaluation practice and evaluation research.

A. SUMMARY

This study was an investigation of the practice of educational program evaluation in British Columbia school districts. The three stages of the study were, developing a framework for the analysis of evaluation practices; applying this framework to school district practices in British Columbia so as to provide a description of program evaluation in the province; and using this knowledge of program evaluation practice as a means of adding to the existing evaluation knowledge base. A review of the pertinent literature led to the conclusion that it would be more useful to focus on the major issues addressed in the literature than on particular approaches to evaluation.

Issue-based questions proposed by Nevo (1983), recast so as to illuminate four basic descriptive questions, served as the framework for the identification of the coding categories used in the analysis of program evaluation in practice. All British Columbia school districts were canvassed for program evaluation reports. Eighty-five program evaluation reports from twenty-eight school districts were analyzed and coded according to the scheme, and the results tabulated. General findings about the evaluations follow. These findings are based on the questions asked at the start of the analysis.

Evaluation – to what end?

1. How was evaluation defined?

Two major definitions were evident: evaluator as judge and evaluator as information broker.

2. What were the intents of the evaluations?

The most common purposes were improvement, decision-making and awareness and knowledge. Approximately half the reports fulfilled both formative and summative functions.

3. Why were the evaluations undertaken?

Evaluations were usually undertaken in response to a particular request or as a school district requirement.

4. What were the objects evaluated?

A wide variety of instructional programs at all grade levels were evaluated. These programs usually formed part of the regular school curriculum; they were generally sponsored by the school district and were based at school sites.

Evaluation – by what means?

5. What kinds of information were collected regarding the objects?

In the main, information on program processes and program outcomes was obtained from a variety of sources. Most often, opinion and descriptive information was collected from stakeholders and written materials. Program personnel, school-based administrators, students and their parents were frequently polled.

6. What criteria were used to judge the merit and worth of the objects?

The key outcome criterion was stakeholder satisfaction; process criteria primarily concerned the administration of the program.

7. What methods of inquiry were used in the evaluations?

Evaluators used a variety of social science methods and data collection techniques in their evaluations such as on-site visits, surveys and interviews.

Evaluation – for whom and by whom?

8. **To whom were the evaluation reports submitted?**
Reports were usually submitted to senior school district administrators or to school trustees.
9. **Who were the designated evaluators?**
In most cases, the evaluators were school district personnel.

Evaluation – with what conclusion?

10. **What recommendations (if any) were made?**
Of the reports, 40% did not include recommendations. In those cases where evaluators did make recommendations, almost half were concerned with suggesting changes to the administration of the program.

Program evaluation as practiced in British Columbia school districts was then considered in general terms and the case made that program evaluation in these districts had the effect of maintaining the status quo.

B. CONCLUSIONS

The conclusions of the study result from the findings associated with the first two stages (i.e., the development of a framework for the analysis of evaluation practice; and the application of this framework to school district practices in British Columbia). The implications, which are given in the final section of the chapter, deal with the third stage of the purpose (i.e., the provision of information in order to add to the existing evaluation knowledge base). The conclusions are general and concern the usefulness of the framework and coding categories for analyzing evaluation reports and about the picture of evaluation as practiced in British Columbia which emerges from the analysis.

1. Usefulness of the Framework and Coding Categories

The framework of four basic questions and ten sub-questions which was developed for the purpose of describing program evaluation in school districts was found to be useful. The questions provided ways to organize the content of the reports. At the broadest level of conceptualization, the four major questions (Evaluation — to what end? Evaluation — by what means? Evaluation — for whom and by whom? and Evaluation — with what conclusion?) painted an overall picture of present evaluation practice in school districts. The ten sub-questions focussed the attention of the researcher on more specific things to watch for. As a result, categories for coding were identified. For example, it became clear that the question "What were the intents of the evaluation?" included both "functions" and "purposes," with the former identifying intents in terms of the formative and summative distinction made in the literature, and the latter identifying intents in terms of the two levels of purpose statement discussed in Chapter IV. The first level of purpose statement referred to what the evaluator was doing, and the second level referred to why the evaluator was doing it, for example, ". . . in order to improve the program, or to inform the decision-making process." Thus, the basic categories for coding were based in the literature and were developed and amplified through the examination of practice as reported in the documents.

The coding categories which were based on the framework and derived inductively from the content of the reports examined, were found, for the most part, to be useful. They accommodated the various kinds of information provided in the documents, and were particularly germane to the school district context, in that they included references to variables such as grade level or curriculum content area.

Moreover, the inductive nature of the task of identifying categories for coding and decision rules for assigning content to these categories, revealed interesting aspects which would not necessarily have become apparent if the categories for coding had been preconceived. For example, as shown in Chapter IV, the term "summative" is not always used in the sense of making a judgement about the value of an object, sometimes evaluators use the word in the sense of "summary," i.e., with no implicit value component.

However, some of the categories were problematic. For example, it was occasionally necessary to create a category "Other" because certain content defied categorization according to the existing categories. Some of the categories included were not found to be as useful as had originally been anticipated. For example, given the difficulty of identifying criteria for judgement used by the evaluators, a category was created concerning the location in the reports in which criteria might be identified. This included locations such as "Introductory Statements" and "Summary." Because the form and style of the reports were so variable, in some cases there was no identifiable introduction or summary. In other cases, criteria for judgement were implicit throughout the documents and, consequently, reference to particular locations was not useful.

Developing a framework for the analysis of evaluation practices as reported in documents and applying this framework to school district practices in British Columbia in order to provide a description of these practices in the province were the first two stages in fulfilling the purpose of this study. The framework which was developed and applied in this way was useful in providing information about program evaluation

practice in British Columbia school districts.

2. The Picture of Program Evaluation: On Not Rocking the Boat

Chapter VI contained a discussion of four aspects of program evaluation which emerged from the findings of the study, and which taken together provided evidence for the observation that present program evaluation practices tend to maintain and reinforce, rather than challenge, the status quo (i.e., these evaluations were unlikely to rock the proverbial boat). The four aspects are summarized below.

Consideration of stakeholder involvement led to the conclusion that stakeholders played a relatively passive role in evaluation. Although they were canvassed for information, about the programs and about their opinions of the programs, they rarely had control over the direction and conduct of the evaluation. Consideration of the role of the evaluator resulted in the observation that most of the evaluations were carried out by internal evaluators and as such these evaluators were subject to the influence of the organization. In those cases in which evaluators did take the role of judge, and in which they did make recommendations for action, it was found that the recommendations tended to perpetuate rather than change the status quo in that the suggestions effected incremental, low risk changes to existing processes rather than large-scale, innovative changes with high risk and cost implications. As such, it would seem that the critical function of evaluation could become a monitoring one which would tend to perpetuate rather than change the status quo.

The unexpected emphasis on raising awareness and increasing knowledge prompted the observation that evaluations may fulfil socio-political purposes unrelated to improving or

changing the programs. In addition, however, it was also noted that evaluators tended to emphasize formative rather than summative functions of evaluation. The positive aura that appears to surround school district evaluations (in that the evaluation process does not appear to be at all threatening to the participants, and in that the reports are supportively written), is another factor which tends to reinforce rather than challenge the status quo.

Finally, the identification of criteria was discussed. It was noted that criteria were rarely made explicit although it was possible to identify both process criteria (in terms of the areas on which the evaluators chose to focus), and outcome criteria. It was apparent that evaluators who did make judgements tended to focus on program processes in terms of making existing programs function more effectively and on program outcomes in terms of whether the stakeholders were satisfied with the program. Given that the majority of evaluators were school district employees, and that the source of criteria used as the basis for judgement appeared to be the implicit criteria and standards of educators, major challenges to existing norms seem unlikely.

Consideration of the four issues together with the observation that so few of the cross-tabulations demonstrated significant associations, lead to the conclusion that school district evaluations tend to keep the existing system functioning with incremental changes as necessary in small pockets of the system (i.e., in particular program contexts). With reference to the cross-tabulations, it seems that evaluators who examine school district programs, cover a wide variety of bases. They focus, for example, on a number of different areas, they poll diverse groups of stakeholders and conduct evaluations in a way which is consistent with the notion of eclecticism in

evaluation, i.e., using methods and data collection techniques which are appropriate for a particular evaluation situation. This study does not provide evidence that external evaluators behave differently from internal evaluators; or that evaluations carried out with assistance from a committee, are substantially different from those without a committee. In addition, the areas on which evaluators chose to focus, were similar in those cases where the evaluator was judge and where the evaluator was information broker.

