

THE DEVELOPMENT AND APPLICATION

OF A METHODOLOGY FOR

PROGRAM EVALUATION

by

Norman E. Gleadow,

B.Sc., University of B.C. (1967)

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Science Education

We accept this thesis as conforming to the
required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May, 1976.

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 representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Education

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date April 1976

Abstract

The purpose of this study was to develop a practical methodology of program evaluation from considerations of what the evaluative act required on logical grounds. The methodology developed was based on P.W. Taylor's (1961) analytical treatment of evaluation and M. Scriven's extensive writings on the subject.

The study involved two broad phases. The first phase was a discussion of Taylor's and Scriven's thoughts on evaluation showing where they are compatible and overlap. The second phase showed how the ideas developed in the first phase were operationalized and adapted to the evaluation of a teacher training program at the University of British Columbia.

A general conclusion of this study was that Taylor's analysis of the method and product of evaluation provided a suitable framework for the evaluation of an educational program. More specifically when Taylor's analysis was combined with Scriven's extensive work, a feasible model of evaluation resulted which readily produced a defensible, overall estimation of worth for an educational product.

Although the methodology of evaluation developed in this study was only applied to a teacher training program, it could easily be extended to the evaluation of other educational programs or products.

TABLE OF CONTENTS

| | Page |
|----------------------------|------|
| LIST OF TABLES | vi |
| LIST OF FIGURES | vii |
| ACKNOWLEDGEMENTS | viii |

Chapter

| | | |
|----|---|----|
| I | THE INTRODUCTION | 1 |
| | 1.00 Purpose of the Study | 1 |
| | 1.10 Statement of the General Problem | 1 |
| | 1.20 Definition of Terms | 3 |
| | 1.21 Educational Evaluation | 3 |
| | 1.22 Summative Evaluation | 3 |
| | 1.23 Formative Evaluation | 3 |
| | 1.30 Statement of the Specific Problem | 3 |
| | 1.40 Overview of the Study | 4 |
| | 1.50 Delimitation of the Study | 5 |
| II | THE THEORETICAL FRAMEWORK FOR AN EVALUATION MODEL | 7 |
| | 2.00 Introduction | 7 |
| | 2.10 Present Difficulties in Evaluation | 7 |
| | 2.11 Tyler's Approach to Evaluation | 10 |
| | 2.12 The Accreditation Model | 11 |
| | 2.13 Stufflebeam's Context, Input, Process and Product (CIPP) Model | 12 |
| | 2.14 Stake's Model | 15 |
| | 2.15 Summary | 16 |
| | 2.20 Evaluation: The Theoretical Framework | 17 |
| | 2.30 The Preconditions | 19 |
| | 2.31 The Evaluatum | 19 |
| | 2.32 Point of View | 19 |
| | 2.33 Grading or Ranking | 21 |
| | 2.40 Level I of S-Grading: Adoption of Standards | 28 |
| | 2.41 Intrinsic Value | 28 |
| | 2.42 Extrinsic Value | 28 |
| | 2.50 Level II of S-Grading: Operational Clarification of the Standards in Grading | 31 |
| | 2.60 Level III: Specification of the Class of Comparison in Grading | 32 |

Table of Contents (Cont.)

| | Page |
|--|------|
| 2.70 Level IV: Determination of the Good and Bad Characteristics of the Evaluatum in S-Grading | 33 |
| 2.71 Gathering Data on the Characteristics | 33 |
| 2.72 Determining if the Characteristics are Good or Bad | 34 |
| 2.80 Level V: The Overall Judgement of Worth | 36 |
| III APPLICATION OF THE MODEL TO THE DESIGN OF AN ACTUAL EVALUATION STUDY | |
| 3.00 Introduction | 38 |
| 3.10 Establishing the Preconditions | 38 |
| 3.11 The Evaluatum | 38 |
| 3.111 Letter to Potential Students for Program B | 39 |
| 3.112 Program Description | 39 |
| 3.113 Description of the Offered Courses in Program B | 40 |
| 3.12 The Point of View | 41 |
| 3.13 Grading | 43 |
| 3.20 Adoption of Standards | 44 |
| 3.21 Standard 1: The Worth of the Program Goals | 45 |
| 3.22 Standard 2: Goal Achievement | 48 |
| 3.30 Operational Clarification of Standards | 49 |
| 3.31 Standard 1: The Worth of Program B's Goals | 49 |
| 3.32 Standard 2: The Degree of Goal Achievement | 52 |
| 3.33 The Checklist for Determining Good Achievement | 54 |
| 3.331 Checkpoint 1: Market | 54 |
| 3.332 Checkpoint 2: True Field Trials | 55 |
| 3.333 Checkpoint 3: True Consumer | 57 |
| 3.334 Checkpoint 4: Long Term | 59 |
| 3.335 Checkpoint 5: Side Effects | 59 |
| 3.336 Checkpoint 6: Process | 63 |
| 3.337 Checkpoint 7: Causation | 64 |
| 3.338 Checkpoint 8: Statistical Significance | 68 |
| 3.40 Specification of the Class of Comparison in Grading | 70 |
| 3.50 Determining the Good and Bad Characteristics of the Evaluatum in S-Grading | 71 |
| 3.51 Gathering Data on the Goal Achievement Characteristics | 71 |
| 3.52 Determining if the Characteristics of the Program are Good or Bad | 75 |
| 3.521 Rating the Worth of the Goals | 75 |

Table of Contents (Cont.)

| | Page |
|---|------------|
| 3.522 Goal Achievement and its Rating | 78 |
| 3.60 Overall Determination of Worth | 81 |
| 3.70 Extension of the Model | 85 |
| 3.71 Introduction | 85 |
| 3.72 The Standards to be Used in Ranking | 86 |
| 3.721 Costs | 87 |
| 3.722 Extended Support | 88 |
| 3.73 The Precedence of the Standards in Ranking | 89 |
| 3.74 Ranking the Evaluata | 92 |
| IV DATA COLLECTION TECHNIQUES IN THE APPLICATION OF THE EVALUATION MODEL | 93 |
| 4.00 Introduction | 93 |
| 4.10 Initial Stage of Data Collection | 93 |
| 4.20 Second Stage of Data Collection | 96 |
| 4.21 Evaluation of U.B.C. Instructors' Teaching Performance | 97 |
| 4.22 Questionnaire to Student Teachers or their Alternate Placement | 98 |
| 4.23 Interview of Student Teachers | 98 |
| 4.24 Final Questionnaire for the Student Teachers | 98 |
| 4.25 Observation of Student Teaching | 99 |
| 4.26 Side Effects of Program B | 100 |
| 4.27 Interview with Program B's Director | 101 |
| 4.28 Questionnaire for School Pupils | 101 |
| 4.29 Interview with U.B.C. Administrators | 101 |
| 4.210 Other Information Sources | 102 |
| 4.30 Practical Procedures for the Organization of Data | 105 |
| 4.40 Confidentiality of Data | 108 |
| V SUMMARY AND CONCLUSIONS | 109 |
| 5.00 Summary | 109 |
| 5.10 Conclusions | 113 |
| REFERENCES | 118a-e |
| APPENDICES | 119 |
| A. Questionnaire for the Program Developers, Subject: Evaluation Priorities | 119 |
| B. Interview of School Administration | 125 |

Table of Contents (Cont.)

| | Page |
|---|------|
| C. Likert Attitude Scale (to measure general attitudes toward the teacher training process of the student teachers) | 127 |
| D. Reaction Form. | 130 |
| E. Evaluation of U.B.C. Instructors' Teaching Performance: Form A | 132 |
| F. Questionnaire to Student Teachers on Their Alternate Placement | 135 |
| G. Evaluation of U.B.C. Instructors' Teaching Performance: Form B | 139 |
| H. Interview of Student Teachers | 142 |
| I. Final Questionnaire for Student Teachers | 145 |
| J. Discipline Techniques | 148 |
| K. Side Effects of Program B (Teacher's Questionnaire) | 150 |
| L. Interview with Program B's Director | 154 |
| M. Questionnaire for School Pupils | 156 |
| N. Interview with U.B.C. Administrators | 160 |
| O. Student Teacher's Questionnaire (Side Effects) | 162 |
| P. U.B.C. Staff Questionnaire on Courses | 167 |
| Q. Student Teachers' Questionnaire on the Evaluation of U.B.C. Instructors Teaching Performance | 170 |
| R. Teachers' Questionnaire | 177 |
| S. Interview with Director of Elementary Education at School Board Offices | 187 |
| T. Proposal for the Evaluation of Student Teachers Teaching Effectiveness | 189 |
| U. Summary of Evaluation Steps in Ranking | 203 |

LIST OF TABLES

| Table | Page |
|---|---------|
| 1 Taxonomy of Evaluation Types | 25 |
| 2 Calendar of Data Collecting Events | 72-74 |
| 3 Data Sources for the Goal Achievement Checkpoints . . | 103 |
| 4 Comparison of Evaluation Procedures | 113-114 |

LIST OF FIGURES

| Figure | Page |
|--|------|
| 1 Developing Evaluation Designs | 14 |
| 2 An Outline of the Logical Process | 27 |
| 3 The Standards and their Ratings | 69 |
| 4 Determining a Rating for Goal Achievement | 79 |
| 5 Product Evaluation Profile for Program B | 80 |
| 6 Comparing Goal Worth and Goal Achievement Rating | 82 |
| 7 Expansion of Figure 6 | 84 |
| 8 Combining the Goal Achievement and Cost Ratings | 91 |
| 9 Format for Organization of Data | 105 |
| 10 Example of Format for Organization of Data | 107 |
| 11 Steps in S-Grading in Formative Evaluation | 112 |

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Professor Walter B. Boldt, who introduced me to the field of educational evaluation, and whose scholarship and intellectual integrity were a constant inspiration.

DEDICATION

To the memory of my father

Frank R. Gleadow

CHAPTER I

The Introduction

1.00 Purpose of the Study

Many present evaluation models are unsatisfactory on theoretical grounds because they lack clarity as to the nature of the evaluative act. The purpose of the present study is to shed some light on this problem by developing a practical model of evaluation based on an analysis of theoretical ideas about evaluation procedures and to demonstrate how the resulting model could be used in a particular case. The problem is important because much of what is currently called evaluation fails to fulfill the evaluation functions to be served, on theoretical grounds.

1.10 Statement of the General Problem

"One can be against evaluation only if one can show that it is improper to seek an answer to questions about the merit of educational instruments, which would involve showing that there are no legitimate activities (roles) in which these questions can be raised." (Scriven, 1967, p. 41)

The general problem to be investigated in the study is to determine the nature of the procedures of evaluation on theoretical grounds and to attempt to specify the evaluative functions to be served in practical terms through illustration and application to the evaluation of a teacher-training program.

Evaluation as practised for the past few years is not

succeeding in its perceived function. Evans (1974, pp. 7, 11, 12) indicates that though evaluation is ". . . now all the rage . . ." it faces many new problems. The problems revolve around difficulties of collecting data, unrealistic expectations on the part of policymakers, unwillingness of people to be evaluated and so on. Guba (1969) indicts evaluation for its failure to find any differences among programs, even when there are obvious differences, and suggests that the whole technique of evaluation should be questioned and examined. Bloom (1969, p. 46) who equates evaluation with testing and measuring, writes of the potentially destructive aspects of evaluation when teachers and students believe the procedures to be unfair (p. 45). In a recent paper, Glass (1975) criticises evaluation as taking weakly defined stances when threatened and then reacting to this inherent weakness with either a humanistic approach of gentle self-evaluation; or by regarding "resistances" to evaluation as mere technical problems overcome by technical means; or by retreating into a vast gray area between.

These are some of the criticisms which have been leveled at evaluation. They raise two fundamental questions: what is evaluation; and how can evaluation be carried out? Many evaluation models deal with the first question by posing another which is generally of the form; "what is it that we are doing today that we call evaluation?" From the answer to this question is distilled a model which is then used for evaluation. In other words the model is a clarification of present practise, and not necessarily evaluation. The second part of the question is usually answered

in a similar way and embroils researchers in the evaluation vs. research imbroglia (Provus, 1970: Hemphill 1969, p. 190; Welch 1969, p. 440). Therefore to work from present practise is not a satisfactory method of defining evaluation and will not be used in this thesis except where present practise is the same as the theoretical position developed in Chapter 2.

1.20 Definition of Terms

1.21 Educational Evaluation: Educational Evaluation is the procedure of justifying a value claim about the merit or worth of an educational product (course, curriculum, etc.) The result of this procedure is a conclusion of the worth of the product.

1.22 Summative Evaluation: This is an overall evaluation of a completed educational product and serves the needs of the client, and more importantly, the ultimate consumer(s) of the product.

1.23 Formative Evaluation: This type of evaluation takes place at intermediate points in the development of a product and serves the needs of the developers for modifying and improving the product.

1.30 Statement of the Specific Problem

The specific problem to be dealt with in this thesis is to

develop a practical methodology of evaluation based on the logical considerations presented by Taylor (1961) and Scriven (1967) and the evaluative functions to be served, in the form of the checklist for the evaluation of products, producers and proposals proposed by Scriven (1974b); and to apply the results to the evaluation of a teacher-training program at the University of B.C.

The specific problem can be further broken down to a number of sub-problems for the purposes of the study:

- a) An analysis of the evaluation process based on Taylor's philosophical position (Chapter II).
- b) Establishing the evaluation functions to be served and the validity of Scriven's Checklist.
- c) Adaption of Scriven's Checklist to the evaluation of a teacher-training program (Chapter III).
- d) Development of instruments and techniques for collecting the necessary evaluation data. (Chapters III and IV).
- e) Making a final evaluative claim about the teacher-training program (Chapter III).
- f) Critique of the effectiveness of the checklist for this purpose (Chapter V).

1.40 Overview of the Study

The thesis comprises the development and illustration of a methodology of evaluation based on P.W. Taylor's (1961) analytical treatment of evaluation and M. Scriven's (1967, 1974b) extensive

operationalization of a compatible model of evaluation. There are three broad phases in this process. The first phase involves a discussion of P.W. Taylor's and Scriven's thoughts on evaluation showing where they are compatible and overlap. This is done in Chapter II.

The second phase shows the method by which the ideas developed in Chapter II were operationalized and adapted to a particular evaluation project; that is the evaluation of a program in teacher training. This was done in Chapter III. Chapter III also illustrates how the developed model can be extended.

The third phase is a discussion of the data collection techniques used in applying the evaluation model developed. Chapter IV illustrates those techniques. Chapter V summarizes the previous four chapters and makes final recommendations on the overall usefulness of the scheme.

1.50 Delimitation of the Study

The Evaluation technique is developed around the evaluation of an educational product. Therefore the extension of the scheme to other areas, such as teaching effectiveness, has not been undertaken in the present study.

The evaluation applied was a formative evaluation. The overall statement of worth arrived at in Chapter III, for the particular program being evaluated, represents an interim conclusion and not a final conclusion on the worth of the program. Data were collected

representing good performance and bad performance. To present a balanced picture both types of data should be revealed; however, one of the functions of a formative evaluation is to provide information, in a summarized and judged (as to validity and worth) format so that program changes can be made to better fulfill the program goals. To reveal only the good performance data and not the bad would be misleading; but to also present the bad would not contribute to the evaluation model being developed and could violate confidentiality. Therefore the actual data have not been included in this study.

Finally, the instruments developed for the evaluation of the program being evaluated are not generalizable. Since this is discussed in more detail in Chapter IV, it will only be briefly noted here that the questionnaires are specific for the particular program.

CHAPTER II

The Theoretical Framework for an Evaluation Model

2.00 Introduction

Chapter Two explicates the theoretical framework for evaluation. The first part of the chapter briefly examines a number of commonly used approaches to educational evaluation, and points out the shortcomings of those approaches. The second part of the Chapter combines Taylor's (1961) logical examination of evaluative procedures with Scriven's very practical descriptions of evaluation. The resulting amalgam provides the structure of evaluation underlying this study.

2.10 Present Difficulties in Evaluation

Many of the presently used models of evaluation have been developed from an experience base. Taylor and Maguire (1966, p. 12) wrote:

"In many ways, the efficiency of evaluation has been hampered by a lack of clear guiding principles indicating what evaluation is potentially able to accomplish. What principles there are, exist mainly as experientially induced procedures and techniques held by the few who work in the area."

The danger in developing evaluation models from experience is that the meaning of "evaluation" is never examined. The result is a broad range of models which purport to be doing the same thing, i.e. evaluating, but when examined are really doing something very different such as describing, or data gathering, or measuring or

simply rubber stamping. These may all be procedures in evaluation, but not one of them can singly be labelled as evaluation.

The weakness of experientially based evaluation models, and the lack of understanding of the meaning of evaluation has resulted in a great deal of criticism of "evaluation" as presently applied. On this point Guba (1969, p. 29) noted that:

"The traditional methods of evaluation have failed educators in their attempts to assess the impact of innovations in operating systems. Indeed, the evidence produced . . . has contradicted the experiential evidence of the practitioner. Innovations have persisted in education not because of the supporting evidence of evaluation but despite it."

This indictment is due to the narrowness with which some practitioners define evaluation. Guba is suggesting that many evaluations are not complete or comprehensive, and they tend to be limited to variables which can be "scientifically" measured and statistically treated.

Another criticism is ". . . one of the primary latent functions of many evaluation studies [is] a legitimization process for predetermined policies." (House, 1973, p. 80).

House's criticism points out one of the most serious omissions in many "evaluations"; that is, failing to determine whether the goals of the program are worthwhile. The acceptance of program goals as being prima facie worthwhile goals is a major weakness in many evaluations. Instead of legitimizing predetermined policies, the evaluator should be determining if the predetermined policies are legitimate.

The unclear nature of evaluation procedures, and the linking of evaluation with criticism has resulted in defensive posturing by those being evaluated and by the evaluators. Stake (1967, p. 524) wrote that the educator's disdain of evaluation was due to his sensitivity to criticism. The educator uses smoke screens such as "national norm comparisons" and "academic freedom" to avoid exposure through evaluation. Provus (1970, p. 54) suggested that the educational administration has contributed to the failure of educational evaluation by, among other things, not explicating his value assumptions, or by not setting performance standards; but goes on to say that part of the reason is the unclear nature of analysis, research and evaluation in the problem solving process.

Glass (1975, p. 10) strongly criticizes the evaluator for retreating to three "straw men" tenets when under attack. The tenets are identified as being the sanctity of science, the public's right to know and all feedback is beneficial. In spite of his criticism, Glass says (ibid) ". . . evaluation can find a stronger faith, even if the scope of its application is somewhat reduced."