These observations made in the previous paragraphs suggest that evaluation has the effect of maintaining the school district status quo and not of bringing about fundamental change. To see evaluation in the light of a district's change processes is not something which was envisaged at the start or in the design stages of this study, but it does nevertheless shed an interesting light on a relatively unexplored dimension of the evaluation field. The conclusion noted above, that evaluation does not seem to be a vehicle for bringing about fundamental change, is not unrelated to the association between politics and evaluation as discussed by House (1973) and Sroufe (1977). Both writers believe that evaluation is usefully seen as part of the political processes of society and of the institutions within society. Cochran (1978) uses this idea to suggest why evaluation does not lead to fundamental change: social mores may militate against innovation and large-scale change.

C. IMPLICATIONS

The third part of the purpose of the present study was to add to the existing evaluation knowledge base. This section attempts to fulfil this purpose by discussing the implications of the findings both for evaluation practice and for evaluation research.

1. Guidelines for Writing Evaluation Reports

The reading and rereading of the evaluation reports examined for this study has provided this researcher with some insights about how evaluations might usefully be conducted in school districts and, more particularly, on how evaluation reports might be written in order to achieve their desired end, i.e., to inform the reader about the evaluation undertaken. Evaluation reports are intended to communicate information to the client and possibly also to a variety of other interested parties. Although the provision of a formal written report is not the only way of communicating, it is one of the most commonly accepted and efficient ways of providing information.

The reports examined represented a variety of approaches to the communication of information. Reports produced in some school districts followed a standard format developed in those districts. Reports produced in other districts appeared to be a product of individual evaluator preferences, as form, length, and style were found to vary a great deal. Sometimes executive summaries were appended to technical reports and sometimes shortened forms of comprehensive reports were produced for circulation to interested groups. Some reports were narratives; others took the form of a series of tables with introductory and concluding remarks; and, in some cases, a variety of formats were combined. What is important in this discussion, however, is not the reporting format, per se, but rather, the kinds of information which would be useful to include. This, of course, relates to how the evaluations might be conducted.

It is likely that the information included in the evaluation report is a product of the particular evaluation circumstances, and as such is likely to differ considerably from report to report. However, from the examination of the reports included in this study,

it would seem that some kinds of information are common to the majority of reports. In addition, it would seem that there is information which might usefully be included in more evaluation reports than evidenced in this study. In some, for example, there was a paucity of descriptive information about the object under consideration. Indeed, in a few cases, the object of the evaluation could not be clearly identified. In others, the purpose was unclear and, in still others, it was very difficult to determine the criteria used for judgement. In contrast, there were other reports which included sufficient information for this researcher to answer all the questions listed in the framework and to suggest other components which could be included within the categorization scheme. It is plausible to assume that types of information which are included in the majority of reports tend to be important, and should be at least considered by all school district evaluators for inclusion in their reports.

Since the coding categories used in this study are derived partly from the literature and partly from practice, they make reference to what both researchers and practitioners consider to be useful. Although the complexity of the coding scheme may limit its utility for practitioners, the series of questions about evaluation on which they were based, can direct the attention of the evaluator to the kinds of factors to be considered in any evaluation study and can provide the evaluator with a way of organizing the information to be collected. The purpose of this section is to maximize the usefulness of the coding categories for evaluation practitioners in school districts by using them as the foundation for a series of guidelines of the kinds of things that school district evaluators might consider useful for including in their evaluation studies and their evaluation reports. Suggestions for items to be considered are given in the following paragraphs. They are arranged under the headings of the four basic questions which

structured the identification of the coding categories.

Evaluation – To What End?

The role of the evaluator can be considered in two ways, judge or information broker. As such, it might be useful to clarify the limits of evaluator responsibility, i.e., is the evaluator responsible for judgement or the provision of information? In order to substantiate the interpretation of information when the evaluator is judge, it might be useful to ask if supporting information has been included.

With regard to the intents of the evaluation, it might be useful to specify purposes; for example, is the evaluation intended to result in any of the following: improvement and change, decision-making, raised awareness, increased knowledge, understanding, long-term development and planning, informed policy-making, provision of data for comparison, or is the report intended to meet funding requirements or to meet accountability demands? In addition, is the function of the evaluation specified? Is the evaluation formative, summative or both? Is the reason for the evaluation specified? Is the evaluation required by school district policy or established practice? Is it required by program design or short term objective? Has the evaluation been requested? If so, who made the request? Did the request come from school district personnel or a parent group?

It is useful to specify exactly what is to be evaluated. For example, is the object of the evaluation an area of content, a way of organizing for instruction, an organizational unit or particular facilities? Is the program continuing or temporary? What aspects

are considered? Who is the sponsor? Where is the program located? What grade levels are affected?

Evaluation — By What Means?

This section focusses on the information to be collected and reported. Questions could include: From whom is the information to be collected? Sources might be: senior administrators, program operating personnel, school-based administrators, trustees, parents, members of local organizations (community groups and agencies), student program participants, non-program participants, or post-program participants. Is information to be collected from people outside the district? Sources might be: district staff from other districts, faculty members (and graduate students) from colleges and universities, private consultants, representatives from provincial ministries, or from outside experts in the particular program area. What kinds of written materials are to be used as information sources? These might include documents and records or literature reviews. Is the evaluator to be used as an information source? Ways this would occur could include making observations of the program in action or assigning tasks or tests. What kind of information is to be collected? This could include opinion or descriptive information, about such things as: process, outcomes, participants, or comparable programs.

Given that in so few cases were criteria for judgement made explicit, it might be useful to consider whether or not criteria should be made explicit or whether the evaluator should rely on criteria implicit in the identification of areas of focus. The prospective evaluator might ask whether the areas of focus which are indicative of implicit criteria

are adequate. Other criteria related questions might include: Is the source of criteria identified? Identified sources might include: written school district or Ministry guidelines or program goals and objectives. Is the nature of the criteria identified? Are the criteria process oriented? On what do they focus? Areas of focus might include: philosophy, program and procedures, policy and administrative practices, personnel, instructional practices, professional development, buildings and facilities, materials and equipment, evaluation and research or community relations, or these in any combination. Are the criteria outcome-oriented? On what do they focus? Areas might include: student achievement, student behaviour and attitudes, indications of a change in state, or stakeholder satisfaction.

The technical aspects of an evaluation are also pertinent. Here questions might include: Are methods of inquiry specified? What is the most appropriate approach given the purposes of the evaluation? For example, would an experimental/quasi-experimental approach be more useful than a descriptive one or vice versa? What types of data collection techniques are to be used? Types of data collection technique might include: questionnaires, interviews and meetings, on-site observation, checklists, rating scales, attitude scales and inventories, or achievement measures.

Evaluation – By Whom and For Whom?

The third basic question concerns the evaluators and the recipients of the evaluation information. Useful questions might include: Are the designated evaluators named? Are their positions given? Are biographical data included? Are advisory groups involved? If there are to be advisory groups, who are the members? What is the

group's mandate?

In terms of the report recipients, it might be useful to make explicit their identities, and those of other pertinent audiences. Questions might include: Is the recipient of the report indicated in the report itself? Is the date of report submission given? Is the school district identified?

Evaluation – With What Conclusion?

This question is closely linked with the overall definition of the evaluation. If the evaluator takes a judgemental role, is he or she responsible for making recommendations? If yes, further questions might include: What kinds of actions are recommended? Are these actions for continuation, termination, change? Are they for the program as a whole? Are they for particular program activities? In which areas are the recommendations? Possible areas of focus are: philosophy, program goals, procedures, policy, administrative practices, personnel, instructional practices, professional development, facilities, materials, further evaluation or research, and relations between program personnel and the community.

2. Implications for Further Research

The framework developed in this study and the information collected can facilitate further research on program evaluation. The study has provided a data base for those who want to test hypotheses about such issues as what gets evaluated; what audiences are served; what approaches, models, and designs are used; who conducts evaluations; and what questions are addressed. Implications for further research are grouped in

seven areas discussed below.

First, as noted in the delimitations of the study, the data base consisted solely of the written reports submitted. Even if the framework developed for this study is used in other studies, the question of the extent to which written evaluation reports accurately reflect evaluation practice remains. This is difficult to investigate. Further exploration is necessary; the literature on case studies might provide fruitful points of departure. In addition, comparison of written evaluation reports with verbal reports from other data sources such as the participants in the evaluation process could serve to illuminate this question.

Second, although there would appear to be no other provincial picture of school district evaluation available, anecdotal evidence from school district administrators and from those involved as evaluation consultants suggests that the picture in other provinces is not dissimilar. The framework can serve usefully as a basis for organizing data collected on program evaluation in school districts outside British Columbia in order that comparisons of British Columbia school district practices can be made with other jurisdictions.

Third, the data contained little systematic evidence about the structures used by school districts for managing their evaluation processes. There was no notable variation by district. However, initial inquiries yielded two kinds of response from districts which did not have formal evaluation reports. Respondents from small districts commented that there were insufficient staff to engage in formal evaluation, while a respondent from a large district said that no evaluations per se were produced, as programs were

monitored on a continuous basis which was not conducive to the production of written evaluation reports. No information was collected about whether or not the school districts submitting evaluation reports had established evaluation and research departments. Research studies could address questions about whether school districts with established evaluation and research departments engage in different kinds of evaluative activities than school districts without such departments. Other research questions might address the reasons why only some school districts engage in formal evaluation activities which result in the production of formal evaluation reports.

Fourth, this study adumbrated a possible relationship between use of internal, as opposed to external, evaluators and evaluation outcomes which reinforce the status quo. Because there were insufficient numbers of external evaluators in this study (and even most of those were school district personnel), no conclusions can be drawn. However, the effect on evaluation outcomes of using external, rather than internal, evaluators warrants further investigation.

Fifth, it was noted that there was scant evidence of full stakeholder participation in the design and implementation of the evaluations. Research is needed on the extent to which such participation affects the form and content of written reports; in particular, whether such reports tend to evaluate programs by the same criteria and whether they suggest more sweeping changes.

Sixth, no information on the dissemination or utilization of the evaluation results was obtained. More could be learned about the impacts of the evaluations if information were collected in these areas. Such research might include a study of the factors

affecting implementation of evaluation recommendations; for example, by considering whether the type of evaluation report produced had any effect on the utilization of the information presented. Such studies would also contribute to an understanding of the relationship between program evaluation and the status quo.

Finally, the relationship between politics and evaluation, already referred to, warrants further study: to what extent is evaluation politically motivated? According to Easton (1965), politics is concerned with the authoritative allocation of values. Evaluation, as it appears to be practiced in the school districts of British Columbia, seems to be less a mechanism for the allocation of values than for the reaffirmation of the way values have already been allocated. That is to say, a change in the allocation of values does not appear to be likely from these evaluations. It would be interesting to look at school districts which have initiated major change (i.e., that have not perpetuated the status quo) to see what role, if any, evaluation played in such change. It might be hypothesized that a decision for such change would precede evaluation and that any evaluation undertaken would be an examination of ways of implementing the change, rather than an evaluation to determine whether or not the change should be introduced. In the present study, some support for such a hypothesis can be found in the evaluation of pilot programs. Although not discussed at great length in this text, it would appear that pilot studies concerned, for example, with technical innovation or with the prevention of child abuse, focussed on ways by which these programs could be implemented most effectively (e.g., some reports of pilot studies contained recommendations for improvement), rather than with informing the decision about whether to introduce the programs into schools.

The coding scheme developed in this study for the analysis of program evaluation practices has illuminated program evaluation practices in British Columbia school districts. The picture which emerges from the documents studied is that program evaluation, in those British Columbia school districts which do produce formal written reports, is essentially conservative. Evaluations are predominantly internal; they provide information and, as appropriate, the evaluators make judgements and suggest incremental changes. Although the evidence is still inconclusive, this study has raised important questions with regard to the extent to which the evaluation process is a mechanism which maintains, reinforces and reassures those involved in the implementation of school district programs.

REFERENCES

- Alkin, M.C., Daillak, R., and P. White
1979 Using Evaluations: Does Evaluation Make a Difference? Beverly Hills: Sage Publications.
- Auman, J.
1987 A Documentary Analysis of the British Columbia School Health Programme (Secondary). Unpublished Doctoral dissertation, The University of British Columbia.
- Babbie, E.
1986 The Practice of Social Research (4th ed.) Belmont: Wadsworth.
- Borg, W.R., and M.D. Gall
1983 Educational Research: An Introduction (4th ed.). New York: Longman.
- Borich, G.D.
1983 "Evaluation models: A question of purpose not terminology." Educational Evaluation and Policy Analysis, 5(1):61-63.
- Boruch, R.F., McSweeney, A.J., and E.J. Soderstrom
1978 "Randomized field experiments for program planning, development, and evaluation: An illustrative bibliography." Evaluation Quarterly, 2(4):655-695.
- Braskamp, L.A., and R.D. Brown
1980 "Utilization of evaluative information." In New Directions for Program Evaluation, 5. San Francisco: Jossey-Bass.
- Bryk, A.S. (ed.)
1983 Stakeholder-Based Evaluation. New Directions for Program Evaluation, 17. San Francisco: Jossey-Bass.
- Bryk, A.S., and J. Light
1981 "Designing evaluations for different program environments." In R.A. Berk (ed.), Educational Evaluation Methodology: The State of the Art. Baltimore: The Johns Hopkins University Press.

- Bryk, A.S. and S.W. Raudenbush
1983 "The potential contribution of program evaluation to social problem solving: A view based on the CIS and Push/Excel experiences." In A.S. Bryk (ed.), Stakeholder-Based Evaluation. New Directions for Program Evaluation, 17. San Francisco: Jossey-Bass.
- Campbell D.T.
1969 "Reforms as experiments." American Psychologist, 24:409-29.
- Campbell D.T., and J.C. Stanley
1966 Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.
- Carney, T.F.
1972 Content Analysis: A Technique for Systematic Inference from Communication. Winnipeg: University of Manitoba Press.
- Catterall, J.S. (ed.)
1985 Economic Evaluation of Public Programs. New Directions for Program Evaluation, 26. San Francisco: Jossey-Bass.
- Ciarlo, J.A. (ed.)
1981 Utilizing Evaluation: Concepts and Measurement Techniques. Beverly Hills: Sage Publications.
- Cochran, N.
1978 "Cognitive processes, social mores, and the accumulation of data: Program evaluation and the status quo." Evaluation Quarterly, 2(2):343-358.
- Cook, T.D., and D.T. Campbell
1979 Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.
- Conner, R.F.
1981 Methodological Advances in Evaluation Research. Beverly Hills: Sage Publications.
- Conner R.F., Altman, D.G., and C. Jackson (eds.)
1984 Evaluation Studies Review Annual, 9. Beverly Hills: Sage Publications.

- Cronbach, L.J.
1963 "Course improvement through evaluation." In Worthen, B.R., and J.R. Sanders, *Educational Evaluation: Theory and Practice*. Worthington, Ohio: Charles A. Jones, 1973.
- Cronbach, L.J.
1982 *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.
- Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., and S.S. Weiner
1980 *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*. San Francisco: Jossey-Bass.
- Easton, D.
1965 *A Framework for Political Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Eisner, E.W.
1976 "Educational connoisseurship and criticism: Their form and function in educational evaluation." *Journal of Aesthetic Education*, 10:135-50.
- Eisner, E.W.
1979a *The Educational Imagination: On the Design and Evaluation of School Programs*. New York: Macmillan Publishing Company.
- Eisner, E.W.
1979b "The use of qualitative forms of evaluation for improving educational practice." *Educational Evaluation and Policy Analysis*, 1(6):11-19.
- Eisner, E.W.
1981 "On the differences between scientific and artistic approaches to qualitative research." *Educational Researcher*, 10(4):5-9.
- Eisner, E.W.
1985 *The Educational Imagination: On the Design and Evaluation of School Programs* (2nd ed.) New York: Macmillan Publishing Company.

- Farrar, E., and E.R. House
1983 "The evaluation of Push/Excel: A case study." In A.S. Bryk (ed.), Stakeholder-Based Evaluation. New Directions for Program Evaluation, 17. San Francisco: Jossey-Bass.
- Feldman M.S., and J.G. March
1981 "Information in organizations as symbol and signal." Administrative Science Quarterly, 26:171-186.
- Gardner, D.E.
1977 "Five evaluation frameworks: Implications for decision making in higher education." Journal of Higher Education, 48(5):571-593.
- Glass, G.V., and F.S. Ellett
1980 "Evaluation research." Annual Review of Psychology, 31:211-28.
- Greene, J.G.
1988 "Stakeholder participation and utilization in program evaluation." Evaluation Review, 12(2):91-116.
- Guba, E.G.
1969 "The failure of educational evaluation." Educational Technology, 9:29-38.
- Guba, E.G.
1978 "Toward a methodology of naturalistic inquiry in educational evaluation." Monograph Series, 8. Los Angeles: University of California. ED164599.
- Guba, E.G., and Y.S. Lincoln
1981 Effective Evaluation. San Francisco: Jossey-Bass.
- Hamilton, D., Jenkins, D., King, C., Macdonald, B., and M. Parlett (eds.)
1977 Beyond the Numbers Game. Basingstoke and London: Macmillan Education.
- Hammond, R.
1969 "Context evaluation of instruction in local school districts." Educational Technology, 9(1):13-18.