The concept of evaluation has been used to describe many different procedures. Most of the present models of evaluation are derived from concepts of educational measurement. They avoid putting value judgements on data or products by using rigorous research design which supposedly is "value free", or by bundling-up the evidence and giving it to someone else to judge. Brief descriptions of the most common models of evaluation follow in subsequent

sections of this chapter.

2.11 Tyler's Approach to Evaluation

Tyler first proposed his model for the evaluation of learning experiences during the 1930's. It evolved from his participation in the, "Eight-year Study" (Aiken, 1942) and was originally published in 1949. His model involves the following six steps (Tyler, 1942):

- 1) Establish the broad goals or objectives of the program.
- 2) Classify the objectives.
- 3) Define the objectives in behavioral terms.
- 4) Suggest situations in which achievement of the objectives can be shown.
- 5) Develop or select measurement techniques, and,
- 6) Gather student performance data and compare it with the behaviorally-stated objectives.

Approaches similar to Tyler's are found in Metfessel's and Michael's (1967), "A Paradigm Involving Multiple Criterion Measures for Evaluating the Effectiveness of School Programs", and W. James Popham's (1972) An Evaluation Guidebook.

All of these methods would have the evaluator looking almost exclusively at the objectives of the program. The bases of the objectives are not the focal point in the evaluation. Popham (1972), in his guidebook, does mention a needs assessment, but it really is mainly a resources inventory and clarification of predetermined needs,

goals or objectives.

Welch used the Tylerian approach when he evaluated a course called "Physical Science for the non-science Student." In his evaluation he was attempting to determine if the objectives of the course had been achieved. He calls this a summative evaluation (Welch, p. 140, 1972). But he omits determining whether the course objectives are worthwhile goals, and hence, whether the program has merit.

2.12 The Accreditation Model

The following quote is a typical definition of accreditation:

"Accreditation, as applied in education, is the recognition accorded to an institution that meets the standards or criteria established by a competent agency or association. Its general purpose is to promote and insure high quality in educational programs." (U.S. Dept. of Health, Education and Welfare, 1959, p. 3)

The article (1959, p. 4) goes on to outline four steps in fully developing accrediting procedures. They are:

- 1) Establishment of standards or criteria.
- 2) Inspection of institutions by competent authorities to determine whether they meet the established standards or criteria.
- 3) Publication of a list of institutions that meet the standards or criteria.
- 4) Periodic reviews to ascertain whether accredited institutions continue to meet the standards or criteria.

Accreditation appears to be very similar to evaluation,

particularly the first two steps above. However, the invalid assumption made in accreditation is that the fulfillment of the standards and criteria in Step 1 necessarily benefit the learner in the institution. As Glass rightly points out:

"The accreditation criteria reflect the interests of administrators; attention is given to the processes or means of education and is opposed to its consequences on learners." (Glass, 1969, p. 20)

2.13 Stufflebeam's Context, Input, Process and Product (CIPP)

Model

In this model, evaluation is defined as,

". . . the provision of information through formal means, such as criteria, measurement, and statistics, to serve as rational bases for making judgements in decision situations." (Stufflebeam, 1958, p. 6)

In other words, the evaluator is essentially an appendage to the decision-maker. This definition is operationalized through the CIPP model, about which Worthen (1968, p. 3) writes:

"Deceptively simple when viewed in outline form, the structure consists of a list of no less than twenty-two decision situations which are common to most evaluation designs. Upon even cursory inspection, however, it is apparent that the structure implies much more than a simple list of twenty-two items which evaluators must keep in mind, for at each of the decision points identified in the structure, the evaluator needs to make a choice among the available alternatives."

Stufflebeam identifies four classes of decisions: planning, programming, implementing, and re-cycling; and suggests four

evaluation strategies for those decisions. The strategies are:

- a) Context evaluation: which identifies the goals and needs to be fulfilled by the program, and the environment where the program will operate.
- b) Input evaluation: which identifies and assesses alternative procedures and designs for attaining program objectives.
- c) Process evaluation: which monitors the on-going program for problems or deviations from program design.
- d) Product evaluation: which is summative evaluation, and determines the extent to which the goals and objectives have been met.

Each strategy is then sub-divided into twenty-two categories--the same for each strategy--to give the evaluation design. This is illustrated in Figure 1 (Worthen 1968, p. 4).

There are two objections to this model. The first is a practical consideration. There could be as many as 88 decisions situations (four classes of decisions with twenty-two decision situations in each class) which seemingly would take an army of data gatherers to fulfill.

The second objection is that use of the CIPP model would fail to provide an overall evaluation. The evaluator using the CIPP model would provide information to decision makers, but would make no decision on the worth of that information or the worth of the product being evaluated. As Glass (1969, p. 36) wrote:

"Being of assistance to the program personnel--so they may better conduct their business--is a proximate aim of evaluation; the ultimate aim of an evaluation is to decide questions of worth."

Figure 1

Developing Evaluation Designs

The logical structure of evaluation design is the same for all types of evaluation, whether context, input, process or product evaluation. The parts, briefly, are as follows:

- A. Focusing the Evaluation
 1. Identify the major level(s) of decision-making to be served, e.g., local, state, or national.
 2. For each level of decision-making, project the decision situations to be served and describe each one in terms of its locus, focus, timing, and composition of alternatives.
 3. Define criteria for each decision situation by specifying variables for measurement and standards for use in the judgment of alternatives.
 4. Define policies within which the evaluation must operate.
- B. Collection of Information
 1. Specify the source of the information to be collected.
 2. Specify the instruments and methods for collecting the needed information.
 3. Specify the sampling procedure to be employed.
 4. Specify the conditions and schedule for information collection.
- C. Organization of Information
 1. Specify a format for the information which is to be collected.
 2. Specify a means for coding, organizing, storing, and retrieving information.
- D. Analysis of Information
 1. Specify the analytical procedures to be employed.
 2. Specify a means for performing the analysis.
- E. Reporting of Information
 1. Define the audiences for the evaluation reports.
 2. Specify means for providing information to the audiences.
 3. Specify the format for evaluation reports and/or reporting sessions.
 4. Schedule the reporting of information.
- F. Administration of the evaluation
 1. Summarize the evaluation schedule.
 2. Define staff and resource requirements and plans for meeting these requirements.
 3. Specify means for meeting policy requirements for conduct of the evaluation.
 4. Evaluate the potential of the information design for providing information which is valid, reliable, credible, timely, and pervasive.
 5. Specify and schedule means for periodic updating of the evaluation design.
 6. Provide a budget for the total evaluation program.

It is precisely the question of worth that evaluators using the CIPP model avoid.

Provus (1969) and Alkin's (UCLA, 1968) models both closely resemble Stufflebeam's CIPP model.

Scriven (1972) strongly criticized the CIPP model. He stated that what is called evaluation in the CIPP model is not evaluation at all. Context evaluation is essentially a market-survey, input evaluation is a survey of resource options, process evaluation is usually social monitoring and bookkeeping, and product evaluation appears to be a dilute mixture of summative and formative evaluation --claiming to do both, but actually, not doing either (Ibid, p. 134-135). Scriven's most general criticism of the CIPP model is (Ibid, p. 36):

"[the CIPP model is] about the most complicated and confusing way of analysing the practical procedures of evaluation that I can imagine, and it's certainly the most complicated one that I've ever seen."

2.14 Stake's Model

In 1967 Stake proposed his evaluation model (Stake, 1967). The model focussed on the description and judgement of educational programs. The descriptive part is divided into "intents" (the goals and/or objectives) and observations (what the evaluator learns through empirical measures). The judgement matrix is also divided into two parts--standards and judgements. The standards can either be absolute (i.e. criterion levels) or relative (i.e. compared to other, similar programs), but must pertain to the thing being

evaluated. The judgements are the assigning of weights to the standards and judging the merit of the product under consideration.

Stake also says that the program should have a rationale; but it is not clear from his paper (The Countenance of Educational Evaluation, Stake, 1967), whether the evaluator makes any judgements of the rationale. He says "The evaluator asks himself or other judges whether the plan developed by the educator constitutes a logical step in the implementation of the basic purposes." (Stake, 1967, p. 13). However, this is not a judgement of the worth of the program goals as they have been outlined in the program's rationale--rather, it is a judgement of the logic of the implementation plans.

In an application of his model Stake (1971, p. 4) writes:

"Evaluators have an obligation to raise the questions, 'Were the right goals pursued?' Different people have different ideas, of course, as to what the right goals are. Still, goals and priorities should be evaluated."

He goes on to say, about the particular project he is evaluating "In the eyes of the evaluators (my emphasis), the Tcity goals are worthy goals, suitably discussed and reasonably operationalized." (Stake, 1971, p. 4) However, the logical connection between the goals and worth of the goals has still not been explicated, for one can still ask, "The goals are worthwhile in his eyes, but are they really worthwhile?"

2.15 Summary

The problem of deciding what one should do when one

evaluates springs from an imprecise idea of what the word "evaluation" means. This has caused many of the shortcomings of the models described above. Unfortunately, in none of the recent compilations of evaluation theory (Popham; 1974, Worthen & Sanders, 1973; Tyler, 1969) is there a serious effort to deal with this problem. Therefore the rest of this Chapter will be an attempt to provide a logical base for the meaning and use of the word "evaluation".

2.20 Evaluation: The Theoretical Framework

"The process of evaluation consists in trying to determine the value of something. As a product or outcome of that process, an evaluation is a settled opinion that something has a certain value." (Taylor, 1961, p. 3)

Taylor's views are paralleled by Scriven (1967), whose contribution is to provide a methodological framework for the analytical treatment. Scriven considers that the goal of evaluation is to determine the merit or worth of some enterprise, and suggests we can attain that goal by gathering and combining needs related performance data, weighing and combining that data and justifying our conclusion of worth by showing that our data-gathering instruments and selection of criteria are valid (Scriven, 1967, p. 40-41).

Taylor (1961, p. 9, 10) has identified five steps in evaluation. To these five steps one could add a preliminary set of preconditions to clarify the five steps which follow. Therefore an

evaluation on logical grounds requires:

1) The Preconditions:

- a) Identify the evaluatum (what it is that is being evaluated).
- b) Decide on whether you are going to:
 - (i) Grade according to standards (S-Grading) (that is evaluating in terms of some absolute standards);
 - (ii) Rank (that is S-Grade and then compare and rank order the evaluatum in a comparison group of similar evaluata which have also been S-Graded);
 - (iii) Grade according to rules (R-Grading) (that is grade the evaluatum as obeying or not obeying a set of applicable rules)

c) Determine the Point of View or Points of View.

2) Adoption of a standard or rule, or set of standards or rules, for evaluating the evaluatum.

3) Operational clarification of the standards or rules.

4) Specification of the class of comparison.

5) Determining the good or bad characteristics of the evaluatum.

6) Deducing, from (3) and (5), the degree to which the evaluatum on the whole fulfills or fails to fulfill the standards; or, in the case of rules, the relative importance of the rules the evaluatum passes or fails (A rule can either be obeyed or violated.

Therefore there are no degrees of compliance or violation.)

These six steps will not be discussed in more detail.

2.30 The Preconditions to an Evaluation

If an evaluator is to perform an evaluation, then three preconditions must be satisfied. The first of these is that there must be something to evaluate (i.e., an evaluatum); the second is that the evaluator must adopt a point of view, and the third is that the evaluator must decide if he is going to grade or rank according to rules or standards.

2.31 The Evaluatum

That which is to be evaluated has been labelled by Taylor as the evaluatum (Taylor, p. 4, p. 23) and is anything which can be ranked or graded (discussed later in this chapter), or which has bad and/or good-making characteristics, to be determined by the evaluator. For example, the evaluatum for this thesis is a teacher-training program at U.B.C.

2.32 Point of View

Taylor says that:

"Taking a certain point of view is nothing but adopting certain canons of reasoning as the framework within which value judgements are to be justified; the canons of reasoning define the point of view." (Taylor, 1961, p. 109)

and

"This means precisely that the judge or evaluator has adopted a set of rules of relevance and valid inference that recognize only certain reasons as relevant and good." (Taylor, 1961, p. 109)

The above explanations were exemplified by Taylor as follows:

"We cannot evaluate a car as good or bad, or as better or worse than another car, unless we know what point of view we are to take . . . Is it to be evaluated as a piece of mechanical engineering? As an object of aesthetic contemplation? . . . Each point of view we take makes a difference, since different standards (norms) are appropriate to different points of view." (Taylor, 1961, p. 5, 6)

It would be unusual to be able to decide the overall worth of any program by evaluating it from only one point of view. Even in Taylor's example above, a careful shopper would not buy an automobile only on the basis of mechanical engineering, or only on the basis of aesthetic contemplation. Nor is it reasonable to compare teacher-training programs only on the basis of their organization, or only on the basis of their relative costs. All reasonable points of view must be taken into consideration.

Taylor has identified eight basic points of view in a civilized culture; they are: the moral, the aesthetic, the scientific, the religious, the economic, the political, the legal, and the point of view of etiquette or custom. He derives these points of view from the eight corresponding major institutions and activities plus the educational institutions; but as he points out:

"There is no single point of view corresponding to the educational institutions of a society, since education is a process which may take place within any point of view. Thus there is moral education, aesthetic education, intellectual education, religious education, and so on." (1961, p. 300)

Taylor's reasoning does not sound the death knell for the evaluation of educational institutions. It simply alerts the evaluator to the multiplicity of points of view which must be at least considered in the evaluation. For example, is the teacher-training program immoral (moral point of view), crass (aesthetic point of view), illegal (legal point of view), expensive (economic point of view), and so on.

2.33 Grading or Ranking

The third pre-condition is to decide what form the evaluation will take. There are three possibilities: grading according to standards, grading according to rules, and ranking according to standards. Daniels (1971, p. 6) gives a particularly lucid account of these three forms, so his description is reproduced in its entirety below:

"Form 1 Grading according to standards. (Let us call this S-grading.)

Form 2 Grading according to rules. (Let us call this R-grading.)

Form 3 Ranking.

Before we see how these three forms, S-grading, R-grading and Ranking, could be used to evaluate curricula, let us look at each in a somewhat simpler context.

S-grading: Suppose we are judging cars for their comfort. It makes sense to say that Car 1 is uncomfortable, Car 2 is fairly comfortable, Car 3 is comfortable, Car 4 is very comfortable and Car 5 is extremely comfortable. This shows us the two main features of S-grading: a) The criterion we use as the basis for our evaluation is a standard. b) Things can fulfill standards in differing degrees. Thus, we can evaluate things on a simple two-part scale (e.g., comfortable or uncomfortable) or on a multiple-level scale (as in our example of the five cars).

R-grading: Suppose, however, that we are concerned to discover whether or not the headlights on our cars are correctly adjusted as set down by a law. Now we are evaluating according to rules and the notion of degrees is not applicable; either the headlights conform with the rules or they do not; they are right or wrong, correct or incorrect.

Ranking: In order to rank things we must go through two steps: a) we first of all S-grade the things to be ranked, as we did for instance with the five cars. b) We then compare the degree to which the things fulfill the standard (R-grading cannot be used for Ranking), and rank them. Thus, in ranking our cars from the point of view of comfort, we might say something like the following: Car 1 - worst, Car 3 - average, Car 5 - best."

It is not unreasonable to suggest that any educational product can undergo all three forms, R-grading, S-grading and Ranking, in its evolution from an initial inspiration to the finished product. For example, suppose a group of professors wished to initiate a teacher-training program based on a particular model of teacher training. Firstly, they would have to determine, among other things, if their proposed program (the evaluatum) was feasible in terms of the university. In other words, would the students be taught the required number of courses (specified number of units); do the prospective students have the required minimum entrance qualifications (for this particular year of university study); will the professors' teaching responsibilities be at least the minimum required? These are rules which are either fulfilled or not fulfilled. In the example given, this first step, though usually informally done, could be interpreted as R-GRADING.

Once the program has been through informal R-grading, it is operationalized and is no longer a proposal. The program is in a developmental period or formative stage. The main concern of the program proponents is whether or not the standards they set for the program are being met. They ask: "Is the operational program the same as that program we originally conceptualized?" The standards to be used in the evaluation come from this version. Various character-

istics of the program are then graded according to those standards. Stake (1967) has proposed this method when he discusses judging the characteristics of a program with respect to absolute standards. He says that each set of absolute standards, when formalized, would indicate acceptable and meritorious levels of performance for each characteristic. This formalization is illustrated in Chapter III of this thesis.

The process of determining whether a program is good or bad with respect to absolute standards is S-GRADING.

The final stage in the program's evaluation, is to determine how it compares to the other, similar teacher-training programs available (i.e. to the members of the class of comparison). There are two criteria which must be taken into account to decide whether the evaluatum is better or worse than something else, or is the best or worst in the class of comparisons. They are:

" . . . whether the evaluatum is good on the whole, or bad on the whole, when graded according to the given standards, and whether its overall goodness or badness outweighs the overall goodness or badness of each member of the class of comparison." (Taylor, 1961, p. 15, 16)

Of course, each member of the class of comparison has also been graded according to the same set of standards.

Therefore, an evaluatum can be graded as "bad" on the whole, yet be better than any of the members of the class of comparisons, or the best of a bad lot. Or, similarly, it could be graded as good, and ranked as average, etc.

The process of determining the comparative value of a program is called RANKING.

Scriven has argued that we should determine the superiority or inferiority of the evaluatum to the competition before we have completely evaluated the evaluatum. (Scriven, 1967) In other words we must have gone through a process of RANKING. In a later article he presented a taxonomy of different evaluative conclusions, (Scriven, 1974a, p. 146) presented here as Table 1.

In practise, a particular evaluative conclusion is aimed for (or arrived at) somewhere between a "Preevaluative," and a "Best-Buy" evaluative conclusion. Anything less than a "Best-Buy evaluative conclusion", constitutes less than a complete evaluation; therefore it is important that the client for the evaluation knows exactly what sort of evaluation he is getting, and how much it falls short of a complete evaluation (if, indeed, it does).

An evaluative conclusion of "Commendatory" in Scriven's taxonomy, is the highest level possible using S-grading. The control group implied in the, "usually adequate premises" of Table 1, is the ideal form of the evaluatum. It is against the standards of that ideal that the evaluatum is graded.

The last three evaluative conclusions: Laudatory, Ideal, and Best-Buy, involve ranking the evaluata, or making overall comparative statements. Therefore, these three conclusions illustrate increasingly comprehensive RANKING.