- Hatch, J.A.
1983 "Applications of qualitative methods to program evaluation in education." *Viewpoints in Teaching and Learning*, 59(1):1-11.
- Hills, R.J., and C. Gibson
1988 *Problem Analysis and Reformulation Skills for Administrators*. Unpublished manuscript, The University of British Columbia.
- Holsti, O.R.
1969 *Content Analysis for the Social Sciences and Humanities*. Don Mills, Reading, MA: Addison-Wesley.
- Hoole, F.W.
1978 *Evaluation Research and Development Activities*. Beverly Hills: Sage Publications.
- House, E.R. (ed.)
1973 *School Evaluation: The Politics and Process*. Berkeley: McCutchan Publishing Corporation.
- House, E.R.
1978 "Assumptions underlying evaluation models." *Educational Researcher*, 7(3):4-11.
- House, E.R.
1980 *Evaluating with Validity*. Beverly Hills: Sage Publications.
- House, E.R. (ed.)
1986 *New Directions in Educational Evaluation*. East Sussex, U.K.: The Falmer Press.
- Hutchinson, B., Hopkins, D., and J. Howard
1988 "The problem of validity in the qualitative evaluation of categorically funded curriculum development projects." *Educational Research*, 30(1):54-64.
- Joint Committee on Standards for Educational Evaluation
1981 *Standards for Evaluation of Educational Programs, Projects, and Materials*. New York: McGraw-Hill Book Company.

- Kennedy, M.M.
1983 "The role of the in-house evaluator." *Evaluation Review*, 7(4):519-41.
- Koppelman, K.L.
1979 "The explication model: An anthropological approach to program evaluation." *Educational Evaluation and Policy Analysis*, 3(1):59-64.
- Kosecoff, J., and A. Fink
1982 *Evaluation Basics: A Practitioner's Manual*. Beverly Hills: Sage Publications.
- Krippendorff, K.
1980 *Content Analysis: An Introduction to its Methodology*. Beverly Hills: Sage Publications.
- Lawler, E.E.
1985 "Challenging traditional research assumptions." In Lawler *et al*, *Doing Research that is Useful for Theory and Practice*. San Francisco: Jossey-Bass.
- Levin, H.M.
1975 "Cost-effectiveness analysis in evaluation research." In M. Guttentag and E.L. Streuning (eds.), *Handbook of Evaluation Research*, 2. Beverly Hills: Sage Publications.
- Leviton, L.C., and R.F. Boruch
1983 "Contributions of evaluation to educational programs and policy." *Evaluation Review*, 7:563-598.
- Leviton, L.C., and E.F.X. Hughes
1981 "A review and synthesis of research on the utilization of evaluations." *Evaluation Review*, 5:525-548.
- Lindkvist, K.
1981 "Approaches to textual analysis." In K.E. Rosengren (ed.), *Advances in Content Analysis*. *Annual Review of Communicative Research*, 9. Beverly Hills: Sage Publications.
- Madaus, G.F., Scriven, M.S., and D.L. Stufflebeam (eds.)
1983 *Evaluation Models: Viewpoints on Educational and Human Services*. Boston: Kluwer-Nijhoff Publishing.

- Marcus, L.R., and B.D. Stickney
1981 Race and Education: The Unending Controversy. Springfield, Ill.: Charles C. Thomas Publisher.
- Miles, M.B., and A.M. Huberman
1984 Qualitative Data Analysis: A Sourcebook of New Methods. Beverly Hills: Sage Publications.
- Nevo, D.
1983 "The conceptualization of educational evaluation: An analytical review of the literature." *Review of Educational Research*, 53(1):117-128.
- Nevo, D.
1984 "The conceptualization of educational evaluation: An analytical review of the literature." In E.R. House (ed.), *New Directions in Educational Evaluation*. East Sussex, U.K.: The Falmer Press.
- Owens, T.R.
1973 "Educational evaluation by adversary proceedings." In E.R. House (ed.), *School Evaluation: The Politics and Process*. Berkeley: McCutchan Publishing Corporation.
- Parlett, M., and D. Hamilton
1976 "Evaluation as illumination: A new approach to the study of innovatory programs." In G.V. Glass (ed.), *Evaluation Studies Review Annual*, 1. Beverly Hills: Sage Publications.
- Patton, M.Q.
1978 *Utilization Focused Evaluation*. Beverly Hills: Sage Publications.
- Patton, M.Q.
1980 *Qualitative Evaluation Methods*. Beverly Hills: Sage Publications.
- Patton, M.Q.
1982 *Practical Evaluation*. Beverly Hills: Sage Publications.
- Pietro, D.S. (ed.)
1983 *Evaluation Sourcebook for Private and Voluntary Organizations*. New York: American Council of Voluntary Agencies for Foreign Service.

- Popham, W.J.
1975 Educational Evaluation. Englewood Cliffs, NJ: Prentice-Hall.
- Provus, M.
1969 "Evaluation of ongoing programs in the public school system." In Worthen, B.R., and J.R. Sanders, Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones, 1973.
- Provus, M.
1971 Discrepancy Evaluation for Educational Program Improvement and Assessment. Berkeley: McCutchan Publishing Corporation.
- Rayborn, R.R.
1986 An Examination of the Role and Position of the Program Evaluator in Washington State Public Schools. Everett, WA: Washington Educational Research Association.
- Riecken, H.W., and Boruch, R.F. (eds.)
1974 Social Experimentation. New York: Academic Press.
- Rippey, R.M. (ed.)
1973 Studies in Transactional Evaluation. Berkeley: McCutchan Publishing Corporation.
- Rivlin, A.M.
1971 Systematic Thinking for Social Action. Washington, D.C.: The Brookings Institution.
- Rossi, P.H., and H.E. Freeman
1985 Evaluation: A Systematic Approach (3rd. ed.) Beverly Hills: Sage Publications.
- Rothenberg, J.
1975 "Cost-benefit analysis: A methodological exposition." In M. Guttentag, and E.L. Streuning (eds.) Handbook of Evaluation Research, 2. Beverly Hills: Sage Publications.

- Rutman, L. (ed.)
1984 Evaluation Research Methods: A Basic Guide (2nd ed.) Beverly Hills: Sage Publications.
- Scriven, M.
1967 "The methodology of evaluation." In R.E. Stake (ed.), Curriculum Evaluation. American Educational Research Association Monograph Series on Evaluation, 1. Chicago: Rand McNalley.
- Scriven, M.
1972 "Pros and cons about goal-free evaluation." Evaluation Comment, 3(4):1-7.
- Scriven, M.
1973 "Goal-free evaluation." In E.R. House (ed.), School Evaluation: The Politics and Process. Berkeley: McCutchan Publishing Corporation.
- Scriven, M.
1983 "Evaluation ideologies." In G.F. Madaus, M.S. Scriven, and D.L. Stufflebeam (eds.), Evaluation Models: Viewpoints on Educational and Human Services. Boston: Kluwer-Nijhoff Publishing.
- Shapiro, J.Z.
1986 "Educational research and educational decision-making." In J. Smart (ed.), Higher Education Handbook of Theory and Research, 11. New York: Agathon Press.
- Shavelson, R.J.
1988 "The 1988 presidential address, contributions of educational research to policy and practice: Constructing, challenging, changing cognition." Educational Researcher, 17(7):4-11,22.
- Smith, E.R., and R.W. Tyler
1942 Appraising and Recording Student Progress. New York: Harper and Row.
- Smith, N.L.
1979 "Requirements for a discipline of evaluation." Studies in Educational Evaluation, 5:5-12.

- Smith, N.L.
1980 "The progress of educational evaluation: Rounding the first bends in the river." Proceedings of the 1980 Minnesota Evaluation Conference on Educational Evaluation. In G.F. Madaus, M.S. Scriven, and D.L. Stufflebeam (eds.), Evaluation Models: Viewpoints on Educational and Human Services. Boston: Kluwer-Nijhoff Publishing, 1983.
- Smith, N.L. (ed.)
1981a Metaphors for Evaluation: Sources of New Methods. Beverly Hills: Sage Publications.
- Smith, N.L.
1981b "Evaluating evaluation methods." Studies in Educational Evaluation, 7:173-181.
- Smith, N.L.
1982 "Evaluation design as preserving valued qualities in evaluation studies." Studies in Educational Evaluation, 7:229-237.
- Sroufe, G.E.
1977 "Evaluation and politics." In J.D. Scribner (ed.), The Politics of Education. The Seventy-sixth Yearbook of the National Society for the Study of Education. Chicago: The University of Chicago Press.
- Stake, R.E.
1967 "The countenance of educational evaluation." Teachers College Record, 68:523-540.
- Stake, R.E.
1975a Program Evaluation, Particularly Responsive Evaluation. Paper #5 in Occasional Paper Series. Kalamazoo: Western Michigan University. ED163060.
- Stake, R.E. (ed.)
1975b Evaluating the Arts in Education: A Responsive Approach. Columbus, Ohio: Merrill.
- Stake, R.E.
1976a "A theoretical statement of Responsive Evaluation." Studies in Educational Evaluation, 2(1):19-22.

- Stake, R.E. (ed.)
1976b Evaluating Educational Programs: The Need and the Response. Washington, D.C.: OECD Publications.
- Stufflebeam, D.L.
1975 "Evaluation as a community education process." Community Education Journal, 5(2):7-12,19.
- Stufflebeam, D.L.
1981 A Review of Progress in Educational Evaluation. Paper presented at the annual review of the Evaluation Network, Austin, TX. ED216031.
- Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., and M.M. Provus
1971 Educational Evaluation and Decision Making. Itasca, Ill.: F.E. Peacock Publishers.
- Stufflebeam, D.L., and W.J. Webster
1980 "An analysis of alternative approaches to evaluation." Educational Evaluation and Policy Analysis, 2(3):5-20.
- Stufflebeam, D.L., and W.L. Welch
1986 "Review of research on program evaluation in United States school districts." Educational Administration Quarterly, 22(3):150-170.
- Suchman, E.A.
1967 Evaluative Research: Principles and Practice in Public Service and Social Action Programs. New York: Russell Sage Foundation.
- Talmage, H.
1982 "Evaluation of programs." In H.E. Mitzel and Associates (eds.), Encyclopedia of Educational Research (5th ed.), 2. New York: The Free Press, Macmillan.
- Thompson, M. S.
1980 Benefit-Cost Analysis for Program Evaluation. Beverly Hills: Sage Publications.

- Tyler, R.W.
1942 "General statement on evaluation." *Journal of Educational Research*, 35:492-501.
- Tyler, R.W.
1949 *Basic Principles of Curriculum and Instruction*. Chicago, Ill.: University of Chicago Press.
- Weber, R.P.
1985 *Basic Content Analysis. Quantitative Applications in the Social Sciences*, 49. Beverly Hills: Sage Publications.
- Weiss, C.H.
1972 *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.
- Weiss, C.H.
1982 "Measuring the use of evaluation." In E.R. House and Associates (eds.), *Evaluation Studies Review Annual*, 7. Beverly Hills: Sage Publications.
- Werner, W.
1979 "Evaluation: Making sense of school programs." *Curriculum, Media, and Instruction Publication Series*, 11. Edmonton: Department of Secondary Education, University of Alberta.
- Willis, G. (ed.)
1978 *Qualitative Evaluation*. Berkeley: McCutchan Publishing Corporation.
- Wolf, R.L.
1979 "The use of judicial evaluation methods in the formulation of educational policy." *Educational Evaluation and Policy Analysis*, 1(3):19-28.
- Worthen, B.R., and J.R. Sanders
1973 *Educational Evaluation: Theory and Practice*. Worthington, Ohio: Charles A. Jones.

APPENDIX 1. CORRESPONDENCE

1.1 Initial letter mailed to superintendents of school districts in British Columbia

1.2 Enclosure



DEPARTMENT OF ADMINISTRATIVE, ADULT AND HIGHER EDUCATION
THE UNIVERSITY OF BRITISH COLUMBIA

October 31, 1985

The purpose of this letter is to ask for your help in my study of program evaluation in school districts in British Columbia.

The study is being conducted as part of my doctoral programme at the University of British Columbia under the supervision of Dr. Graham Kelsey. At this stage I am doing no more than assessing the availability of data. To that end I enclose a form which I would be most grateful if you would complete and return to me in the enclosed stamped, addressed envelope.

Thank you for your attention.

Yours sincerely,

Enclosure

Trisha Wilcox

Program Evaluation in B.C. School Districts

Name of District: _____

1. Has your district engaged in any formal program evaluation activities in the last five years?
Yes [] No []
2. If yes, did any of these studies result in the production of a written report?
Yes [] No []
3. If yes, would you be willing to provide me with copies of some or all of these reports?
Yes [] No [] Perhaps []

If perhaps, please explain: _____

4. If yes, would you or a member of your staff be willing to answer further questions about them?
Yes [] No [] Perhaps []

If perhaps, please explain: _____

Name of Contact Person: _____

Telephone Number: _____

Thank you for your cooperation. Please return this sheet in the stamped, addressed envelope provided.

APPENDIX 2. REPORTS EXCLUDED IN THE FIRST STAGE OF ANALYSIS

List of Excluded Reports

Report #003 and Report #063

Reports #003 and #063 describe the development of tests to be used for student placement. As both reports focus on test development for individual student assessment, they are not included in the present study.

Report #050

This document reports on a province-wide study sponsored by the Ministry of Education. It focusses on standard setting and lists recommendations for the Ministry and for school districts across the province.

Report #062

This is an administrative overview indicating which skills should be taught at which grade level (K-12). Although there is some indication of the developmental process engaged in by district personnel, the document presents the final stage of the process, i.e., the finished skills continuum.

Report #065

Report #065 is a description of the third stage of a long-term assessment. The purpose of this stage is to establish if the students from two districts, the target and the control district, have comparable levels of skill development in a specific area. As the long-term assessment is not described in its entirety, report #065 has been excluded.

Report #066

This document was excluded because it reports an analysis of student results on Ministry sponsored achievement tests.

Report #069

One of the four sections of this report was not submitted. As incomplete reports are not included in the study, this report was rejected.

Report #085

This report describes three dimensions of a project planned for the following school year. Each dimension includes strategies for the improvement of services. As the document does not report on actual evaluative practices which have already taken place, it has been excluded.

Report #090

This document takes the form of a proposal to decentralize the management of a number of similar programs. Issues involved in decentralization are listed and discussed but there is no information on a specific object or objects.

Report #093 and Report #094

These evaluations form part of a province-wide study. Although the data are collected

in a specific school district, the evaluations were conceived and implemented by evaluators contracted by the Ministry of Education. Given that the purpose of this thesis is the examination of school district program evaluations, reports #093 and #094 have been excluded.

Report #097

Much of this report is illegible. On pages where sections have been highlighted and then photocopied the marked sections cannot be read. Much of the description and all of the recommendations are marked in this way.

APPENDIX 3

CODING INSTRUMENT

Throughout this coding instrument the following notes apply:

1. (R) denotes the use of real numbers.
2. Code values of "8" ("88" in a two column code) and "9" ("99" in a two column code) always carry the same special meanings. "8" ("88") denotes "Other," and "9" ("99") denotes "Not Stated."
3. Code values of "0" and "1" denote "No" and "Yes" respectively unless otherwise specified.

PART I

General Information

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
1. Document ID number (ID)	3	1-3	001	001-110(R)
2. Year document produced (YEAR)	2	5-6	80	80-86(R)
3. School District (SD) number (SDNO)	2	7-8	18	01-92(R)
4. SD Size (SDSIZE)	1	9	1	1-5
a. Very small (1-1000 students)				a=1
b. Small (1001-3000)				b=2
c. Medium (3001-8000)				c=3
d. Large (8001-15000)				d=4
e. Very large (15000-up)				e=5
5. Number of documents submitted by SD (DOCNO)	2	10-11	01	01-99(R)

PART II

Question 1:

How was Evaluation Defined?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
6. Definition: Evaluator as judge or information provider (Implied) (DEFN)	1	14	1	1-5
a. Judgement (no supporting information)				a=1
b. Judgement (some supporting information)				b=2
c. Judgement (much supporting information)				c=3
d. Provision of information (some judgement)				d=4
e. Provision of information (no judgement)				e=5

PART III

Question 2:

What were the Intents of the Evaluation?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
Purpose (Stated or Implied)				
7. Improvement and Change (PURPIMP)	1	17	1	0,1
8. Decision Making (PURPDEC)	1	18	1	0,1
9. Development and Planning (PURPDEV)	1	19	1	0,1
10. Awareness and Knowledge (PURPAWR)	1	20	1	0,1
11. Accountability (PURPACCT)	1	21	1	0,1
12. Requirements of Funding Agency (PURPFUND)	1	22	1	0,1
13. Informing the Policy Process (PURPPOL)	1	23	1	0,1
14. Understanding (PURPUNDS)	1	24	1	0,1
15. Provision of Comparative Data (PURPCOMP)	1	25	1	0,1
16. Other (See hand notes to list) (PURPOTHR)	1	26	1	0,1
17. Function: Formative or Summative or both (FUNCTION)	1	28	1	0-8
a. No statement made, no inference possible				a=0
b. Formative (specified)				b=1
c. Formative (implied)				c=2
d. Summative (specified)				d=3
e. Summative (implied)				e=4
f. Formative and Summative (both specified)				f=5
g. Formative and Summative (both implied)				g=6
h. Formative specified, Summative implied				h=7
i. Formative implied, Summative specified				i=8

PART IV

Question 3:

Why was the Evaluation Undertaken?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
Reason-Required				
18. Required by SD policy or established practice (RSREQRPP)	1	31	1	0,1
19. Required by program design or SD short term objective (RSREQRRDO)	1	32	1	0,1
20. Reason-Requested (by): (RSREQUEST)	1	33	1	0-9
a. Program Sponsors				a=1
b. School Staff				b=2
c. Trustees (no indication that eval. done as part of policy or practice)				c=3
d. Senior SD Admin.				d=4
e. Representative Cttee.				e=5
f. Evaluation Cttee.				f=5
21. Other/Not Stated (RSOTHR)	1	34	1	0,1,9