This partitioning of Scriven's taxonomy does not exclude the first four evaluative conclusions (Preevaluative, Minimal evaluative, Overall evaluative and Commendatory) from the process of RANKING. If

TABLE 1
Taxonomy of Evaluation Types

| <u>Evaluation Type</u> | <u>Usual Verbal Expression</u> | <u>Usually Adequate Premises</u> |
|--|---|---|
| Preevaluative (Goal or criterion achievement) | "The treatment X had the effect Y on the population of students, S, in conditions C; and Y was the goal or shows that the goal was achieved." | 1. X was the treatment. 2. X caused Y. 3. Y implies that the goal was achieved. |
| Minimal evaluative | "X had a good effect (on S in C)." | 1. X caused Y. 2. S or non-Ss (desired, ¹ enjoyed, were benefited by) Y. |
| Overall evaluative | "X had an overall good effect." | 1. As for minimal, plus, 2. Y had no harmful effects on Ss or non-Ss, or, 3. Y had much less significant harmful effects on Ss. |
| Commendatory | "X was worth doing." This is almost a sub-case of overall evaluation, if costs are taken as a harmful effect. | 1. As for overall, plus, 2. The cost of X was manageable. 3. Y was worth the cost. |
| Laudatory | "X was the best choice." | 1. As for commendatory, plus, 2. No other treatment, on data which was available, appeared as cost-effective. |
| Ideal | "X was the best possible treatment." | 1. As for commendatory, plus, 2. No other treatment was in fact as cost-effective. |
| Best-Buy | "X was a Best-Buy." | 1. As for commendatory, plus, 2. X is a member of a group which offers the best or almost the best performance for significantly less cost than their performance peers. |

¹ These premises give prima facie support for evaluative conclusions, not deductive support, but evaluation--like science--only needs prima facie inference.

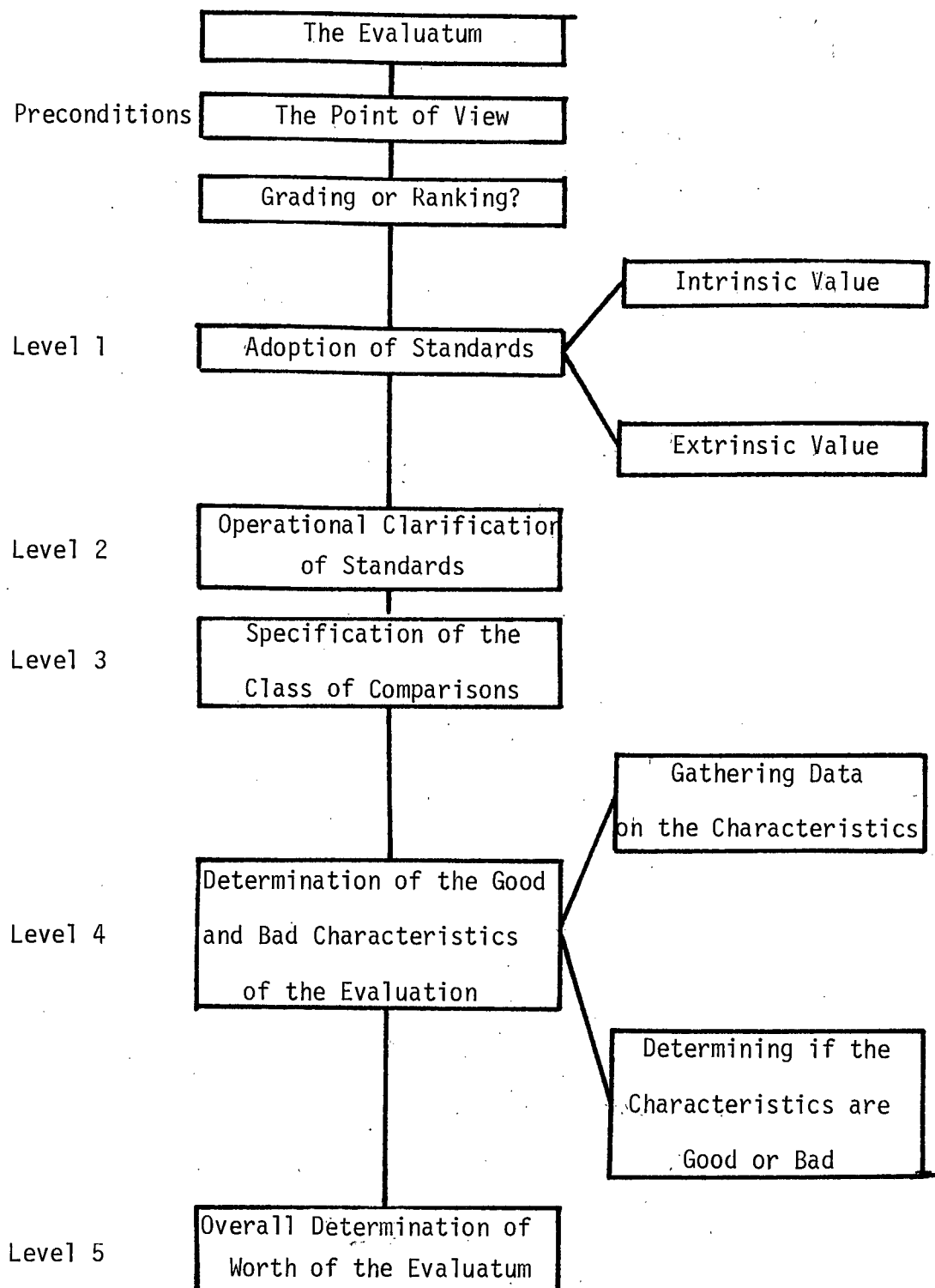
the evaluator is only RANKING, then he would base his rank order of the various evaluata on an increasingly comprehensive list of standards, as he goes from a preevaluative to a Best-Buy conclusion.

An internal, formative evaluator (i.e. the early developmental stage of the project is being evaluated by someone involved in, and committed to the project (Scriven, 1967, p. 45) is usually concerned with the early stages of S-GRADING. In other words, he is grading the evaluatum on the basis of a series of standards derived from some ideal, or best possible, form of the evaluatum. On the other hand, if the evaluator is an external, summative evaluator (i.e. the final marketed version is being evaluated by someone external to the project), then he is ranking the evaluatum in the class of comparison. There is an intermediate step near the end of the formative stage of the evaluation. At this time, an external formative evaluator could be engaged to perform the last stages of S-grading and early stages of ranking before the evaluatum went to the summative evaluation.

The particular problem examined by this thesis was an evaluation of a teacher-training program. This program was in the developmental stage, and the author of this thesis was primarily an internal, formative evaluator. Therefore, the form of the evaluation was S-GRADING, and aimed at an overall evaluative conclusion (see Table 1). Since this was the case, then the rest of Taylor's logical process of evaluation will be examined as it applies to S-GRADING.

Figure 2 has been prepared to aid in the following description of Taylor's philosophical position. This figure should not be rigidly

Fig. 2

An Outline of the Logical Process of Evaluation

interpreted as a flow chart depicting a temporal sequence of events. The process of evaluation does begin with an evaluatum, and end with an overall estimation of goodness or badness; however, the intervening steps may occur in any order and repetitively, before arriving at the final evaluative conclusion. Nevertheless, it is a necessary condition in Taylor's view of evaluation, that all the steps in Figure 1 must be completed before an evaluation can be logically verified.

2.40 Level 1 of S-Grading: Adoption of Standards (See Fig. 1)

Taylor identifies two types of standards. They are standards of intrinsic value and standards of extrinsic value. (Taylor 1961, pp. 20-25).

2.41 Intrinsic Value

Standards of intrinsic value are used to evaluate things which are valuable in and for themselves (Daniels, 1971, p. 9) and thereby satisfying the following two conditions. Firstly they must be used only in evaluating immediately felt or perceived qualities of our own experience; and secondly, they must be standards of nonderivative value (Taylor, 1961, p. 23). If an evaluatum produced feelings of pleasantness, then those feelings would have intrinsic value.

2.42 Extrinsic Value

Whereas standards of intrinsic value are of nonderivative

value, standards of extrinsic value are of derivative value, that is, dependant on the value of other things. Taylor has sub-divided extrinsic value into inherent value, instrumental value and contributive value. (Taylor, 1961, pp. 26-32, 310-311)

If an evaluatum is judged in terms of standards of inherent value, then the evaluator would look for evidence that it produces experiences which are judged as intrinsically good. Inherent value differs from intrinsic value in that the former must be other than a quality of our own experience.

"Works of art and things of natural beauty may also be said to be good on the ground that one who looks at them normally has a good or rewarding experience. Then, we may say that they have inherent goodness". (Frankena, 1973, p. 81)

In other words an object's inherent value depends on its capacity to produce intrinsic value of someone's experience in response to it. (Taylor, 1961, p. 28)

To have instrumental value, the evaluatum must be effective in bringing about a given end which is intrinsically or extrinsically valuable. As Daniels said, (1971, p. 9).

"A thing has instrumental value if it can be used to obtain some thing or condition other than itself which is, for some reason or other, considered to be of value."

In the broadest sense of instrumental value an evaluator would grade or rank a teacher-training program according to its effectiveness in enabling students to become good teachers. In a narrower sense

he could examine methods courses (or parts of those methods courses) for their success in teaching the student some particular technique of value.

To fulfill a standard of contributive value, a "part" must contribute to a good whole. The more necessary it is to the whole, the more contributive value it has. For example, professors of education should contribute greatly to the production of good teachers, therefore they should have high contributive value in a teacher-training program and would be judged in terms of standards of "contribution".

As Taylor indicates, the above three kinds of standards of extrinsic value are dependent upon intrinsic value. He says (1961, p. 32):

"First, a thing may have instrumental value in virtue of the fact that it is a means to an end which has contributive value as part of a whole. This whole in turn either may have inherent value in itself or may have instrumental value as a means to an end which has inherent or intrinsic value."

In other words, the final appeal for the raison d'être of anything is that some one or some group of people derives intrinsic value from it. "Thus we arrive finally at things which we judge to be desirable in themselves. Their value is intrinsic to them." (Taylor, 1961, p. 32) Or as Frankena (1973, p. 81-82) wrote:

"We also sometimes say that things are good; desirable, or worthwhile in themselves, as ends, intrinsically. When someone asks 'What is _____ good for; the answer may be given by trying to

exhibit its usefulness, extrinsic value, or inherent goodness; but one may also try to show (and here the final appeal must be 'Try it and see') that it is enjoyable or otherwise good in itself . . . In fact, it is hard to see how money, cars, and other material possessions, even paintings, can have any goodness or value at all, extrinsic or inherent, if the experiences they make possible are not in some way enjoyable or good in themselves."

The above statements about intrinsic value do not mean that an educational product is not worthwhile if it doesn't have intrinsic value. However, if the program does not have intrinsic value, then it should be shown that achieving the goals of the program is instrumental in fulfilling some aspect of living which will have intrinsic value. If there is no apparent connection between the program goals and future attainment of intrinsic value, then the worthwhileness of the program goals should be seriously questioned.

2.50 Level II of S-GRADING (See Fig. 1): Operational Clarification of the Standards in Grading

"This consists in a set of statements to the effect that if an object O has characteristics C, it fulfills a certain standard S to a certain degree D; that if O has characteristics C', it fulfills S to a greater (lesser) degree D'; etc." (Taylor, 1961, p. 10)

Once a standard has been generally identified as a standard of intrinsic or extrinsic value, then it must be elaborated in such terms that permit the evaluator to say: "characteristic C fulfills standard S to degree D." This intermediate step in the process of evaluation has been called "conceptualizing" by Scriven (Scriven, 1974a, p. 142). He sees this step in evaluation as a process of

both "devising an appropriate taxonomy and measuring the specific performance in terms of the dimensions of the taxonomy." (Scriven, 1974a, p. 137). In Chapter III it will be shown how each standard chosen in the evaluation of a teacher-training program is elaborated so as to allow some rating of the degree to which the evaluatum met that standard.

2.60 Level III--Specification of the Class of Comparison in Grading

In S-GRADING, the class of comparison is some ideal form of the evaluatum. Taylor's example (1961, p. 9) of the S-GRADING of paintings ended ". . . is [the painting] good as compared with what a painting of this type should be?" In the instance of the S-GRADING of a teacher-training program the parallel question becomes: "Is the program good as compared with what a teacher-training program at this level should be?"

S-grading, as argued earlier, is only an intermediate step in the complete evaluation of a product. The strongest evaluative conclusion which can be realized is "Commendatory" (See Table 1). The choice of S-GRADING or RANKING is an option in the first four evaluative conclusions of Table 1, and is primarily based on circumstance. If there is a readily identified comparison group(s) present, and if the evaluator has access to the comparison group(s) early enough to plan a parallel evaluation scheme for the evaluata, then he should probably use RANKING. Scriven (1967, p. 64) argues:

"Comparative evaluations are often very much easier than noncomparative evaluations, because we can often use tests which yield differences instead of having to find an absolute scale and then eventually compare the absolute scores."

If, on the other hand, the comparison group(s) is not easily identifiable and/or the evaluator cannot gain access to it early enough to perform evaluations on it then S-GRADING is the best alternative in the early stages. However, the evaluator should be considering methods of using, or creating (Scriven, 1967, p. 69), comparison groups if anything like a complete evaluation is to be accomplished.

2.70 Level IV--Determination of the Good and Bad Characteristics of the Evaluatum in S-GRADING

This particular step in a logical evaluation can be broken down into two subprocesses; they are:

- a) Gathering data on the characteristics.
- b) Determining if the characteristics are good or bad.

2.71 Gathering Data on the Characteristics

This sub-process involves using various psychometric techniques to gather data. In this particular study, the general techniques used were: a Likert attitude scale, questionnaires, observation, formal and informal interviews, examinations of records of meetings, and examinations of other internal evaluations (overall and of student teaching). Metfessel and Michael (1967, p. 937-943)

provided a very comprehensive list of possible techniques for the evaluation of school programs. The techniques used in this study will be elaborated in Chapter III, Section 3.51.

2.72 Determining if the Characteristics are Good or Bad

Each characteristic, on which data has been collected, is judged as good, bad or neutral (equal amounts of goodness and badness). This judgement depends not only on the point of view being taken, but also the value system in the particular point of view. Thus while a characteristic which is judged as good from the intellectual point of view, may be bad from the economic point of view, it is also the case that a course in home economics which dealt with the preparation of pork for the table may be religiously quite acceptable for an Anglican, but religiously repugnant to an orthodox Jew.

The problems of evaluator and client working at cross-purposes due to conflicting value systems or points of view are magnified in international education. The following quote is included as a scaled-up, but particularly good, example of these concerns:

"There are frequently several interests in a single development project; the interest of the responsible organization (government agency, business enterprise, private institution, etc.) being, in theory, paramount and usually also in practice. If intergovernmental organizations or foreign governments or private foundations also participate in the project, they may be assumed to have their objectives to achieve, and these may or may not be identical with those of the responsible organization. Moreover, individual experts, although nominally, and usually governed by the goals of employing organizations, sometimes have their own and different objectives (such as testing out a pet theory, or demonstrating the superiority of a particular

method or product). The people likely to be affected by a development project also have their own goals, which may or may not be reflected in the policies of their government. A government may change, too, or a new official be appointed during the life of a project, and as a result different goals may be established." (Hayes, 1969, pp. 29-30)

In a similar vein, Scriven (1974a, p. 138) wrote:

"The developer or teacher always has one conceptualization of the data in mind: If he (or she) feels he has been successful, he or she naturally sees the data as 'demonstrating success in achieving such and such goals,' and hence (since those goals would not have been adopted had it not been felt they had merit) the project is judged meritorious. But there are many other ways to see most projects. It is just as important for the evaluator, as opposed to the developer, to retain an open mind about the legitimacy of radically different interpretations of data, as it is for the scientist reading a research paper in which the author proposes that certain experimental results support his theory."

It is for these reasons that the evaluator must be sure that the client for the evaluation understands the point of view being taken in the evaluation; and that the value systems of the client and evaluator are fairly congruent. When the evaluator makes a judgement that a particular characteristic, "C", is "good" according to standard "S", then he must have arguments which will support his contention; he must be able to point at various evidences for "C" and, through them, verify his value judgement. (Taylor, 1961, p. 70)

In other words, the evaluator must have complete and valid evidence about a particular characteristic and he must have good reason (given the evidence) for saying the characteristic is good, bad or neutral. If the evaluator is challenged on his judgement, he would first show that his evidence was complete and valid

(verification). If he was still challenged, he would then show that his reasons for judging that that data on that characteristic meant the characteristic was good, bad or neutral, were good reasons (validation).

Scriven (1974a, p. 139) has called this process, which leads from the conceptualized intermediate conclusions to the eventual conclusions of worth or merit, credentialing.

2.80 Level V-- The Overall Judgement of Worth

There are two factors that determine whether an evaluatum is to be graded as good or bad on the whole. The first factor, is the degree to which the evaluatum fulfills all the standards applied to it. As Taylor (1961, p. 8) writes:

"That depends on what standards one is appealing to, . . . how clearly those standards are defined, to what extent the degrees to which they can be fulfilled are measurable, and how "distant" from reality is one's ideal."

The second factor is the relative precedence of the standards. When the evaluatum is evaluated according to a set of standards, those standards may be ranked in order of decreasing importance (i.e. S_1 , S_2 , S_3 , . . . , S_n where S_1 is the most important standard and S_n is the least important standard). That is to say, when we are making an overall judgement of value, greater weight is given to the fact that the evaluatum fulfills S_1 than to the fact that it fulfills S_2 . If the evaluatum fulfills S_1 and S_2 to the same degree, its fulfillment

of S_1 adds more to its overall goodness than its fulfillment of S_2 (Taylor, 1961, p. 13).

CHAPTER III

Application of the Model to The Design of an Actual Evaluation Study.

3.00 Introduction

Chapter Three applies the model of evaluation developed in Chapter Two to the design of an evaluation study of an undergraduate teacher training program. The chapter illustrates the choice of standards, data collection on the standards, and how an evaluative conclusion can be reached. Each section of the Chapter illustrates a practical application of the similarly headed, but theoretical sections in Chapter Two..

3.10 Establishing the Preconditions

3.11 The Evaluatum

The evaluatum is an undergraduate teacher training program subsequently referred to as Program B. A description of the Program does not necessarily portray the Program as it actually operated but rather as it was intended to operate, including a time scale of operation, the geographical location, the personnel requirements, etc. Determining the degree to which the intended program and the actual program were congruent is a function of the evaluator.

The following is a description of the evaluatum based on the materials prepared for the Program.