PART V

Question 4:

What was the Object Evaluated?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
PROGRAM OR PROGRAM-RELATED PRACTICE				
22. Curriculum And Instruction (OBJCI)	2	37-38	01	00-99
a. French				a=01
b. Counselling				b=02
c. Computers				c=03
d. Personal Safety				d=04
e. Career Preparation and Access				e=05
f. Science				f=06
g. Music				g=07
h. Library				h=08
i. Learning Assistance				i=09
j. Reading				j=10
k. Physical Education				k=11
l. International Baccalaureate				l=12
m. Police Liaison				m=13
n. Consumer Education				n=14
o. Art/Science				o=15
p. Continuing Adult				p=16
q. Instructional Innovation				q=17
r. Grading Practices				r=18
s. Other				s=88

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
Testing				
23. Writing (OBJWR)	1	39	1	0,1
24. Mathematics (OBJMA)	1	40	1	0,1
25. Reading (OBJRD)	1	41	1	0,1
26. Chemistry (OBJCH)	1	42	1	0,1
27. Other (OBJOTHR)	1	43	1	0,1
28. Special Services (OBJSS)	2	44-45	01	00-99
a. Special Programs (composite)				a=01
b. Native Programs (composite)				b=02
c. Native Program				c=03
d. Alternate/Rehabilitation				d=04
e. Gifted and Talented				e=05
f. Learning Disabled (including diagnostic and extended skills centres)				f=06
g. Behaviour Disturbed				g=07
h. Hospital				h=08
i. Residential				i=09
j. Multiple Handicaps				j=10
k. Hearing Impaired				k=11
l. TMH (interface with regular program)				l=12
m. Other				m=88
29. Organization for Instruction (OBJORG)	1	46	1	0-9
a. Split-grade Classes				a=1
b. Computer Aids				b=2
c. Other				c=8
30. ORGANIZATIONAL UNIT (ORGUNIT)	1	47	1	0-9
a. School				a=1
b. Centre				b=2
c. Other				c=8

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
31. FACILITIES (FACILITS)	1	48	1	0-9
a. Portable Classrooms				a=1
b. Optimum Use of Space				b=2
c. Other				c=8
ADDITIONAL CHARACTERISTICS				
Permanence				
32. Continuing (PERMCONT)	1	50	1	0,1
33. Temporary (PERMTEMP)	1	51	1	0-8
a. Pilot				a=1
b. Project				b=2
c. Display				c=3
d. Equipment				d=4
e. Other				e=8
Aspect Evaluated				
34. General, Global View (ASPGEN)	1	53	1	0,1
35. Specific Aspects (ASPSPEC)	1	54	1	0-8
a. Delivery of Services				a=1
b. Impact				b=2
c. Needs Assessment				c=3
d. Forward Planning				d=4
e. Student Life				e=5
f. Consistence, Prediction				f=6
g. Reading				g=7
h. Other				h=8
36. Object Sponsor (ORGLOCSP)	1	56	1	1-9
a. School or Centre				a=1
b. District				b=2

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
c. Joint Sponsorship				c=3
d. Non-district Agency				d=4
e. Other/Not Stated				8,9
37. Object Base (ORGLOCBS)	1	57	1	1-9
a. School or Centre				a=1
b. District				b=2
c. School and District				c=3
d. Non-district Facilities				d=4
e. Other/Not Stated				8,9
38. Grade Level (ORGLOCGL)	2	58-59	01	00-99
a. Primary				a=01
b. Intermediate				b=02
c. Elementary				c=03
d. Elementary/Junior Secondary				d=04
e. Junior Secondary				e=05
f. Senior Secondary				f=06
g. Secondary				g=07
h. Elementary/Secondary				h=08
i. Adult				i=09
j. Other/Not Stated				j=88/99

PART VI

Question 5:

**What Kinds of Information Regarding each Object
were Reported?**

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
SOURCE:				
PEOPLE OTHER THAN EVALUATOR				
Stakeholders (SD employees)				
39. Senior SD Administrators (INFSTADM)	1	1	1	0,1
40. District Office Staff (INFSTOFF)	1	2	1	0,1
41. Program-operating Personnel (INFSTPOP)	1	3	1	0,1
42. Program-using Personnel (INFSTPUP)	1	4	1	0,1
43. Uninvolved Personnel (INFSTUNP)	1	5	1	0,1
44. School-based Administrators (INFSTSAD)	1	6	1	0,1
45. Other/Not Stated (INFSTOTH)	1	7	1	0,1,9
Stakeholders (non SD employees)				
46. Trustees (INFSTRS)	1	8	1	0,1
47. Parents (INFSTPRT)	1	9	1	0,1
48. Members of Local Organizations (community groups and agencies) (INFSTORM)	1	10	1	0,1
49. Community Members (at large) (INFSTCMM)	1	11	1	0,1
50. Faculty (universities or colleges) (INFSTFAC)	1	12	1	0,1
51. Provincial Ministry Employees (INFSTMIN)	1	13	1	0,1
52. Student Program Participants (INFSTSP)	1	14	1	0,1
53. Non-program Participants (students) (INFSTNPP)	1	15	1	0,1
54. Post-program Participants (students who have graduated or left program) (INFSTPPP)	1	16	1	0,1
55. Other/Not Stated (INFSTOTR)	1	17	1	0,1,9

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
Externals				
56. District Staff from other districts (INFEXDIS)	1	18	1	0,1
57. Faculty Members (and graduate students) from colleges and universities (INFEXFAC)	1	19	1	0,1
58. Private Consultants (INFEXCON)	1	20	1	0,1
59. Representatives from Provincial Ministries (INFEXMIN)	1	21	1	0,1
60. Outside Experts (in program area) (INFEXEXP)	1	22	1	0,1
61. Other/Not Stated (INFEXOTH)	1	23	1	0,1,9
SOURCE: WRITTEN MATERIALS				
62. Documents and Records (INFWRDOC)	1	24	1	0,1
63. Review of the Literature (INFWRRLIT)	1	25	1	0,1
64. Other/Not Stated (INFWRROTH)	1	26	1	0,1,9
SOURCE: EVALUATOR				
65. Observation of program in action (obs. recorded in report) (INFEVOBS)	1	27	1	0,1
66. Evaluator assigned tasks or tests (INFEVTSK)	1	28	1	0,1
67. Other/Not Stated (INFEVOTH)	1	29	1	0,1,9
NATURE				
General				
68. General--Collection of Information (INFGEN)	1	31	1	0-2
a. Not applicable				a=0
b. Exclusively general				b=1
c. Predominantly general				c=2
Specific				
Opinion Information				
69. Process (INFOPPRO)	1	32	1	0,1

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
70. Outcomes (INFOPOUT)	1	33	1	0,1
71. Participants (INFOPPAR)	1	34	1	0-3
a. Not applicable				a=0
b. Students				b=1
c. Staff				c=2
d. Students and staff				d=3
72. Similar objects in other sites (INFOPSIM)	1	35	1	0,1
73. Other (INFOPOTH)	1	36	1	0,1
Descriptive Data and Information				
74. Process (INFDSPRO)	1	37	1	0,1
75. Outcomes (INFDSOUT)	1	38	1	0,1
76. Participants (INFDSPAR)	1	39	1	0-3
a. Not applicable				a=0
b. Students				b=1
c. Staff				c=2
d. Students and staff				d=3
77. Similar objects in other sites (INFDSOOTH)	1	40	1	0,1
78. Other (INFDSOTH)	1	41	1	0,1

Part VII

Question 6:

What Criteria were Used to Judge the Merit and Worth of the Object?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
LOCATION In which sections of the report are the criteria evident?				
79. Criteria identified as such (CRILOIDT)	1	44	1	0,1
80. Checklists (sentence completion) (CRILOCHK)	1	45	1	0,1
81. List of goals and objectives (CRILOGLO)	1	46	1	0,1
82. Evaluation Questions (CRILOEQU)	1	47	1	0,1
83. Evaluator's introductory statements/abstracts (CRIULOINT)	1	48	1	0,1
84. Evaluator's summary statements (CRILOSUM)	1	49	1	0,1
85. Interview protocols or questionnaire items (CRILOITM)	1	50	1	0,1
86. Recommendations (CRILOREC)	1	51	1	0,1
87. Measurement data (CRILOMES)	1	52	1	0,1
88. Other (CRILOOTH)	1	53	1	0,1

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
SOURCE				
89. Written SD or Ministry guidelines (CROSOGDL)	1	55	1	0,1
90. Program goals and objectives (CRISOGLO)	1	56	1	0,1
91. Terms of reference for the evaluation (identified by clients and evaluator (CRISOTOR)	1	57	1	0,1
92. Identified by evaluator(s) (CRISOEID)	1	58	1	0,1
93. Alternate Objects (CRISOALT)	1	59	1	0,1
94. Other (CRISOOTH)	1	60	1	0,1
NATURE				
Process				
95. Adherence to external guidelines (CRINAGDL)	1	62	1	0,1
96. Other (CRINAOTE)	1	63	1	0,1
Areas of Focus:				
97. Philosophy (CRINAPHL)	1	64	1	0,1
98. Program and Procedures (CRINAPRP)	1	65	1	0,1
99. Policy and Admin. Practices ("fit" with SD) (CRINAPLA)	1	66	1	0,1
100. Personnel (CRINAPER)	1	67	1	0,1
101. Instructional Practices (CRINAINS)	1	68	1	0,1
102. Professional Development (CRINAPRD)	1	69	1	0,1
103. Buildings and Facilities (CRINAFAC)	1	70	1	0,1
104. Materials and Equipment (CRINAMAT)	1	71	1	0,1
105. Evaluation and Research (CRINAEVR)	1	72	1	0,1
106. Community Relations (CRINACMR)	1	73	1	0,1
107. Other (CRINAOTH)	1	74	1	0,1
Outcomes				
108. Student Achievement (CRINAACH)	1	75	1	0,1
109. Student Behavior and Attitudes (CRINABEH)	1	76	1	0,1

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
110. Indications of improved or desirable state (in areas other than student performance) (CRINASTA)	1	77	1	0,1
111. Stakeholder Satisfaction (CRINASAT)	1	78	1	0,1
112. Other (CRINAOTR)	1	79	1	0,1

Part VIII

Question 7:

What Methods of Inquiry were Used in the Evaluation?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLES	CODE VALUE
Approach Ascribed by Evaluator(s)				
113. Experimental/Quasi-experimental (METHEXPT)	1	1	1	0,1
114. Descriptive (METHDESC)	1	2	1	0,1
115. Situational Interpretation (METHSITU)	1	3	1	0,1
116. Judgement Matrix (METHJDMT)	1	4	1	0,1
117. Survey (METHSURV)	1	5	1	0,1
118. On-site Visit (METHVSIT)	1	6	1	0,1
119. Other/Not Stated (METHOTHR)	1	7	1	0,1,9
Type of Data Collection Techniques				
120. Questionnaires (DATAQUES)	1	12	1	0,1
121. Interviews and Meetings (DATAINTM)	1	13	1	0,1
122. Documents and Records (DATADOCS)	1	14	1	0,1
123. On-site Observation (DATAOBS)	1	15	1	0,1
124. Off-site Visits and Inquiries (DATAVSIT)	1	16	1	0,1
125. Checklists, Rating Scales, Attitude Scales and Inventories (DATABESC)	1	17	1	0,1
126. Achievement Measures (standardized tests, modified stand. tests, teacher made tests, evaluator assigned tasks, e.g.s of student work (DATAACHM)	1	18	1	0,1
127. Other/Not Stated (DATAOTHR)	1	19	1	0,1,9
No. of Data Collection Techniques				
128. Number of techniques used (DATAANO)	1	20	1	0-9 (R)

Part IX

Question 8:

To Whom Was the Report Submitted?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
Stated Recipient of Evaluation Report				
129. Board of School Trustees (RECTRST)	1	23	1	0,1
130. Superintendent or Senior SD Administrators (RECSUPT)	1	24	1	0,1
131. Evaluation Committee (RECECTE)	1	25	1	0,1
132. Representational Committee (advisory to Board) (RECRCTE)	1	26	1	0,1
133. School Staff (RECSTAF)	1	27	1	0,1
134. Funding Agency (RECRCTE)	1	28	1	0,1
135. Other/Not Stated (RECOTHR)	1	29	1	0,1,9

Part X:

Question 9:

Who Were the Designated Evaluators?

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
EVALUATOR SOURCE:				
STAKEHOLDERS (within SD jurisdiction)				
SD Employees				
136. Senior SD Administrators (EVSTADM)	1	32	1	0,1
137. District Office Staff	1	33	1	0,1
138. Program-operating Personnel (EVSTPOP)	1	34	1	0,1
139. Program-using Personnel (EVSTPUP)	1	35	1	0,1
140. Uninvolved Personnel (EVSTUNP)	1	36	1	0,1
141. School-based Administrators (EVSTSAD)	1	37	1	0,1
142. Other/Not Stated (EVSTOTHR)	1	38	1	0,1,9
Non-SD Employees				
143. Trustees (EVSTTRS)	1	39	1	0,1
144. Parents (EVSTPRT)	1	40	1	0,1
145. Members of Local Organizations (EVSTORM)	1	41	1	0,1
146. Community Members (at large) (EVSTCMM)	1	42	1	0,1
147. Faculty Members (and graduate students) from colleges and universities (EVSTFAC)	1	43	1	0,1
148. Provincial Ministry Employees (EVSTFAC)	1	44	1	0,1
149. Students (EVSTSTU)	1	45	1	0,1
150. Other/Not Stated (EVSTOTH)	1	46	1	0,1,9
EVALUATOR SOURCE:				
EXTERNALS (outside SD jurisdiction)				
151. District Staff on loan from other districts (EVEXDIS)	1	47	1	0,1
152. Faculty Members (and graduate students) from colleges and universities (EVXFAC)	1	48	1	0,1
153. Private Research/Evaluation Consultants (EVEXCON)	1	49	1	0,1

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
154. Representatives from Provincial Ministries (EVEXMIN)	1	50	1	0,1
155. Outside Experts in Program Area (EVEXEXP)	1	51	1	0,1
156. Other/Not Stated (EVEXOTHR)	1	52	1	0,1,9
EVALUATOR NO.				
157. No. of designated evaluators (EVALNO)	2	53-54	01	01-99 (R)
ADVISORY STRUCTURE: NO. OF ADV. GPS				
158. No. of Formal Advisory Groups (ADVGPNNO)	1	56	1	0-9 (R)
ADVISORY STRUCTURE: COMPOSITION OF GPS				
Representatives of:				
159. Senior SD Administrators (ADVGPADM)	1	57	1	0,1
160. District Office Staff (ADVGPOFF)	1	58	1	0,1
161. Program-operating Personnel	1	59	1	0,1
162. Program-using Personnel (ADVGPPUP)	1	60	1	0,1
163. Uninvolved Personnel (ADVGPNUP)	1	61	1	0,1
164. School-based Administrators (ADVGPSAD)	1	62	1	0,1
165. Other/Not Stated (ADVGPOTH)	1	63	1	0,1,9
166. Trustees (ADVGPTRS)	1	64	1	0,1
167. Parents (ADVGPPRT)	1	65	1	0,1
168. Members of Local Organizations (ADVGPORM)	1	66	1	0,1
169. Community Members (at large) (ADVGPCMM)	1	67	1	0,1
170. Faculty Members (and graduate students) from colleges and universities (EVEXFAC)	1	68	1	0,1
171. Provincial Ministry Employees (ADVGPMIN)	1	69	1	0,1
172. Students (ADVGPOSTU)	1	70	1	0,1
173. Other/Not Stated (ADVGPTHR)	1	71	1	0,1,9
174. District Staff on loan from other districts (ADVGPDIS)	1	72	1	0,1
175. Faculty Members (and graduate students) from colleges and universities (EVEXFAC)	1	73	1	0,1

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
176. Private Research/Evaluation Consultants (ADVGPCON)	1	74	1	0,1
177. Representatives from Provincial Ministries (ADVGPMIN)	1	75	1	0,1
178. Outside Experts (content area) (ADVGPEXP)	1	76	1	0,1
179. Other/Not Stated (ADVGPOTR)	1	77	1	0,1,9
ADVISORY STRUCTURE:				
NO. OF MEMBERS				
180. Total Number of Members of Advisory Group(s) (ADVMEMNO)	2	78-79	08	00-40(99)
N.B. "1" indicates whether rep(s) of a particular group is/are present and cols. 78-79 indicate total nos. of members in group.				

Part XI.

Question 10:

What Recommendations (if any) Were Made?

INTENT AND TARGET OF RECOMMENDATIONS

(No. of recommendations of each type in real nos.)