3.111 Letter to Potential Students for Program B.

Dear Student:

The Faculty of Education has initiated several new programs for the professional training of teachers. One of these new programs, titled "Program B" is described in the accompanying letter. Briefly, what we intend to do is have a group of 25 students spend the entire year at school X, a Vancouver elementary school. All university courses and all student teaching will be undertaken there. The administration and staff of School X and faculty members from U.B.C. have cooperatively planned the program and are very enthusiastic about this new approach to teacher education. As a 3rd year transfer student, you are eligible to be a part of this new program. The program is limited to 25 students and we hope to fill the section before the end of the Vancouver school year. If you are interested please contact me immediately so a preliminary meeting or two may be arranged.

Sincerely,

Program B Director,
Faculty of Education
University of B.C.

3.112 Program Description

Program B will be conducted entirely in a participating elementary school, but students will be able to make use of the University libraries. It has been agreed that it will contain:

1. a heavy component of actual classroom experience (Education 397)
2. professional methods courses
3. a psychological component
4. a foundations component
5. a general elementary education course

The form which the classroom component will take will depend on the school staff and their recommendations, but a flexible, varied, teaching program that includes experiences at several levels of school life is envisioned. The methods courses will be interested in the specific requirements of each area, the materials available, and curriculum development. During the time devoted to professional

methods the following courses will be covered:

| | | | |
|---------|-----------------------|-------------|-------------|
| Ed. 304 | Language Arts | 3 units | Professor A |
| Ed. 321 | Science Education | 1-1/2 units | Professor B |
| Ed. 323 | Art Education | 1 unit | Professor C |
| Ed. 325 | Physical Education | 1 unit | Professor D |
| Ed. 370 | Mathematics | 1-1/2 units | Professor E |
| Ed. 371 | Mathematics Education | 1-1/2 units | Professor E |

The psychology component will be designed to increase the sensitivity of the pre-service teachers to the needs and requirements of children not as consumers of science, mathematics, or any other content area, but rather as human beings. This component is also designed to help the pre-service teachers understand some of the problems associated with learning and evaluation. To this end the theoretical material associated with Education 310 and 311 will be covered.

The foundations component will be designed to accomplish three major tasks:

1. To enable pre-service teachers to make clear, well-founded decisions as to what goals they, as teachers, ought to pursue.
2. To enable them to promote, in an ethically responsible manner, the moral development of their students.
3. To give them a clear understanding of the nature of the task of developing the ability to think.

To this end the theoretical material associated with Ed. 400 will be covered.

The general elementary education component (Education 449) will contain many of the common elements of the methods courses. It will also deal with much of what is currently done in student teaching seminars at the University. Further it will deal with such topics as teaching strategies, classroom organization and curricular planning.

3.113 Description of the offered courses in Program B

(Taken from the Calendar of the University of B.C.

1974/75).

Ed. 304: Curriculum and Instruction in the Language Arts. -- A study of (a) the curriculum organization in the language arts particularly in the intermediate grades; (b) techniques of instruction in these subjects and grades.

Ed. 321: Curriculum and Instruction in Elementary Science. -- A study of (a) the curriculum organization in science for the elementary grades; (b) techniques and strategies of instruction in science for these grades.

Ed. 323: Curriculum and Instruction in Art I.

Ed. 325: Curriculum and Instruction in Physical Education. -- A study of (a) the curriculum organization in physical education for the elementary grades; (b) techniques of instruction in physical education for these grades.

Ed. 370: Mathematics for Elementary Teachers. -- A study of the method and structure of elementary school mathematics as a discipline.

Ed. 371: Methods of teaching Elementary School Mathematics. -- A study of materials, skills, strategies and curriculum organization.

Ed. 310: Growth and Development. -- Research as it applies to the elementary school child. Not open to students who have taken Education 331.

Ed. 311: The Nature and Measurement of Learning. -- A study of learning and the techniques of evaluation as they apply to the elementary school child. Not open to students who have taken Education 331.

Ed. 400: Philosophy of Education. -- An introductory course in which consideration is given to the philosophical foundations of education and to the practical bearings of theory upon curriculum content and classroom practice in our schools.

3.12 The Point of View

In Chapter II, Section 2.32, the eight basic points of view as outlined by Taylor (1961; p. 300) were given. They are the moral, the aesthetic, the intellectual, the religious, the economic, the political, the legal and the point of view of etiquette or custom. Taylor considers these points of view to be basic because:

"... the conduct of any given individual in a culture is always subject to a value system belonging to at least one of them and is usually subject to value systems belonging to more than one of them. Second, they are the dominant points of view in a culture, in the sense that they set the values of the major social institutions and activities which carry on the civilization of the culture."
(Taylor, 1961, p. 300)

Moreover, in addition to the eight basic points of view there are many nonbasic points of view corresponding to particular group interests.

" . . . it is appropriate to judge the organization in terms of its group interest, that is, in terms of the purposes of the group as a whole (as distinct from the self-interest of each member of the group)." (Taylor, 1961, pp. 301-302)

A faculty of education exists for the clearly defined purpose of producing good teachers, (using the best techniques possible), and can legitimately be judged according to its ability to perform that function. Similarly, so can any particular program of teacher education in the faculty. Therefore, the evaluator can evaluate a program of teacher education from the institutional point of view, that is, from the point of view of a faculty of education. This point of view would encompass the logistics of teacher education, such as opportunities for practise teaching, available methods course, professors available to teach the students, financing of programs and so on. In other words, the evaluator would use these criteria for judging the activities and policies of a program operating under the auspices of a faculty of education.

It is also necessary for the evaluator to judge a teacher education program from the educational point of view. If we are educating people to be teachers, rather than simply training them to be teachers, then the practises and policies of a program of teacher education must be judged in terms of their educational worth.

It is necessary that all programs be evaluated from the moral point of view and the legal point of view ". . . since [the Program's] purposes are to achieve certain goals or to pursue certain ends without violating the moral code of the society and without breaking Society's laws." (Taylor, 1961, p. 300)

Therefore, Program B will be evaluated from the point of view of the faculty of education (that is, the institutional point of view), from the educational point of view, and from the moral and legal points of view.

Some other basic points of view which could be considered in the evaluation are the economic, the political, the intellectual, and the point of view of etiquette or custom. However, Program costs were not a consideration in this evaluation so no evaluation was done from the economic point of view; the intellectual point of view was included in the educational point of view; and the political and etiquette or custom points of view were beyond the scope of this thesis.

3.13 Grading

The method of evaluation described by this thesis is S-GRADING. (see Chapter II, Section 2.33). That is, Program B is not directly compared to other existing programs in terms of its goals and performance. Rather, Program B is graded according to the degree to which it fulfills a set of standards which are considered by the evaluator to be appropriate to the Program. S-GRADING is reasonable in formative evaluation, where the evaluator must make judgements as to the worth of the program for the purposes of improving it and where comparison with other Programs is not a major concern.

Referring to Table I in Chapter II, it was shown how the most comprehensive evaluative conclusion which could be reached in S-GRADING

(when costs have not been taken into account), is the "Overall Evaluative Conclusion." Therefore, the evaluation described by this thesis attempted to answer the following question: "Did treatment X provided by Program B, have the effect of producing undergraduate student teachers (S), competent in the classroom (Y) (for this level of university training); with the premises that X was the treatment and it caused Y; the S's enjoyed, desired and benefitted by Y, and Y had no harmful effects on S's (or non S's)?"

3.20 Adoption of Standards

It is useful to look at a generally agreed upon description of education in order to establish the broad standards to be used in this evaluation. Peters (1966, p. 3) provided this definition when he wrote: "[Education] implies that something worth-while is being or has been intentionally transmitted in a morally acceptable manner." Thus, there are two criteria for evaluating educational processes. The first is the worth of the process (or product) goals and the second is the "transmission in a morally acceptable manner."

" . . . to educate someone implies not only some sort of achievement, but also one that is worth-while. It also implies that the manner of doing this should not be morally objectionable." (Peters, 1966, p. 4)

It is this essential duality of end product and process which has been often overlooked in educational evaluation. Scriyen (1967, p. 51) wrote,

"One of the reactions to the threat of evaluation . . . was the extreme relativization of evaluation research. The slogan became: How well does the course achieve its goals? instead of How good is the course? But it is obvious that if the goals aren't worth achieving then it is uninteresting how well they are achieved . . . Thus evaluation proper must include as an equal partner with the measuring of performance against goals, procedures for the evaluation of the goals."

Therefore, for the evaluation model developed in this thesis, two standards will be used. The first standard is goal worth. The degree to which Program B fulfills or fails to fulfill this standard is dependent on the worth of the Program's goals. The second standard is goal achievement. The degree to which Program B fulfills or fails to fulfill this standard is dependent on whether or not the operation of the Program results in the goals being achieved.

There is one restriction placed on the standard of goal achievement. That is, its degree of fulfillment should only be determined if the goals have been shown to be worthwhile.

3.21 Standard 1: The Worth of the Program Goals

One approach to establishing the worth of a set of goals is to determine the extent to which they are based on defensible needs and/or wants. For some, defensible needs are seen as arising out of the discrepancy between a desired state of affairs in education and the current state of affairs. Both Popham (1969, p. 23) and Provus (1972, p. 37) have advocated this discrepancy as a measure of educational need. Popham wrote that desired learner outcome minus current learner status equals an educational need. Provus, similarly,

wrote that the discrepancy between idealized performance of citizens and actual performance equals need. However, this does not necessarily determine need in education. Suppose the "desired learner outcomes" or "idealized performance of citizens" have not been shown (in the context of a particular educational environment) to be worthwhile goals; or, suppose they are in fact worthless goals; in that case, the discrepancy between current learner status and those goals could not be described as an educational need. This argument could be countered, of course, by saying that desired learner outcomes would never be desired unless they were worthwhile. This can be shown to be fallacious by using the "open-question argument."

Frankena (1973, p. 99) wrote about this type of argument:

"Suppose that a definist holds that "good" or "right" means "having the property P," for example, "being desired" or "being conducive to the greatest general happiness." Then, the argument is that we may agree that something has P, and yet ask significantly, "But is it good?" or "Is it right?" That is, we can sensibly say, "This has P, but is it good (or right)?" But if the proposed definition were correct, then we could not say this sensibly for it would be equivalent to saying, "This has P, but has it P?" . . . Therefore the definition cannot be correct."

Since we can ask the question "Is a desired learner outcome a worthwhile outcome?" then we cannot assume that "desired" equals "worthwhile". Need in this sense of being a discrepancy between a desired and actual state is essentially an indication of degree of goal achievement without necessarily seeing if the goals are worth achieving.

Scriven has identified needs as biological. He wrote (Scriven,

1974b, p. 13)

"True needs assessments involve establishing that the product actually facilitates survival, health, or some other defensible need that is not adequately serviced . . ."

However, in education the criteria of survival etc., for establishing the worth of the program's goals, is often too restrictive. As Komisar wrote (Komisar, 1961, p. 37):

"Survival . . . is too miserly a conception, since there are deficiencies which arouse us without the threat of death (e.g. affection)."

To determine if the goals of the program, or the desired learner outcomes of the program are worthwhile requires validation of the goals or outcomes. This is accomplished by showing: (Taylor, 1961, p. 84-85, Daniels, 1971, p. 12).

- 1) That fulfilling the goals enables one to fulfill higher goals.
- 2) That fulfilling the goals does not conflict with another principle of our way of life which is of importance to us.
- 3) That fulfilling the goals leads to beneficial consequences.
- and 4) That there are no grounds for making an exception to the goals in this situation.

Scriven (1966a) offered the following similar method for determining if something is worth doing:

- 1) If something will bring about a state of affairs that people

value, that is a good prima facie reason for doing it.

2) If there are prima facie reasons for doing something and none against, we should do it.

3) If there is a conflict of supportable prima facie reasons, appeal must be made to a general moral principle.

3.22 Standard 2: Goal Achievement

Goal achievement is a measure of the effectiveness of transmitting something in a morally acceptable manner to achieve the worthwhile goals. In other words it is a check on the performance of a particular evaluatum in achieving the worthwhile goals. Moral considerations, in considering the standard of goal achievement, must take precedence over any rating based on performance. If the program is immoral then it should be stopped regardless of how well it appears to be achieving the goals.

It is this check on morality which is often neglected in the evaluation of educational products and proposals. It is a useful exercise to evaluate Jonathan Swift's (1667-1745) "A Modest Proposal for Preventing the Children of Poor People from Being a Burthen to their Parents or Country, and for Making them Beneficial to the Public," using the major evaluation models, to illustrate the consequences of omitting a check on morality. Most of the models would find the proposal quite acceptable. This is not to suggest that cannibalism would be overlooked in the evaluation of an educational

product; but, other transgressions such as indoctrination might be conveniently ignored or not looked for, unless there is an explicit statement requiring a morality check.

3.30 Operational Clarification of Standards

In this section each standard will be explicated in the terms used to evaluate the teacher training program, Program B.

3.31 Standard 1: The Worth of Program B's Goals

The goals of the program can be derived from the type of program being evaluated. The evaluator does not look at the program's plans or performance to determine if its goals are worthwhile; instead he looks at the characteristics of the overall program goals and validates them as being or not being worthwhile. Program B is a particular method of training teachers, but regardless of the techniques used, the ultimate goal of the Program is to train people to be good teachers. It is that goal which is validated.

As Taylor (1961, p. 301-302) suggested, if you are evaluating the activities and policies of a group in terms of that group's interest then you would evaluate the worth of the goals and the degree of goal achievement as they pertained to only that group. The point of view to be adopted would be the point of view of the institution or, in this case, the faculty of education under whose auspices Program B is functioning. On the other hand, you can evaluate

Program B outside of the group's framework. It is now legitimate to ask how the Program's activities and policies conform to the rules and standards of an enlightened public. The point of view to be adopted would be the educational point of view.

Program B, as described in the Section 3.11 (The Evaluatum) did not have a particularly detailed list of goals; therefore a set of worthwhile goals was constructed using a well thought-out model of teacher education. This model was supplied by Broudy (1965). In his article he identified the teacher as being a member of the educational profession and as a specialist.

As a member of the educational profession a potential teacher requires:

- (1) Basic courses in history, philosophy, psychology and the social sciences, as part of his general education.

- (2) Study of the historical, philosophical, psychological and sociological foundations of educational policy, curriculum, organization and support and strategies of teaching and learning.

The teacher as a specialist requires:

- (3) The historical, philosophical, psychological and sociological foundations of the specialty.

- (4) Study of subject matter over and above that taken in general education as needed for the specialty.

- (5) Laboratory exercises: abstracted and often schematized samples of a whole class of real tasks such as a made-up grouping

task, a make-believe test, a demonstration, etc. Broudy refers to this as laboratory exercises.

(6) Clinical experience: especially constructed situations to demonstrate actual examples of studied theory. For example, the classic case of the slow learner or discipline problem.

(7) Internship: the student-teacher spends time working on a real task under a minimum of supervision, i.e. the practice-teaching segments of teacher education.

(8) Research: The teacher ought to be a consumer of research and may be a producer of research. Therefore he should be introduced to to his methods, canons, status and prosperity.

The first four requirements refer to the foundations of the educational field and area of specialization in this field. As such they constitute educational goals of a faculty of education. Requirements (5) - (7) refer to the technology (i.e. the way and means) of applying theory to practice. Thus, they can be seen as institutional goals, because they primarily serve the interests of the institution (e.g. improving classroom practice). The technology is not unique to 'education', but it is of interest to the field of education. Research, requirement 8, is an institutional goal in that it serves the interests of the field.

Broudy's paper, as briefly described above, forms the basis for determining the degree to which Program B has worthwhile goals from the educational and institutional points of view.

It is important to point out that Program B is only one year of a four year teacher training program. Therefore parts of Broudy's criteria will be fulfilled either before or after Program B. This would be taken into consideration in the final, overall evaluation of the worth of the program.

3.32 Standard 2: The degree of Goal Achievement

Assuming that the goals of a program have been identified and are worthwhile, then the evaluator would attempt to determine if the program is achieving the goals. Often, the program developers will have written out, in great detail, how they hope to achieve the (worthwhile) goals. It is useful for the evaluator to examine these written intents to check their validity. Some reasons for this, as Scriven (1974a, p. 154) points out could be: (1) new programs on the market could make a switch to them worthwhile. (2) The performance evidence of the existing program, or others to which it was compared, may have changed. (3) Unanticipated difficulties (political, moral, economic) may arise which have not occurred in other alternate programs or methods (4) the original decision to run the program may have been wrong due to poor data or poor logic.

If this is done, then the next step is to identify areas of investigation which will provide the evaluator with the data necessary to determine whether the goals are being achieved. Scriven (1974b) has contributed a great deal to this end by developing

a checklist which can be used to evaluate products, producers and proposals. He supports the validity and utility of the checklist by making the following four points (Scriven, 1974b, p. 8-9):

(1) Every checkpoint has a clear a priori rationale. Failure to meet any one of the checkpoints immediately causes doubt about the quality of the evaluation.

(2) Medical and industrial routinely pass, and often are required to pass, every checkpoint.

(3) The checklist has been developed out of the Product Review Panels of 1971-72 and 1972-73 done for the National Center for Educational Communication, on subcontract to the Educational Testing Service.

(4) It has been used by several hundred school administrators in evaluating graduate programs, and by students in the evaluation training seminar at the University of California, Berkeley. They frequently state that it is more value to them doing evaluations than anything else available in the literature.

Each checkpoint will be described in general terms, and in terms of Program B. This description will be followed by summary version of the checkpoint. The summary version consists of a methodological check and a substantive check.

The methodological part is a check on whether data is available (or whether it can be obtained), its source, and the quality of that

data. The substantive part is a degree of fulfillment scale. It indicates how well the checkpoint has been met. In order for an evaluatum to receive a high rating on a particular checkpoint there must be good evidence of good performance; i.e. the data is complete, and it indicates good performance by the evaluatum. A low score on a checkpoint implies that either good evidence or good performance or both is lacking.

"It does not require, for example, that there is good evidence of bad performance, for otherwise products which turned in no data would do better than those that were known to fare badly." (Scriven, 1974b, p. 22)

3.33 The Checklist for Determining Good Achievement

3.331 Checkpoint 1: MARKET (Disseminability)

This is not an economic checkpoint. It requires that, given worthwhile goals, there are dissemination plans that ensure consumers for the educational product. It would consider the clarity, feasibility and ingenuity (for utilizing available resources) of the implementation plans.

To fulfill this checkpoint to a high degree Program B must show that it has specific, feasible plans to provide the students with (following Broudy's (1965) criteria):

a) The historical, philosophical, psychological and societal foundations of the specialties.

b) The professional content of the specialties over and above that taken in previous courses.

c) Laboratory exercises, clinical exercises and internships.

d) Methods of procedure followed in practise, professional tasks (grading, marking, making tests), and methods of dealing with the apparatus and materials of teaching.