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLES	CODE VALUE
ACTION RECOMMENDED: CONTINUATION				
181. Continuation-Philosophy (CONTPHIL)	2	1-2	01	00-99
182. Continuation-Program and Procedures (CONTPROG)	2	3-4	01	00-99
183. Continuation-Policy and Admin. Practices (CONTPOLC)	2	5-6	01	00-99
184. Continuation-Personnel (CONTPERS)	2	7-8	01	00-99
185. Continuation-Instructional Practices (CONTINST)	2	9-10	01	00-99
186. Continuation-Prof. Development (In-service) (CONTPROD)	2	11-12	01	00-99
187. Continuation-Building and Facilities (CONTFACL)	2	13-14	01	00-99
188. Continuation-Materials and Equipment (CONTMATS)	2	15-16	01	00-99
189. Continuation-Evaluation and Research (CONTEVLR)	2	17-18	01	00-99
190. Continuation-Community Relations (CONTCOMR)	2	19-20	01	00-99
191. Continuation-Other (CONTOTHR)	2	21-22	01	00-99
ACTION RECOMMENDED: MODIFICATION				
192. Modification-Philosophy (MODPHIL)	2	24-25	01	00-99
193. Modification-Program and Procedures (MODPROG)	2	26-27	01	00-99
194. Modification-Policy and Admin. Practices (MODPOLC)	2	28-29	01	00-99

		NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
195.	Modification-Personnel (MODPERS)	2	30-31	01	00-99
196.	Modification-Instructional Practices (MODINST)	2	32-33	01	00-99
197.	Modification-Professional Development (MODPROD)	2	34-35	01	00-99
198.	Modification-Buildings and Facilities (MODFACL)	2	36-37	01	00-99
199.	Modification-Materials and Equipment (MODMATS)	2	38-39	01	00-99
200.	Modification-Evaluation and Research (MODVLR)	2	40-41	01	00-99
201.	Modification-Community Relations (MODCOMRL)	2	42-43	01	00-99
202.	Modification-Other (MODOTHR)	2	44-45	01	00-99
	ACTION RECOMMENDED: INNOVATION				
203.	Innovation-Philosophy (INNPHIL)	2	47-48	01	00-99
204.	Innovation-Program and Procedures (INNPPROG)	2	49-50	01	00-99
205.	Innovation-Policy and Admin. Practices (INNPOLC)	2	51-52	01	00-99
206.	Innovation-Personnel (INNPERS)	2	53-54	01	00-99
207.	Innovation-Instructional Practices (INNINST)	2	55-56	01	00-99
208.	Innovation-Professional Development (INNPROD)	2	57-58	01	00-99
209.	Innovation-Buildings and Facilities (INNFACL)	2	59-60	01	00-99
210.	Innovation-Materials and Equipment (INNMATs)	2	61-62	01	00-99
211.	Innovation-Evaluation and Research (INNEVLR)	2	63-64	01	00-99
212.	Innovation-Community Relations (INNCOMMRL)	2	65-66	01	00-99
213.	Innovation-Other (INNOTHR)	2	67-68	01	00-99

	NO. OF COLUMNS	COLUMN NOS.	DATA EXAMPLE	CODE VALUE
ACTION RECOMMENDED: TERMINATION				
214. Termination-Philosophy	2		01	00-99
215. Termination-Program and Procedures				
216. Termination-Policy and Admin. Practices				
217. Termination-Personnel				
218. Termination-Instructional Practices				
219. Termination-Professional Development				
220. Termination-Buildings and Facilities				
221. Termination-Materials and Equipment				
222. Termination-Evaluation and Research				
223. Termination-Community Relations				
224. Termination-Other				
NUMBER OF RECOMMENDATION				
Rec. no. (total) (RECNO)	2	70-71	01	00-99

APPENDIX 4

CODING CATEGORIES

SHOWING FREQUENCIES FOUND IN DATA BASE

General Information

CATEGORY	FREQUENCY IN DATA BASE
1. Document ID number (ID)	N/A
2. Year document produced (YEAR)	
a. 1980	4
b. 1981	6
c. 1982	7
d. 1983	14
e. 1984	24
f. 1985	23
g. 1986	7
3. School District (SD) number (SDNO)	N/A
4. SD Size (SDSIZE)	
a. Very small (1-1000 students)	1
b. Small (1001-3000)	7
c. Medium (3001-8000)	11
d. Large (8001-15000)	4
e. Very large (15000-up)	5
5. Number of documents submitted by SD (DOCNO)	N/A

**Question 1:
How was Evaluation Defined?**

CATEGORY	FREQUENCY IN DATA BASE
6. Definition: Evaluator as judge or information provider (Implied) (DEFN)	See Table 5.2

**Question 2:
What were the Intents of the Evaluation?**

CATEGORY	FREQUENCY IN DATA BASE
Purpose (Stated or Implied)	
7-16	See Table 5.3
17. Function: Formative or Summative or both (FUNCTION)	
17	See Table 5.4

**Question 3:
Why was the Evaluation Undertaken?**

CATEGORY	FREQUENCY IN DATA BASE
Reason-Required	
18-21	See Table 5.5

Question 4:
What was the Object Evaluated?

CATEGORY	FREQUENCY IN DATA BASE
PROGRAM OR PROGRAM-RELATED PRACTICE	
22. Curriculum And Instruction (OBJCI)	
a. French	8
b. Counselling	6
c. Computers	4
d. Personal Safety	4
e. Career Preparation and Access	3
f. Science	2
g. Music	1
h. Library	2
i. Learning Assistance	1
j. Reading	1
k. Physical Education	1
l. International Baccalaureate	1
m. Police Liaison	1
n. Consumer Education	1
o. Art/Science	1
p. Continuing Adult	1
q. Instructional Innovation	1
r. Grading Practices	1
s. Other	0

CATEGORY	FREQUENCY IN DATA BASE
Testing	
23. Writing (OBJWR)	1
24. Mathematics (OBJMA)	2
25. Reading, Chemistry, Math. (Composite)	1
26. Chemistry (OBJCH)	0
27. Other (OBJOTHR)	0
28. Special Services (OBJSS)	
a. Special Programs (composite)	6
b. Native Programs (composite)	3
c. Native Program	5
d. Alternate/Rehabilitation	5
e. Gifted and Talented	4
f. Learning Disabled (including diagnostic and extended skills centres)	3
g. Behaviour Disturbed	2
h. Hospital	1
i. Residential	1
j. Multiple Handicaps	1
k. Hearing Impaired	1
l. TMH (interface with regular program)	1
m. Other	0
29. Organization for Instruction (OBJORG)	
a. Split-grade Classes	1
b. Computer Aids	1
c. Other	0
30. ORGANIZATIONAL UNIT (ORGUNIT)	
a. School	1
b. Centre	3
c. Other	0

CATEGORY	FREQUENCY IN DATA BASE
31. FACILITIES (FACILITS)	
a. Portable Classrooms	1
b. Optimum Use of Space	1
c. Other	0
ADDITIONAL CHARACTERISTICS	
Permanence	
32-33	See Table 5.7
Aspect Evaluated	
34-35	See Table 5.8
36. Object Sponsor (ORGLOCSP)	See Table 5.9
37. Object Base (ORGLOCBS)	See Table 5.9
38. Grade Level (ORGLOCGL)	See Table 5.9

Question 5:

**What Kinds of Information Regarding each Object
were Reported?**

CATEGORY	FREQUENCY IN DATA BASE
SOURCE	
39-67	See Table 5.10
NATURE	
68-78	See Tables 5.12 and 5.13

Question 6:
What Criteria were Used
to Judge the Merit and Worth of the Object?

CATEGORY	FREQUENCY IN DATA BASE
LOCATION In which sections of the report are the criteria evident? 79-88	Not Reported (see P. 145)
SOURCE 89-94	See Table 5.15
NATURE 95-112	See Table 5.15

Question 7:
What Methods of Inquiry were Used in the Evaluation?

CATEGORY	FREQUENCY IN DATA BASE
Approach Ascribed by Evaluator(s) 113-119	See Table 5.16
Type of Data Collection Techniques 120-127	See Table 5.17
No. of Data Collection Techniques 128. Number of techniques used (DATAANO)	See Table 5.18

Question 8:
To Whom Was the Report Submitted?

CATEGORY	FREQUENCY IN DATA BASE
Stated Recipient of Evaluation Report	
129-135	See Table 5.19

Question 9:
Who Were the Designated Evaluators?

CATEGORY	FREQUENCY IN DATA BASE
EVALUATOR SOURCE	
136-156	See Table 5.20
EVALUATOR NO.	
157. No. of designated evaluators (EVALNO)	See Table 5.22
ADVISORY STRUCTURE: NO. OF ADV. GPS	
158. No. of Formal Advisory Groups (ADVGPNO)	See Table 5.23
ADVISORY STRUCTURE: COMPOSITION OF GPS	
159-179	See Table 5.23
ADVISORY STRUCTURE: NO. OF MEMBERS	
180. Total Number of Members of Advisory Group(s) (ADVMEMNO)	See Table 5.23

Question 10:
What Recommendations (if any) Were Made?

CATEGORY	FREQUENCY IN DATA BASE
ACTION RECOMMENDED AND TARGET AREA OF RECOMMENDATIONS	
181-224	See Tables 5.24, 5.25, and 5.26