In addition, there must be students who want to enroll in Program B and procedures to enroll them. There must be instructors able to teach the courses. A school must be available with a staff willing to take on Program B; plus a myriad of minor details such as books and materials being present, films, access to libraries, areas for study, etc., which, cumulatively, can have considerable impact on the success of a program.

Summary version of Market Checkpoint.

| Methodological Evidence | Substantive Fulfillment characteristics | rating |
|-------------------------|---|--------|
| Dissemination plan: | Very large and/or very important market | |
| clarity | will be reached | 4 |
| feasibility | Large and/or important market will be | |
| ingenuity | reached | 3 |
| economy | Significant market will probably be | |
| Size | reached | 2 |
| Importance | Possible, but not probable, that a | |
| Other | significant market will be reached | 1 |
| | Inadequate evidence to suggest that a | |
| | significant market will be reached | 0 |

3.332 Checkpoint 2: TRUE FIELD TRIALS.

How typical are the people, the setting, the time

allotted to the final version of the program?

"It is very tempting to think one can extrapolate from field trials with volunteer schools . . . but this has frequently proved unsound . . . in actual practise, deadlines, overcommitment, and underfinancing combine to render almost all products deficient on this checkpoint." (Scriven, 1974b, p. 11)

The parallel in statistics is the concept of generalizability.

If Program B is to obtain a high score on this checkpoint, then it must be shown that the school, in which the program was held, was not particularly atypical. This is true regardless of whether the school was exceptionally good or exceptionally bad. If it was exceptionally good it is obvious that the program probably received more of its share of help to be a good program. If it was exceptionally bad, then the temptation is to say: "If it can work under those terrible conditions, it can work anywhere!" But that is not necessarily the case, unless it has been shown that it can work anywhere. The safest bet is to run the program in an "average" school and make adjustments to the program's developed framework if it subsequently ends up in a very bad or very good situation.

Not only should the school setting be "average", but so should the university staff and the student teachers. If portability of Program B is one of its aims, then the trial versions should not be composed of exceptionally good or bad staff or exceptionally good or bad student teachers. The arguments against exceptionally good or bad staff and students parallel those given in the preceeding paragraph

for the school.

Summary version of TRUE FIELD TRIALS Checkpoint

| Methodological Evidence | Substantive Fulfillment Characteristics | Rating |
|----------------------------|--|--------|
| Final version? | Perfectly typical | 4 |
| Typical user? | Minor differences | 3 |
| Typical aid? | Reasonable bet for generalization | 2 |
| Typical setting? | Serious weakness | 1 |
| Typical time-frame? | Relevance unclear | 0 |

Comment: Checks should be made on all the evidence sources. This will provide a good base for the rating given.

3.333 Checkpoint 3: TRUE CONSUMER

This checkpoint outlines gains expected of various groups (consumers) involved in Program B. The types of gains will usually depend on the point of view taken in the evaluation. Deciding on the data requires a clear understanding of the evaluative function and " . . . which audiences it is addressed to, commissioned by and--regardless of these two considerations--responsible to." (Scriven, 1974b, p. 12). The different groups of consumers may be interested in different aspects of the project, and, in order to provide data for all the relevant groups, each aspect should be examined. This is evaluating the program from the prudential point of view. That is, each consumer of Program B would in effect be asking "What beneficial effects does this program have on me?"

To fulfill the requirements of this checkpoint, the evaluator

would determine the goals, with respect to the evaluatum, of each consumer. He would then gather data on these goals and feed them back to the consumer for his appraisal of the beneficial and/or harmful effects (with respect to the consumer's goals). The evaluator, however, must have overriding judgement since he is the one who takes responsibility for the conclusions. It is important that the data be objectively gathered so that the consumer can make a verifiable judgement.

Though the evaluator tries to provide the necessary data, it is obviously not possible to provide "Best Buy" information if the evaluation is "Overall" (see Table I). Limitations of this sort must be pointed out early in evaluation, and revised up or down as the evaluation proceeds.

The different consumers in this study are:

- a) The Dean of Education and his representatives.
- b) The U.B.C. staff involved in Program B.
- c) The Student teachers in Program B.
- d) The Principal of the school in which Program B operated.
- e) The Teachers in the above school.
- f) The pupils in the above school.
- g) The School Board.
- h) Future consumers of the program.

The most important consumers for this stage of the evaluation are the student teachers in Program B, and U.B.C. Staff, and ultimately

the pupils in the school being exposed to the student teachers.

Summary version of TRUE CONSUMER checkpoint.

| Methodological Evidence | Substantive Fulfillment characteristics | Rating |
|-------------------------|---|--------|
| District? | Full data on all relevant "consumers" | 4 |
| Principal? | Fair data on all relevant "consumers" | 3 |
| Teacher? | Good data on the most important "consumers" | 2 |
| Student? | Weak data on the most important "consumers" | 1 |
| Taxpayer? | Only speculation about most important | 0 |
| Other? | "consumers" | |

3.334 Checkpoint 4: LONG TERM

This checkpoint outlines follow-up plans. These plans are particularly important when the project being evaluated can be expected to have long range effects--as is the desirable case for most educational projects. The follow-up can also identify any good or bad side effects which may take considerable time to surface.

Summary version of LONGTERM checkpoint

| Methodological Evidence | Substantive Fulfillment characteristics | Rating |
|-----------------------------|--|--------|
| Week to month later | Good direct evidence about the effects at times needed. | 4 |
| Month to year later | Some direct evidence about the effects at times needed. | 3 |
| Year to few years later | Follow-up gives reasonable support to a conclusion about effects when needed | 2 |
| On-job or life-space sample | Follow-up or other data suggests a conclusion about effects when needed. | 1 |
| | Useless or no follow-up; no other grounds for inferring long-term effects. | 0 |

3.335 Checkpoint 5: SIDE EFFECTS

This broad checkpoint provides a general pattern for

a free ranging search for any good or bad program effects on the consumers. It is particularly important that the evaluator makes his search under many different points of view.

Scriven (1967, pp. 77-80) has given examples of the forms these side-effects may take in a new curriculum project. Applying this to the consumers of Program B, then examples of side effects for each group are:

1) The U.B.C. Staff

- a) Increase or decrease in status as a result of being connected with Program B. (Promotion, Demotion).
- b) Problems or benefits of travelling between the university and the school hosting Program B.
- c) The benefits or problems of spending a lot of time on the Program; such as: fatigue, being able to develop his/her own ideas, loss of time for research, etc.

2) The Student-Teachers:

- a) Increase or decrease in status as a result of being connected with Program B.
- b) Problems or benefits of being removed from the University setting (e.g. the library, other professors, students in other courses, etc.)
- c) Fatigue or enthusiasm due to having to teach pupils in the host school, as well as having to do academic work at the same time.

- d) Problems and benefits of studying, and working together (exchanging ideas, being crowded, no quiet areas, encouraging the despondant, etc.).
 - e) Problems and benefits of being in close contact with the host school's teachers (see 2d).
- 3) The Host School's Teachers:
 - a) Increased or decreased work load (having a student teacher to help or to hinder).
 - b) Increased or decreased status.
 - c) Opportunity for professional development.
 - d) Emotional pleasures and problems having student teachers always in the school, staff-room, and their classroom.
- 4) The Principal
 - a) Increase in administrative duties (allocating classrooms, cooperating with the University staff and students, etc.).
 - b) Increase or decrease in status.
 - c) Fewer or greater numbers of discipline problems.
- 5) The Pupils of the Host-School
 - a) Did they find the frequent changing from student-teacher to regular teacher disturbing in any way, or beneficial in any way.
 - b) Could the pupils obtain more individual attention and help as a result of there being teachers, student-teachers and U.B.C. staff in the school.

6) The Dean and his Representatives

- a) Does the program present problems or advantages in time-tabling, allocating staff and rooms, and financing.
- b) Has the Dean's office received accolades and/or abuse as a result of the Program functioning.

7) The School Board

- a) See 6 (a).
- b) Has the school board office received accolades and/or abuse as a result of the program functioning under the Board's jurisdiction.

The above are examples of possible side effects. It is not meant to be an exhaustive list. Side effects differ from process effects in that they are unanticipated and not part of the treatment. If the side effects are part of the treatment rather than a result of it then they are process effects.

Summary version of SIDE EFFECTS Checkpoint.

| Methodological Evidence | Substantive Fulfillment characteristics | Rating |
|-------------------------|--|--------|
| Comprehensive search? | Meets all requirements well; side effects overall positive. | 4 |
| Skilled? | Generally good; side effects overall positive. | 3 |
| Independent? | Barely acceptable; side effects overall positive or neutral. | 2 |
| During/End/ | Some study made, but incomplete. | 1 |
| Later | No worthwhile study. | 0 |

Comments: The evaluator should use all the Evidence Sources. A high rating on this checkpoint indicates that the side effects are good and that the quality of information on the side effects is good.

3.336 Checkpoint 6: PROCESS

The process checkpoint is multifaceted. It is used to substantiate or invalidate descriptions of the product, causal claims involved in the evaluation (that the treatment caused the gains) and/or it may bear on moral questions (e.g. the intrinsically or inherently negative dimensions of cruelty, injustice and unhappiness, as well as their opposites). The pay-off emphasis is of primary importance (Scriven, 1974b, p. 16)

To fulfill this checkpoint to a high degree requires three things, they are:

a) Congruency: The program as described (see Section 3.11 "The evaluatum") and the implementation plans for the program (see Section 3.21, Standard 2: MARKET) must be congruent with what actually happened.

b) Causation: If the Program proponents say that Program B caused some effect Y, then there must be a search to see if Y is present. This is simply a check for the presence of Y, it is not an explanation for the cause of Y. If Y is not present, then the Program's proponents claim that Program B caused Y is not correct. If Y is present then their claim may be correct. The next checkpoint, CAUSATION, determines if or if not it was Program B which caused Y.

c) Justice: The moral dimensions of the program must be observed. For example there must be a search for any injustice, unhappiness, cruelty (and their opposites) which is part of the

program--as opposed to being a side-effect of the program (Scriven, 1974b, p. 16).

Summary Version of PROCESS Checkpoint.

| Methodological Evidence | Substantive Fulfillment characteristics | Rating |
|-------------------------------|---|--------|
| Descriptive congruence check? | Passes with flying colors. | 4 |
| Causal clues check? | Appears satisfactory. | 3 |
| Instrument validity? | Reasonable risk. | 2 |
| Judge/observer reliability? | Significant omission(s). | 1 |
| | Inadequate | 0 |

Comment: All of the Evidence Sources should be used by the evaluator, rated separately and then combined into an overall performance rating for PROCESS.

3.337 Checkpoint 7: CAUSATION.

Whereas the PROCESS checkpoint was a presence check for a particular claimed effect (E) of Program B; CAUSATION attempts to determine if it was really Program B which caused E.

The problem of causation is extremely complex. As Kerlinger (1973, p. 393) wrote:

"... the study of cause and causation is an endless maze ... we agree that causal laws cannot be demonstrated empirically, but we are equivocal about thinking causally. There is little doubt that scientists do think causally and that when they talk of a relation between "p" and "q" they hope and believe that "p" causes "q".:

Scriven (1966b, p. 454) wrote:

"... Can we not proceed further and define "possible causes" in terms of some combination of necessary and sufficient conditions, these being interpreted as simple regularity notions? The answer appears to be that we cannot."

With the limitation that when the word "cause" is used, it means "possible cause", Scriven provides the following useful "developmental sequence" with the comment (1966b, p. 455) "[it] does not establish the common idea that later members are simply complex combinations of the earlier ones."

1) Basic Experimental Case: If whenever and however C is produced, E occurs; and E never occurs unless C is produced then C caused E. C is produced at random intervals.

2) Basic Observation Case: C is observed to occur on various occasions and is accompanied or followed by E, and E never occurs on other occasions. C is the cause of E if we could reduce the situation to a basic experimental case, i.e. if we could randomly control the appearance of C. If this is not possible then the problem is to eliminate the possibility that some other factor X is causing both C and E.

3) Compound Causes: Not only C but also D is needed to bring about E. Scriven refers to C and D as causal factors or co-causes of E.

4) Multiple Causes: "If C and D are each sufficient to bring about E, and nothing else is, then whichever occurs is the cause. If both occur, one of them may not have had any effect on this occasion, a possibility which we check by examining the situation for the presence of known intermediate links which characterize the modus operandi of C and D, i.e. any sets of conditions "C₁ or C₂ or . . ." (or "D₁ or D₂ or . . .") which are necessary for C (or D) to act as the cause of

E. This test does not apply where no such links are known, and since that it is not logically necessary that there be any (C and E may be adjacent links in the chain, or differ only from a certain descriptive standpoint, or represent "action at a distance"), the test is not part of the meaning, . . ." (Scriven, 1966b, p. 455-56).

It is "Multiple Causes" which is of particular interest to this thesis. In S-GRADING, the absence of control groups and the inability of a formative evaluator to randomly assign particular treatments "c" eliminate proving the "Basic Experimental Case", the "Basic Observation Case" and "Compound Causes." However, the modus operandi (MO) methods briefly described by Scriven in "Multiple Causes" and elaborated by him in Evaluation Perspectives and Procedures (Scriven, 1974b, p. 68-84), provides a useful method of determining a causal inference pattern. As he point out, the MO method is routinely used by anthropologists and historians. (Scriven, 1974b, p. 68).

Briefly, the problem is to determine what caused E. The method of determining this is:

Step 1: A, B, C, D . . . can sometimes cause E.

Step 2: Nothing else is known to cause "E".

Step 3: B but not A, C or D was present.

Step 4: The MO of B (i.e. $b_1, b_2, b_3 \dots, b_n$), which is highly distinctive, was present.

then Step 5: B probably caused "E". (Scriven, 1974b, p. 73).

For example, if a student teacher uses a particular method of discipline then possible sources for that method are; his upbringing

(parents, home, school, etc.), his sponsor teacher in the school, his own reading in his education (or other) classes, his contact with other student teachers, his University advisor, and rules of the school he is in. If he was observed to consistently use the rules of the school (the effect) then a search could be made for why he did use these rules of discipline. If it were found that the principal had carefully discussed the rules of discipline used in his school, that he had circulated a written list of these rules, that the sponsor teacher had made a point of telling the student teacher about the rules and the student teacher said that he used those rules because they were the school rules, then probably it was the rules of the school which determined (i.e. was the cause of) the discipline techniques used by the student teacher.

This search for the complete MO obviously would become more complicated if it was found that the discipline rules of the school were the same as those espoused by the student teacher's university advisor.

Summary Version of CAUSATION Checkpoint.

| Methodological Evidence | Substantive Fulfillment Characteristics | Rating |
|---------------------------------------|--|--------|
| Randomized experimental design? | Impeccable | 4 |
| Quasi-experimental design? | Good bet | 3 |
| Ex post facto? Modus Operandi method? | Plausible bet | 2 |
| A priori interpretation of | Weak bet | 1 |
| correlational data? | Hopeless bet | 0 |

Comment: A randomized experimental design is the best source of evidence. The other methods are ranked in descending order of preference.

3.338 Checkpoint 8: STATISTICAL SIGNIFICANCE.

This is the standard that determines if the measurements taken on the project were appropriate, and if they were analysed at an acceptable level of statistical significance. Scriven (1974b), p. 15) says "This requires no great sophistication, and it is frequently the only work of sophistication in an evaluation design." In the next paragraph he comments, ". . . it is all too easily obtained without the results having any educational significance . . .". In other words, simply reporting sophisticated statistical test results, does not mean an evaluation has been carried out.

Summary Version of STATISTICAL SIGNIFICANCE Checkpoint.

| Methodological Evidence | Substantive Fulfillment characteristics | Rating |
|---|---|--------|
| Appropriate analysis? Appropriate significance level? | Flawless analysis, astronomical significance | 4 |
| | High significance, well-tested | 3 |
| | Reasonably significant | 2 |
| | Marginal significance | 1 |
| | Not shown to be significant | 0 |

To determine an overall rating for the standard for goal achievement it is proposed that the individual ratings for each checkpoint be added together and averaged. This average is then used as the rating for goal achievement. It is impossible for a product to receive a high rating in this way, and not be fulfilling the worthwhile goals. Each checkpoint has been rated in terms of its own relevance to the goals and on the basis of good data, (i.e. good evidence of good performance), therefore the sum of the checkpoint ratings will reflect this.

The actual rating procedure for each checkpoint and for the combined rating of overall worth will be discussed later in this chapter.

Figure 3 has been prepared to summarize the preceding discussion. This Figure includes the standard of COSTS.

Figure 3. The Standards and Their Ratings

| STANDARD | CONSIDERATIONS | RATING FOR THE STANDARD |
|--|---|---|
| 1) The worth of the Program Goals. | To be worthwhile, the goals should: -enable one to fulfill higher goals. -not conflict with another principle of our way of life. -lead to beneficial consequences. -not warrant having an exception made to the goals in this situation. | Maximum priority, a desperate need, extremely worthwhile goals 4 Great importance, very worthwhile goals 3 Probably worthwhile goals . . . 2 Possibly worthwhile goals . . . 1 No good evidence of the goals having any worth 0 |
| 2) The degree of goal achievement. | -MARKET -TRUE FIELD TRIALS -TRUE CONSUMERS -LONG TERM -SIDE EFFECTS -PROCESS -CAUSATION -STATISTICAL SIGNIFICANCE | Completely fulfills the goals 4 Fulfills almost all of the goals 3 Fulfills the most important aspects of the goals . . . 2 Fulfills some but not all of the important aspects of the goals 1 Does not appear to be directed to fulfilling the goals 0 |
| 3) COSTS (not included in this thesis) | -Comprehensive cost analysis -Expert judgement of costs -Independent judgement of costs -Costs for all competitors | Breakthrough for comparable products 4 Significantly lower than comparable products 3 Reasonable for comparable products 2 Probably high for comparable products, or somewhat incomplete data . . . 1 Apparently excessive for comparable products of data inadequate 0 |

If the evaluator was going to examine costs, or aim at an evaluative conclusion of "commendatory" or higher (see Table 1), then cost standards would have to be included. The term "cost" is extremely broad, and includes not only dollar costs, but psychological cost, lost opportunity cost, social costs, etc. (Scriven, 1974b, pp. 19-21; Haller, E.J., 1974). In this evaluation cost was not a major consideration as a separate standard and will not be included in subsequent discussions.

The final comment in this section has to do with the type of standards. In Section 2.40 of this thesis, standards were identified as being of intrinsic or extrinsic value. According to the definitions of these standards (see Sec. 2.40) a process or product cannot have intrinsic value; it can only produce feelings and experiences which may have intrinsic value. Both the worth of the goals for Program B and the degree of goal achievement are standards of extrinsic value. In this investigation, goal worth is a standard of instrumental value because it leads to the accomplishment of higher level goals or needs (i.e. being a good teacher); and goal achievement is a standard of instrumental value because it is a rank or grade of the program's effectiveness in fulfilling the goals.

3.40 Specification of the Class of Comparison in grading.

As stated in chapter II, sec. 2.34 and 2.60 the class of comparison should be some ideal form of the program (or the best of all real or imagined similar programs). The "idealized program" referred to in Chapter II is not only what the client hopes to see, but also

what he should hope to see. In other words, the evaluator cannot simply accept the client's expectations of the program as the standards by which the characteristics of the program are judged. This is particularly true in education, where the program developer's goals should be congruent with those of an enlightened public.

Therefore it is the decision of the evaluator as to the criteria defining the standards. He must be able to justify his ratings of criteria and standards to his client. The evaluator can realistically define the acceptable levels through expert opinion, logical analysis and/or through empirical studies by him or others. In other words the program is compared to predetermined standards of excellence (Stake, 1967, p. 18) based on the characteristics of the best of all real or imagined similar programs.

3.50 Determining the Good and Bad Characteristics of the Evaluatum in S-Grading

3.51 Gathering Data on the Goal Achievement Characteristics

Table 2 represents a calendar of events in the evaluation of Program B. It shows the date of the event, the instrument or technique used (including verbal and written presentations by the evaluator), the group being questioned or being addressed and the checkpoints being investigated. The instruments used, and the questioned asked in the interviews, etc., are appended to the thesis.

Table 2. Calendar of Data Collecting Events

| Date (1975) | Instrument or Technique or Presentation | Group being Questioned or Receiving | Checkpoints Being Investigated |
|----------------|--|---|---|
| Jan 15 | Questionnaire (App A) | U.B.C. Staff | True Consumer |
| Jan 31 | Interview (App B) | Principal and Vice-Principal | True Consumer Side Effects Process True Field Trials Market |
| Feb 3 | Likert Attitude Scale (App C) | Student Teachers | Process Comparisons True Consumer |
| Feb 4 | Reaction form (App D) | Student Teachers | Free Ranging Search- an open ended questionnaire |
| Feb 7 | Staff teaching perform- ance (form A) (App E) questionnaire | Student Teachers | Process |
| Feb 12 | Written submission from the teachers | the evaluator and U.B.C. staff | Side Effects Process True Consumer |
| Feb 19 | Verbal Presentation | U.B.C. Staff | True Consumer Process |
| Feb 27 | Questionnaire on Alternate Placement Appendix F | Student Teachers | Process Side Effects |
| Feb 28 | Verbal Presentation | U.B.C. Staff | True Consumer |
| Feb 28 | Likert Attitude Scale (App C) | Student Teachers | See Feb. 3. |
| March 4 | Evaluator chaired meet- ing between U.B.C. teach- ers and student teachers and verbal presentation. | U.B.C. Staff Teachers Vice-Principal Principal | Process Side Effects |
| March 6 | Reaction form (App D) | Student teachers | See Feb. 4. |
| March 7 | Staff teaching perform- ance (Form B) (Appendix G) Questionnaire. | Student teachers | See Feb. 7. |

Table 2 (Cont.)

| Date (1975) | Instrument or Technique or Presentation | Group being Questioned or Receiving | Checkpoints Being Investigated |
|----------------|---|---|---|
| March 20 | Interview (App H) | Student Teachers | True Field trials True Consumer Comparisons Side Effects Process-Congruency, Justice |
| March 25 | Interview (App H) | Student Teachers | See March 20 |
| March 27 | Verbal presentation | Student Teachers | True Consumer |
| March 27 | Questionnaire (App I) | Student Teachers | Process-congruency |
| April 2 | Interview (App I) | Student Teachers | See March 20 |
| April 8 | Observation of Student- teaching (App J) | 6 Student Teachers | Process Causation-Modus Operandi method |
| April 9 | Observation of Student- teaching | 4 Student Teachers | See April 8 |
| April 10 | Observation of Student- teaching | 2 Student Teachers | See April 8 |
| April 10 | Questionnaire (App K) | Teachers | Side Effects |
| April 11 | Observation of Student Teaching | 2 Student Teachers | See April 8 |
| April 13 | Interview (App L) | Program Director | Need Market Process True Field Trials True Consumer |
| April 15 | Questionnaire (App M) | School's pupils: grades 5 & 6 | True Consumer Process Side Effects |
| April 29 | Interview (App N) | University Field Director | Need Side Effects Process True Consumer True Field trials |

Table 2 (Cont.)

| Date (1975) | Instrument or Technique or Presentation | Group being Questioned or Receiving | Checkpoints Being Investigated |
|---|--|---|---|
| April 29 | Interview (App O) | Director of Elementary Education | Side Effects Process |
| April 30 | Interview (App O) | Director of Student Teaching | Side Effects Process |
| May 2 | Interview (App O) | Associate Dean | Side Effects Process |
| End of Academic year; evaluation--continues with different student teachers; different school. | | | |
| Sept. 19 | Questionnaire (App P) | Student Teachers | Side Effects |
| Sept. 25 | Questionnaire (App Q) | U.B.C. Staff | Process (descrip- tion of courses) |
| Oct. 3 | Questionnaire (App S) | Student Teachers | Process - evalua- tion of U.B.C. Staff teaching performance. |
| Oct. 24 | Questionnaire (App R) | Teachers | Side Effects Process True Consumer |
| Oct. 31 | Interview (App T) | School Board | Side Effects True Consumer Process |

In addition to the information collected as indicated in Table 3,
the author also attended Program B's directors' staff meetings on
the following dates:

January 15, 22

February 12, 19, 26

March 5, 12, 20

April 10

Attending these meetings provided information on the checkpoints of side-effects, process, market, and true consumers.

Other data collected were census data (1971) for the area served by the school, and the original application forms of the student teachers. These data helped in determining a rating for the checkpoint of time field trials.

3.52 Determining if the Characteristics of the Program are Good or Bad.

3.521 Rating the Worth of the Goals

The characteristics of the goals are examined in terms of their worth. Therefore in this section arguments will be presented outlining the procedures used in determining a rating for the worth of the goal for training people to be good teachers. A complete validation of the rating given at the end of the arguments for the need for teacher training has not been documented here. The conclusions reached can be judged by their reasonableness and not by their completeness. This is the type of approach one would use for his client in evaluation.

Program B has stated that it is attempting to contribute to the training of good teachers. Perhaps it seems trite to ask at this point "do we need to train and educate people to be good teachers?"

but the question will be posed to maintain continuity. The other side of the question is "or can any enlightened but untrained person step into a classroom and teach?" If the question of needing teachers at all should arise, then it would be necessary to proceed through a complete validation until everyone was generally agreed that teachers were needed. There is no point in asking "Do we need to train people to be good teachers?" if there is widespread disagreement about needing teachers at all.

Assuming that it is generally agreed that we need good teachers then we must address the problem of somehow establishing if training people to be good teachers is a worthwhile goal. The best way to do this would be empirically i.e. randomly assigning a group of prospective teachers to schools where they immediately begin teaching. After a few years compare them to another group who began teaching at the same time but who had been through a program of teacher training. Lacking this, however, one could look (cautiously) at present practice and enlightened present thought on the subject.

Present practice indicates that people must be trained to be good teachers. Present thought suggests the same thing. As Broudy wrote:

"That public school instruction can be manned by bright and devoted amateurs is the illusion which misguides many of the critics of teacher preparation.

.....
The salvation of our society will depend on the ability of education to exploit the blessings of technology in behalf of what makes life worthwhile, viz., the possibility of high grade individual experience in something called 'the good life'. That solving these problems will

require less than a generation of professional educators trained and educated in fully developed professional schools, I find impossible to believe." (Broudy, 1965, p. 415)

Though there are no specific formulas there are general rules, outlined in section 3.21 which will validate the worth of our goals with as much accuracy and objectivity as most measurements in education.

In making our judgement it is convenient to rate the worth of the goals on the following five point scale.

| | |
|--|---|
| Maximum priority, a desperate need, extremely worthwhile goals | 4 |
| Great importance, very worthwhile goals | 3 |
| Probably worthwhile goals | 2 |
| Possibly worthwhile goals | 1 |
| No good evidence of the goals having any worth | 0 |

The evidence suggests that there is more than just "Probably worthwhile goals" in a program to train people to be good teachers. However, as many schools have been operating with some untrained but functional teachers on their staffs it would be difficult to give a rating of "Maximum priority, . . .". Therefore the standard of goal worth for a program to train people to be good teachers is given a rating of "Great Importance" and a rating value of 3.

Notice that the worth of the goals, and its rating, are independent of the program being evaluated, in the sense that the goals have varying degrees of worth regardless of whether there is a program devised to achieve those goals.

3.522 Goal achievement and its rating

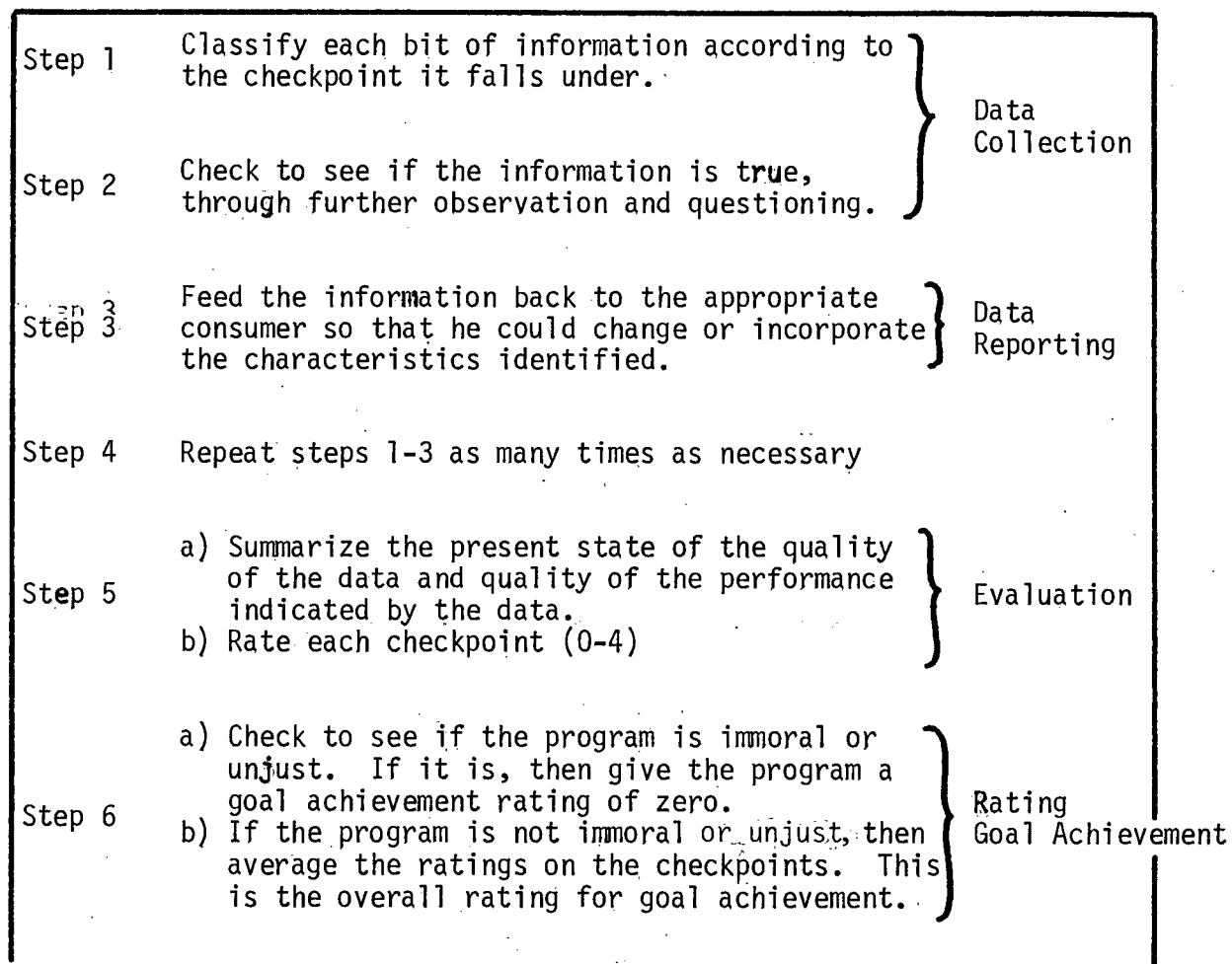
Formative evaluation has the responsibility both of providing valid information for program improvement and of raising questions about the merit of the program at particular points in time (Scriven, 1967, p. 41). In the evaluation of Program B information was collected, validated through observation and further questioning, summarized, and then fed back to the program participants so that changes could be made. The changes would either attempt to eliminate problems or attempt to incorporate advantages of the program. However, even though the program was still in the formative stage, an overall estimation of worth should be made. This could be done at natural breaks in the program (e.g. at Christmas or during the summer holidays of educational programs). The overall evaluation would reflect the state of the program at that point in time, and, therefore, so would the rating on each of the goal achievement checkpoints.

In other words in formative evaluation the evaluator would cycle through the checkpoints a number of times. Each time through he would delete previous problems that had been eliminated through changes in the program, checkmark possible contributions to the achievement of the goals, identify new problems and contributions, check to see if previously identified contributive features are still present (if still of use) and so on. Then, when he made his judgement of the overall worth, he would consider the state of each checkpoint, in terms of the quality of data (methodological) and the quality of performance on the data (substantive) to determine a rating.

To determine the overall rating for goal achievement a simple average is proposed with the conditions that the rating should not be calculated if the program is immoral or unjust.

The above steps are summarized in Figure 4.

Figure 4: Determining a Rating for Goal Achievement.

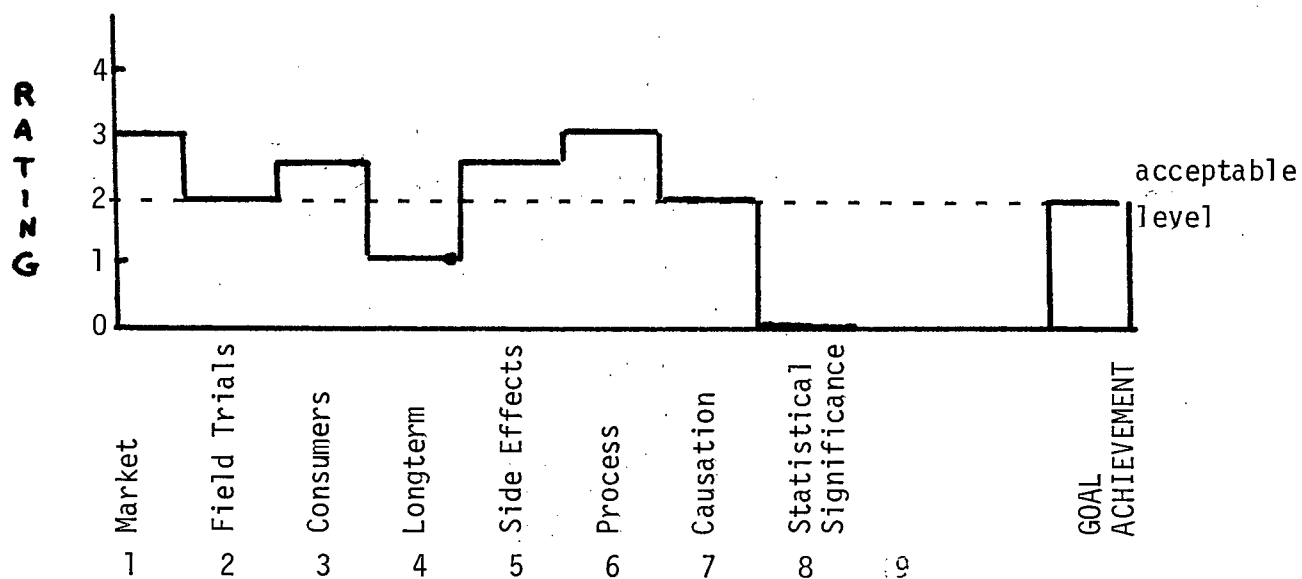


The data collected in the formative evaluation of Program B are considered confidential and cannot be reported here to illustrate the final rating given to the checkpoints.

Scriven (1974b, p. 31) has suggested the use of a "product

evaluation profile". His suggestion is used in Figure 5 to illustrate the ratings of each goal achievement checkpoint and an overall rating of goal achievement for Program B. (see also Figure 2).

Figure 5
Product Evaluation Profile
For Program B (to end of May '75)



Comments on the ratings

- 1) Market is important. Plans for reaching the market are unique, feasible and are economically viable. Detailed program description still needed for portability.
- 2) U.B.C. staff very enthusiastic and hard working, they are not an average group. Teachers seem reasonably average. Student-teachers were selected on basis of high marks. Many immigrant children with language problems at the school--made some student teaching

overly difficult.

- 3) Missing data on the gains by the school's pupils.
- 4) Informal follow-up plans, not finalized.
- 5) Good data on side effects. Most bad side-effects were eliminated by simple adjustments to the program. A number of unanticipated gains by Teachers and Student-teachers.
- 6) Student-teachers using techniques learned in their methods courses. Most are teaching satisfactorily. Some are considered by all consumers to be excellent. Program very conscious of justice and morality. Some difficulties in fulfilling all that was expected due to sketchy nature of program description.
- 7) Modus operandi check on Student-teachers' teaching but more data needed.
- 8) No statistical data collected.

3.60 Overall Determination of Worth

Each standard, goal worth and goal achievement, has now been rated. The problem now becomes to make an overall determination of the worth of the educational endeavor (Program B). This depends on two things (Chapter II, sec. 2.80):

- a) The degree to which the evaluatum fulfills the standards applied to it.
- b) The relative precedence of the standards.

Considering "b" first: The standard of Goal Worth takes precedence over the standard of Goal Achievement. If the goals of the program are

not worthwhile, then, regardless of how well those goals are being achieved, the program cannot receive an overall "good" rating.

Figure 6 summarizes this interpretation of goal worth taking precedence over goal achievement.

Figure 6

Comparing Goal Worth and Goal Achievement

GOAL WORTH RATING (See Fig. 2)

| | | Not Worthwhile | Probably Worthwhile | Extremely Worthwhile | | |
|---|--------------------------------|-------------------|---|-------------------------|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| G O A L A C H I E V E M E N T (See Fig. 2) | Goals not being achieved | 0 | | | The goals are worthwhile; but the Program has not been shown to be achiev- ing the goals; or the program has been shown to be morally unacceptable | |
| | | 1 | | | | |
| | Achieves some goals | 2 | The program goals have not been shown to be worth- while. The program is not fulfilling a useful function. | | The goals are worthwhile; and the program has been shown to be achieving the goals in a morally acceptable manner. | |
| | | 3 | | | | |
| | Achieves all goals | 4 | | | | |

If the rating of goal worth is greater than or equal to "2" and the rating of goal achievement is greater than or equal to "2; then the Program is at varying degrees of readiness for general distribution and could now undergo a summative evaluation.

If the rating of goal worth is greater than or equal to "2" and

the rating of goal achievement is less than "2" then the program needs more development before it is generally distributed. The right hand side of Figure 6 (i.e. Goal Worth has a rating greater than or equal to "2") is expanded in Figure 7 to illustrate the different descriptions which result from the evaluation.

If a program has a goal achievement rating of "2" or less on Figure 7 then it may be unworthwhile; or that is, the program may have undergone an evaluation and found to not be achieving the goals; or the program may be unworthwhile because it is found to be using morally unacceptable methods; or it may not have undergone an evaluation (or have undergone a very poor evaluation) in which case no judgement can be made of its overall worth.

If the program has a goal achievement rating of "2" or more on Figure 7 then it is worthwhile and increases in worth, as it is shown to achieve more of the goals and as the goals increase in importance.

Program B has a goal achievement rating of about "2" and a goal worth rating of "3". Therefore it could be described as a worthwhile program, fulfilling important goals to an acceptable level. However, this statement has the following restriction. The evaluative conclusion being aimed for was an "overall evaluative conclusion" (see Table 1, Chapter II). Cost data were not taken into consideration, nor was the program extensively compared to other teacher training programs. Therefore, rewording the overall determination of worth in terms of the "overall evaluative conclusion", the description becomes:

Figure 7: Expansion of Figure 6

| | | GOAL WORTH RATING | | |
|-------------------------------|---|--|-------------------------------|------------------------------------|
| | | 2 | 3 | 4 |
| GOAL ACHIEVEMENT RATING | 0 | Probably worthwhile goals but: | Very worthwhile goals but: | Extremely worthwhile goals but: |
| | | 1) No data on the program or 2) The data indicates the program is not achieving any part of the goals or 3) The program is morally unacceptable. | | |
| | 1 | Probably worthwhile goals but: | Very worthwhile goals but: | Extremely worthwhile goals but: |
| | | 1) Poor data on the program or 2) Data indicates the program is doing a poor job of achieving the goals or 3) The program has serious moral shortcomings. | | |
| | 2 | Probably worthwhile goals and: | Very worthwhile goals and: | Extremely worthwhile goals and: |
| | | 1) Acceptable data and 2) Data indicates the goals are being acceptably achieved. | | |
| | 3 | Probably worthwhile goals and: | Very worthwhile goals and: | Extremely worthwhile goals and: |
| | | 1) Good data and 2) Data indicates the goals are almost completely being achieved by the program. | | |
| | 4 | Probably worthwhile goals and: | Very worthwhile goals and: | Extremely worthwhile goals and: |
| | | 1) Excellent data and 2) Data indicates the program is completely achieving the goals. | | |

"Program B acceptably fulfilled the goal for producing student teachers, who were competent in teaching (for this level of their teacher training). The student teachers benefitted from Program B; and there were no harmful effects on the program participants."

Note that the above description does not claim that Program B is the best program for training teachers at the third year level. To make that statement it would be necessary to fully evaluate the comparative programs, including the standard of cost, and then rank order them and program B to determine which was best. This would be the job of a summative evaluator. The formative evaluation illustrated by this thesis has simply provided Program B's developers and supporters (financial and otherwise) with a "green light" on subsequent development.

3.70 Extension of the Model

3.71 Introduction

Chapters Two and Three have explicated a logical evaluation model based on R-GRADING. This section "The Extension of the Model", has been included to show that the model developed for R-GRADING can be logically and easily extended to include the most comprehensive evaluations, that is, RANKING the evaluata in a class of comparison to determine if it is the best evaluatum.

Ranking involves grading an evaluatum according to certain standards and then comparing that overall rating with the ratings obtained by other similar programs graded according to the same

standards. The evaluata are then ranked according to how well they met the standards. This tells the client for the evaluation which evaluatum is the best, assuming each evaluatum was evaluated to the same degree. Since each evaluatum must have gone through S-grading before it was ranked in the class of other relevant evaluatum, the evaluation can also inform the client of the overall worth of each evaluatum.

There are three questions which must be discussed for the extension of S-grading to Ranking. Firstly, "What precedence do these standards take in determining the overall worth or merit of the evaluata;" and, finally, "how can we rank the evaluata?"

3.72 The Standards to be Used in Ranking

In Chapter 3 the standards of Goal Worth and Goal Achievement were used as standards for S-grading. The logical question then is "could we use just those two standards for Ranking?" The answer is "yes"; but this assumes that the cost of the project is not of concern. Unfortunately cost is usually of concern. Obviously if programs X and Y fulfilled the standard of Goal Worth to the same degree, and the standard of Goal Achievement to the same degree, yet Program X cost more than Program Y, the Program Y would be a better program. Therefore, though it is possible to rank competitive programs only on the basis of Goal Worth and Goal Achievement, it is more reasonable to rank them by also considering the standard of Cost.

The standard of Extended Support may also be taken into consideration. However this standard is not as necessary as the three previously mentioned (Goal Worth, Goal Achievement and Cost). Extended support and Cost were incorporated into Scriven's (1974b) "Product Checklist". They will now be described in more detail.

3.721 Costs

Scriven (1974b, p. 21) provided three criteria for cost data. The two which are applicable to this discussion are presented below.

a) Cost data must be Comprehensive.

This means that the evaluators should investigate such costs as: covering maintenance costs, capital costs, psychic costs, dollar costs, costs of in-service updating of needed helpers, and so on. Scriven (ibid) also suggests that:

"A qualitative cost-effectiveness analysis should be attempted where possible, and, if it is impossible, then cost-benefit analysis should be done as systematically as possible."

b) Cost data must be verified.

All the cost estimates and real costs should be independently verified, preferably by an accountant skilled in estimating costs of educational products.

A summary version of this standard could be similar to the following: (Scriven, 1974b, p. 28)

| <u>Considerations</u> | <u>Rating</u> | |
|-------------------------------------|--|---|
| ___ Comprehensive cost analysis? | - Breakthrough for comparable products | 4 |
| ___ Expert judgement of costs? | | |
| ___ Independent judgement of costs? | - Significantly lower than comparable products | 3 |
| ___ Costs for all competitors? | - Reasonable for comparable products | 2 |
| | - High for comparable products or somewhat incomplete data | 1 |
| | - Excessive for comparable products or data incomplete | 0 |

A more extensive discussion of "Costs" has been provided by E.J. Haller (1974, pp. 406-450).

3.722 Extended Support

It is highly desirable that there be systematic procedures for updating or upgrading a product once it is on the market in the light of new information which could improve the product. As Scriven (1974b, p. 21) points out: ". . . this implies the necessity for a systematic continuing procedure for collecting field data."

This standard is desirable but not necessary. Its presence should be a plus for the program, but its absence should not count against the program. It would be of great importance if there was a strong need for the program and there were no available alternatives. In that situation there should be a continuous collection of field trial data.

A summary version of this standard could be similar to the following: (Scriven, 1974b, p. 28).

| <u>Considerations</u> | <u>Rating</u> |
|---|-----------------------------------|
| ___ Postmarketing data collection? | - Excellent and comprehensive 4 |
| ___ Postmarketing system for improvement? | - Good and fairly comprehensive 3 |
| | - Minimally acceptable 2 |
| ___ In-service training? | - Weak -- less than adequate 1 |
| ___ Up-dating of aids? | - Negligible -- apparently none 0 |
| ___ New users and user data? | |

3.73 The Precedence of the The Standards in Ranking

As argued in Chapter 3, the standard of Goal Worth takes precedence over all the other standards. If the Goals of the evaluatum are not worthwhile then the evaluatum should not exist. When comparing different evaluata, it is assumed that the goals are the same for all of them. The criterium of similar goals is necessary in order to compare the evaluata. There is no point in comparing Program X to Program Y if they are fulfilling different goals. For example, one could argue that we need methods of transporting people from Point A to Point B and then give a rating to this goal; and then consider the proposed (or operating) methods of fulfilling that goal. Possible methods could be: by private car, by bus, by train or whatever. The goal, regardless of the method, remains the same. What will determine the individual method's overall worth, (assuming the goals are worthwhile), and hence their rank order, will be the degree to which they each achieve the goals, and the cost of each. Therefore, assuming that the evaluata

being compared are all trying to fulfill the same goals, and assuming that the goals are worthwhile; then the standards of goal achievement and Cost will determine each program's relative worth (in the class of comparison) for the purpose of ranking. (The role of Extended Support will be discussed later).

It does not appear that a simple averaging of the rating for Goal Achievement and of the rating for Cost can be used to determine which evaluatum is best. Figure 8 will help clarify this statement. Figure ~~10~~⁸ illustrates the same procedure that was used in determining the overall worth in S-grading using Goal Worth and Goal Achievement. The matrix has been divided into four quadrants. It is assumed in the following discussion that the evaluation was complete, so that a low rating is not due to lack of information on that standard. Each quadrant will now be described:

Quadrant 1: (high cost, low goal achievement)

If the cost is high and the evaluatum is not achieving the goals, then it seems reasonable to reject it from further consideration. The cost of a program is a relative standard, i.e. high cost means that an evaluatum costs much more than some other evaluatum within the available resources. If all the evaluata being compared fall into this quadrant, then they should all be rejected, and a new method for achieving the goals should be searched for.

Quadrant 2: (high cost, goals being achieved)

In this quadrant, the cost is high, but the evaluatum is achieving the goals. The question is: "Why is the evaluatum costing

Figure 8.

Combining the Goal Achievement and Cost Ratings

Rating for Goal Achievement

| | | | | | |
|---|---|--|---|--|---|
| | 0 | 1 | 2 | 3 | 4 |
| R a t i n g f o r c o s t | 0 | <u>Quadrant 1</u> Reject evaluatum. | | <u>Quadrant 2</u> Why is the evaluatum costing so much? | |
| | 1 | | | | |
| | 2 | <u>Quadrant 3</u> Why is the evaluatum not achieving the goals? | | <u>Quadrant 4</u> The evaluatum is acceptable. | |
| | 3 | | | | |
| | 4 | | | | |

so much?" It may be achieving the goals only because of a high investment of resources and materials. An evaluatum which has an equal or greater goal achievement rating (i.e. a rating of "2" or more) and lower cost is obviously superior. If all the evaluata fall in this quadrant, then they should be closely examined as to methods of cutting the costs.

Quadrant 3: (low cost, low goal achievement)

Here the cost is acceptable but the evaluata which fall in this quadrant are not achieving the goals. The reasons an evaluatum

in quadrant 3 may not be achieving the goals may be either because it is not a "good" program (i.e. poorly designed, etc.); or it may be because it has not utilized sufficient resources. If all the evaluata being compared fall in this quadrant then they should be closely examined to determine if it is due to being poorly designed programs, or due to insufficient use of resources.

3.74 Ranking the Evaluata

Quadrant 4: (Reasonable to low cost, reasonable to high goal achievement)

This quadrant defines the area of acceptability. It is only here that a simple average of the ratings for cost and goal achievement should be calculated to determine the rank order of the evaluata. The evaluatum with the highest average rating is the "best buy" (see Table 1, Chapter 2). If two or more evaluata have the same highest average rating it is suggested that that one with the highest goal achievement rating be chosen as the best. If there is still a tie for the "best buy" (i.e. two or more have the same goal achievement rating and the same cost rating) then that one with plans for "extended support" should be given the highest rating. Thus, the standard of extended support is seen as a final tie-breaker.

If none of the evaluata being ranked fall in Quadrant 4 of Figure 8, they can still be ranked either with respect to costs or with respect to goal achievement. However, this rank order will establish the best of a bad lot and does not mean the "best" evaluatum is a "good" evaluatum.

CHAPTER IV

Data Collection Techniques in the Application of the Evaluation Model.

4.00 Introduction

This chapter catalogues the data gathering techniques used in the application of the evaluation model described in Chapters Two and Three. All of the questionnaires and other techniques, used to gather data in the evaluation of Program B, have been appended to the thesis.

The first part of Chapter Four illustrates two stages in the data gathering procedures. The first stage was an overall portrayal of Program B and the second stage was data collection on specific aspects of the Program. These stages are related to the standard of Goal Achievement by Table 3. The final part of the Chapter illustrates how the collected data was recorded and validated.

4.10 Initial Stage of Data Collection

The evaluation of Program B began after the program had been in operation for three months. Therefore the initial stage of data collection had two purposes. Firstly to determine which check-points in Scriven's Evaluation Checklist should be considered first; and secondly, to provide a general idea of the benefits of the Program to the student teachers at this stage in the Program's development.

The first purpose was accomplished by asking each of the members of the development team which of the Checkpoints on Scriven's Evaluation Checklist should be considered first. (see Appendix A) The responses to the Checklist identified major areas of interest and concern, from the point of view of the developers, and provided direction for setting initial priorities in data-collection.

The second purpose, that is to determine the initial benefits of the program, was accomplished by questioning three groups of people - the student teachers, the school administrators and the school's pupils. The student teachers were asked to respond, on a five point scale, to a set of attitude statements on the teacher training process. The attitude scale (see Appendix C) was developed using the usual techniques (Likert, 1932). It was administered twice, over a period of one month, to the student teachers in Program B, to the student teachers in the equivalent year of the regular teacher training program, and to the student teachers in the equivalent year of the transfer teacher-training program (see U.B.C. 1974-75 calendar for more details of the programs). This provided a means of comparing the attitudes of the program B student teachers with the attitudes in the two comparison programs (the regular and the transfer program). Administering the attitude scale twice gave an estimate of stability of group response on the questionnaire.

The student teachers were also asked to fill in a "reaction form" (see Appendix D). This provided the evaluator with their overall rating of the merit of the program, and with a great deal of information on general complaints and praises.

The second group to be questioned was the school administration, that is, the principal and vice-principal of the school. They were interviewed according to the list of questions given in Appendix B.

The third group to be questioned was the pupils of the school. This was accomplished by talking to individuals or small groups during the recess or lunch breaks. The candid responses of this group provided a great deal of general information, not only on the operation of Program B, as seen through the pupils eyes, but also on the general school climate.

There are two obvious groups omitted from this initial survey, the teachers in the school and the program developers. The teachers had been exposed to numerous surveys already. The School Board and the Provincial Government had both been unusually active in sending out questionnaires in the 1974-75 year. Further, an external group of evaluators, commissioned by the Faculty of Education at U.B.C., was evaluating all the alternate programs offered in teacher education in the Faculty. The latter group of evaluators had asked the teachers to fill in an extensive questionnaire on Program B just prior to the internal evaluation beginning. Therefore, it was considered prudent to delay giving any questionnaires to the teachers at the beginning and instead to rely on informal "chats" with individual teachers.

The program developers were also omitted from this initial data gathering because of their frankness in their weekly meetings. These meetings were attended by the evaluator.

This overview of Program B was valuable for a number of reasons. It provided general data on the feelings of all the consumers toward the program; it suggested major areas of concern or neglect for more detailed investigation; it allowed the evaluator to introduce himself and outline the general scheme of evaluation to most of the consumer groups; and it provided general feedback to all the groups involved, particularly to the program developers.

This initial stage of data collection could also be called a 'portrayal' stage of data collection. It is portrayal that Stake (1972a) advocated when he wrote:

"In the more ordinary and modest situation . . . we should limit our evaluation aims to what we can do and to what the client needs most. What many clients need is a credible, thorough representation of what the program is, including information about who likes what about it."

Portrayal is useful (see previous paragraph) but is often only a step toward an evaluation. (For an example of this approach see Gleadow (1975)). There must be a second stage of data collection in order to make conclusions of the worth of the program. An estimate of worth may be possible using portrayal or initial data collection, but it would be a guess and extremely difficult if not impossible to justify.

4.20 Second Stage of Data Collection

After this initial, 'portrayal' stage of data collection the data gathering was directed to specific areas of the Program. Each

of the techniques used will be now briefly discussed. The techniques have been categorized in Table 3 in this Chapter.

4.21 Evaluation of U.B.C. Instructors' Teaching Performance

Scriven (1973) has condensed many of the correlates of good teaching performance into a short list of primary indicators. It was these primary indicators that provided the source of many of the questions asked on Form A (Appendix E) and later in the program, Form B (Appendix G). These two forms were filled out by the student teachers directly onto IBM mark cards for ease of analysis. Seven student-teachers were randomly assigned to the evaluation of two instructors which were randomly paired for each student-teacher. In other words each instructor was evaluated by seven student-teachers. Subsequently, it was felt that this method was too time consuming. Therefore, the questionnaire was revised to the form illustrated in Appendix Q. In this final form, the student-teacher gave an overall rating to the teaching performance of the instructors and indicated if there were any instructors who were exceptions (either better than or worse than the rest of the group) to the overall rating. This simplified the filling out of the questionnaire, and the collection and interpretation of data, as well as identified exceptional instructors. The questionnaires also include an overall evaluation of the teaching performance of each instructor. Scriven (1973, p. 13) wrote about this overall judgment of merit:

"Even though we do not know exactly what its reliability is, at least (a) the results cannot so easily be wrongly interpreted and (b) the chances that it's negatively correlated with good teaching are probably lower than with any other single question (or combination of a few questions)."

This reasoning can also be applied to other areas of concern; as a result an overall judgement was asked for on a number of other aspects of the program. The evaluation of the U.B.C. instructors' teaching performance was a process check.

4.22 Questionnaire to Student teachers on their Alternate Placement (Appendix F)

During the course of Program B, the student teachers were placed in other schools for a two week practicum. The type of school was chosen by the student teacher. This questionnaire provided a check on the perceived success of that experience.

4.23 Interview of Student Teachers (Appendix H)

Three randomly chosen student teachers were interviewed over the space of a few weeks. Each interview took approximately one-half hour. The interviews were recorded on tape with the permission of the interviewee.

4.24 Final questionnaire for the Student Teachers (Appendix I)

The source of all the questions on this questionnaire was the original descriptions of the program (see Chapter 3 Sec. 3.11

"The Evaluatum"). Its purpose was to obtain a final subjective congruency check between what was said was going to occur, and what actually occurred.

4.25 Observation of Student Teaching

Fifteen of the student teachers were observed for at least one lesson each, by the evaluator, during their last classroom teaching experience. There were two areas of concern. Firstly a process check on classroom justice was needed and secondly an attempt to connect observed student teacher behaviors in the classroom and their Program B experience, that is, a causal check using the "modus operandi" method (see Chapter 3, section 3.337) was attempted.

Classroom justice data was collected by direct observation, with particular attention being paid to the student teachers' methods of discipline. It was felt that discipline techniques would be good indicators of the fairness with which the student teachers' dealt with the pupils. Therefore, each time a pupil or pupils were disciplined, a note was made on the reason for the disciplining, the type of discipline technique used, and whether the discipline technique was effective. The discipline techniques were categorized beforehand (see Appendix J). After the lesson, the student teacher was asked by the evaluator why certain techniques were used in particular situations. The pupils were asked in a questionnaire if they felt they were fairly treated by the student teachers (Appendix

M questions 3 and 7) to provide another check of classroom justice.

The modus operandi check was performed by recording all the techniques used by the student teacher and then interviewing the student teacher after the lesson to determine where he/she got the idea for a particular technique. (Was it his/her own idea? Was it a variation of an idea from a professor? a teacher? a textbook? Was it the sponsor teacher's idea? Was the technique freely used or was the student teacher told to use that technique?) Generally, the number of different techniques used in a lesson was quite limited, so that recording them and interviewing the student teacher after the lesson was not a time consuming or onerous task. The result of this was an indication of the contribution Program B made in transferring classroom techniques to the student teachers.

Evaluating the effectiveness of the student teachers was an area of particular difficulty. There were no instruments developed during the period of evaluation described by this thesis, though initial work on the instruments was begun in this period. However methods have now been developed and are being field tested (as this thesis is being written) in the second year of Program B. For the sake of completeness, the proposal and instruments being used are appended to this thesis as Appendix U).

4.26 Side Effects of Program B (Teacher's Questionnaire)

(Appendix K)

As noted in the covering letter to this questionnaire, this

was the only questionnaire given to the teachers. Its purpose was to determine the positive and negative side effects as perceived by the teachers.

4.27 Interview with Program B's Director (Appendix L)

This interview was tape recorded with the permission of the interviewee. It provided information on the initial planning of the program, and the reasons for its initiation. It was also used to determine the validity of the field trial data already collected.

4.28 Questionnaire for School Pupils (Appendix M)

This questionnaire was given to two grade five and two grade six classes. Each question was read to the pupils. If they were not clear on the meaning of a question they were asked to ask for clarification. It provided information on the teaching effectiveness of the student teachers, as perceived by the pupils.

4.29 Interview with U.B.C. Administrators (Appendix N)

The Field Director of Student Teaching, the Director of Elementary Ed, the Director of Student Teaching (Elementary), and the Associate Dean were all interviewed (separately) and each was asked the four questions written in Appendix N.

In the interviews and questionnaires, usually more than one

of the checkpoints of goal achievement was addressed. To illustrate this, Table 3 has been prepared. The standard of goal worth, and a category of "evaluative conclusion" have also been added to the table. The "evaluative conclusion" category refers to a question which asks the respondent to make an overall judgement of some aspect of the program. The numbers under each category of the Checklist in Table 3, refer to the questions of the interview or on the questionnaire which were obtaining information applicable to that category.

4.210 Other Information Sources

Table 3 suggests that there are areas which were neglected by the interviews and the questionnaires. Some of these, such as long term, statistical significance and critical comparisons were neglected because of the character of this evaluation. As stated earlier in this thesis, this study was a formative evaluation in S-GRADING. Therefore comparison groups were not of concern except for establishing a referant point for the attitude questionnaire (Appendix C). There were no long term plans being developed at the time of the evaluation so there was no data to collect on this checkpoint (note that it also received a very low rating on the profile (Chapt. III Section 3.522)). No measures which could have statistical procedures meaningfully applied to them were taken, except for the attitude questionnaire administered to the student teachers; therefore the checkpoint of statistical significance was

TABLE 3 Data Sources for the Goal Achievement Checkpoints *

| Appendix No. | Questionnaire | Interview | Respondant | Goal Worth | Market | Field Trials | True Consumer | Long-term | Side Effects | Process | Causation | Statistical Significance | Evaluative Conclusions |
|--------------|---------------|-----------|------------------------|------------|---------------|----------------------|---------------|-----------|---------------|--------------------------------------|-----------|--------------------------|------------------------|
| A | X | | Developers | | 1 | 2 | 3 | | 5 | 6 | 7 | | |
| B | | X | School Admin. | 1 | | | 6,7 | | 4,5 | 2,3 | | | |
| C | X | | Student teach. | | | | | | | All | | Yes | |
| D | X | | Student teach. | | | | | | 2 | 2 | | | 1 |
| E | X | | Student teach. | | | | | | | 1-14 | | | page 2 |
| F | X | | Student teach. | | | | 4 | | 3 | 1,5-9 | | | 10,11 |
| G | X | | Student teach. | | | | | | | 1-12 | | | page 2 |
| H | | X | Student teach. | 13 | 20 | 3,5,6, 8,9,16, 18,19 | 7,10,11 12 | 17 | 1,2,15,20 | 4,9,10,11 12,15,20 | | | |
| I | X | | Student teach. | | | 9,10 | | | 13 | 1-8,11,12 | | | 14 |
| J | Observation | | Student teach. | | | | | | | All | All | | |
| K | X | | School teach. | | | | | | 1-4 | 5 | | | |
| L | | X | Program Direc. | 2,12 | 1,3,4, 5,6,13 | 5,6,7, 10,11 | 8,10 | | 9 | | | | |
| M | X | | School pupils | | | | 5,9 | | 5,6,14, 16,17 | 1-4,6,7, 8,10,11, 12,13,15, 18,19,20 | | | 21 |
| N | | X | U.B.C. Admin-istrators | 1 | | 5 | | | | 2,3 | | | 4 |

* The numbers under each checkpoint refer to the question numbers on the questionnaires noted for each appendix.

given a zero rating.

The other checkpoints with apparently small amounts of data are market and true field trials. But that is because the data for these checkpoints were collected by means other than questionnaire or formal interviews. Most of the information for market (that is, the plans for implementation of the program) was collected from the original recruitment materials sent to the student teachers, and through the weekly meetings with the program developers, during which the program planning was revised and/or planned. Much of the true field trial information was obtained through (a) examining census data for the area served by the school and comparing this with the rest of the city; (b) examining the transcripts of marks and original application forms of the student teachers who applied and were accepted, as well as those who applied and were rejected; and (c) asking the administrators in the faculty of education to compare the ability of Program B's professors as a group, to the rest of the faculty.

The checkpoint of true consumers was unsatisfactorily developed. Gains or losses of the student teachers were measured through their marks on their teaching ability and courses, and through their attitudes and overall evaluations of the program. However, school pupil gains or losses, teacher gains or losses, U.B.C. professors' gains or losses, school administration gains or losses, University gains or losses, and School Board gains or losses were not established in sufficient detail. There was some evidence on this checkpoint for all of these groups however it was scanty. Only estimates of specific

gains or losses could be made.

4.30 Practical Procedures for the Organization of Data

A difficult and important task of the evaluator is to organize and condense the data he collects. The following procedure is presented as a feasible and workable method developed for the type of evaluation model discussed in this thesis. For each checkpoint (market, true consumer, etc.) a chart is prepared as illustrated in Figure 9 below:

Figure 9.

Format for Organization of Data

Checkpoint: _____

| Date | Data Source | Information | Validation Check | Comment |
|------|-------------|-------------|------------------|---------|
|------|-------------|-------------|------------------|---------|

Each of the headings in Figure 9 will now be discussed:

Date: The date the information was collected.

Data Source: The data source would indicate the person(s) who supplied the data; the technique used to collect the data (e.g. questionnaire, interview, informal discussion, etc.); and the question asked or responded to (identified by number if it was on a questionnaire or in a structured interview).

Information: An accurate summary statement of the information collected, or, if the information cannot be summarized, a reference to the raw data should be presented. There should also be a judgement on the part of the evaluator as to whether the information is positive, negative or neutral (equal amounts of positiveness and negativeness) with respect to a particular checkpoint in Need Fulfillment.

Validation Check: It is important that there be a check to determine if the information obtained was valid. This usually means further questioning of program participants or observers, and/or observation by the evaluator.

Comment: This could be a comment by the evaluator about what should be done with the information and whether or not what should be done has been done. For example, if the information was negative then the evaluator might make a note that that aspect of the program should be reported to the program developers and the problem should be eliminated. Later, he would check to see if it was eliminated. If the program was changed so as to eliminate the problem then the evaluator in his final report would use the information as a caveat to future program users,

but he would not give a lower rating to the checkpoint because of it.

An example (Figure 10) is now offered to clarify the above description. All the data is fictitious.

Figure 10.

Example of Format for Organization of Data

Checkpoint: Side Effects

| Date | Data Source | Information | Validation Check | Comment |
|---------|--|--|--|---|
| Feb. 15 | Student teachers, Questionnaire 1, Question 3. "Were there any unexpected positive or negative experiences in this program." | 12 of 20 student teachers indicated that the teachers asked them to participate in Parent-teacher night. The student-teachers thought the experience very worthwhile (judged <u>positive</u>) (see summary of all the comments on this, on page 125 of data notebooks) | a) The teachers were asked for their opinions of this participation by the student-teachers. They indicated they thought that it was valuable for the S.T.'s to participate - no negative feeling. b) Some of the parents who came to the parent-teacher interviews were questioned. They indicated they did not find the presence of the student-teachers distracting. | Report this to the program directors. (done Feb. 20) Decision was made to implement this in future programs. |

The advantage of this approach to recording the data is two-fold. Firstly, it tremendously simplifies the condensing of information when it comes time to make a final report; and secondly, it clearly outlines all the data which went into giving a rating to a particular checkpoint. This is extremely important, for the evaluator must be able to verify each conclusion, recommendation and rating he makes. This is relatively easy to do using the above format.

4.40 Confidentiality of Data

If the data sources are people or their records then strict confidentiality must be maintained. All of the people involved in an evaluation must be assured, and believe that assurance, that the information collected will remain anonymous as to the individual. In other words the evaluator could identify the source of information as the teachers, but not indicate what individual teachers had to say.

This can decrease the power of statistical techniques which rely on pre-post testing or correlation measures; but the loss in statistical power due to maintaining anonymity was thought to be justified by the gain in candid information.

CHAPTER V

Summary and Conclusions

5.00 Summary

This study has explored the possibility of devising a method of evaluation, based on the logical method of Paul W. Taylor (1961) in combination with the more practical approach of Scriven (1974b) which could be used in the evaluation of a teacher training program. The goal of the study was to bridge the gap between the theoretical concepts of Taylor (1961) and Scriven, and the practical problems of evaluating an educational program.

A fundamental assumption of this study was that evaluation is a logical procedure regardless of what is being evaluated. ". . . it is the logical method which a rational person would follow if he were trying to come to a careful reflective decision about the value of something." (Taylor, 1961, p. 4). The method of evaluation includes adopting a set of standards within a particular point of view, the operational clarification of those standards, specifying a class of comparison, gathering data on the characteristics of the evaluatum, and deducing from the data gathered and the standards chosen the degree to which the evaluatum fulfills the standards. The product of an evaluation is a conclusion that the evaluatum has a certain value, worth of merit. That conclusion depends on the relative precedence of the standards, and the degree to which the evaluatum

fulfills the standards.

The standards used in this study were: (1) The standard of goal worth and (2) the standard of goal achievement. These standards were used for broadly categorizing the two main components of any good educational program, (i.e. something worthwhile is being transmitted in a morally acceptable manner). The standards were used in the evaluation of a teacher training program (Program B) from the educational, institutional, moral and legal points of view.

The methodology developed in the study, was illustrated by applying it to the formative evaluation of a school-based, teacher training program. The evaluation was an example of S-GRADING. That is, the teacher training program was not directly compared to other, existing, similar programs; rather, it was graded according to the degree to which it fulfilled a set of appropriate standards.

When GRADING, the evaluator determines a set of criteria or fulfillment characteristics which operationally clarify the standards used in the evaluation. The criteria are derived from an ideal form of the program being evaluated, and will therefore differ from one type of educational product to the next. In this study, the criteria used were those suggested by Broudy in his paper on the professional preparation of teachers. (Broudy, 1965). The teacher training program (Program B) was evaluated by comparing both the goals of Program B and the method and degree of achieving those goals to Broudy's criteria; and, on the basis of that comparison, judging whether the

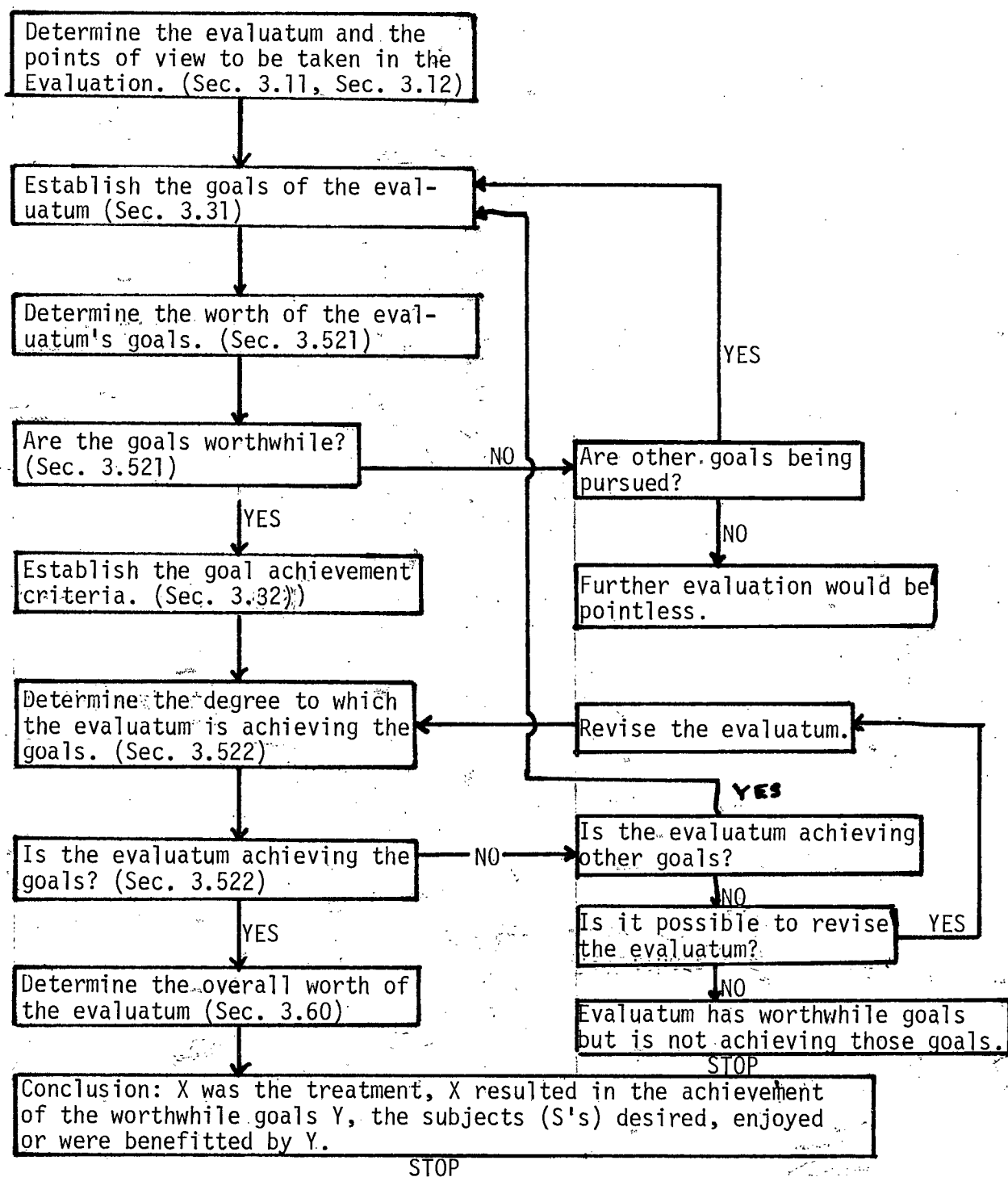
program's goals were worthwhile and being achieved in a morally acceptable manner.

If the client for the evaluation questions the final judgement of worth, then the evaluator would first show that the program met the criteria to the degree indicated by that final judgement of worth. If the client questioned the criteria used in defining the standards, then the evaluator must show that those standards were appropriate for the program being evaluated, and that meeting the criteria used to operationalize the standards indicated meritorious performance.

Figure 19 clarifies the evaluation procedures developed in this study. More information can be obtained by referring to the sections of the thesis noted in each step of the Figure. Figure 19 should be cautiously interpreted as a temporal sequence. In practise, the evaluator may enter the program after the goals have been established and after the program has begun operation. In that case, he would have to gather data on the goals and the goal achievement criteria at the same time.

Section 3.70 showed how the evaluation procedures developed for S-GRADING could be logically extended for RANKING. Appendix U has been included to summarize that extension.

Figure 11

Steps in S-GRADING in FORMATIVE EVALUATION

5.10 Conclusions

A general conclusion of this study is that Taylor's (1961) analysis of the method and product of evaluation can provide a framework for the evaluation of an educational program. More specifically when Taylor's analysis is combined with Scriven's extensive work and checklist, a feasible model of evaluation results, which readily produces a defensible, overall estimation of worth for an educational product.

Table 4 has been prepared to point out how the model of evaluation developed in this thesis differs from the more popular models of educational evaluation. On the left of the table are the limitations of some of the present educational evaluation models (see Section 2.10). On the right of the table are the contributions of this thesis toward overcoming those limitations.

Table 4.

Comparison of Evaluation Procedures.

| Limitations of some present models of evaluation (see Sec. 2.10). | Contributions of this thesis toward overcoming the limitations. |
|--|---|
| 1) <u>Tyler's Model</u> | |
| a) Focuses on the product and not on the process of evaluation. | a) Both product and process are of concern in the evaluation (see particularly Sec. 3.336, p. 63). |
| b) Focuses directly on measurable goals and objectives but makes no direct judgments on the worth of those goals | b) Determining the worth of the goals is one of the most important steps in the evaluation. If a program has worthless goals then it has no merit (see Sec. 3.21, p. 45, 3.60, p. 81) |

Table 4 (Cont.)

| | |
|---|--|
| <p>2) <u>Accreditation.</u></p> <p>a) Does not connect the observed characteristics of the program to their consequences on the program participants.</p> | <p>a) All data gathered is judged as to whether it is contributing to the achievement of the worthwhile goals. (Sec. 3.521, p. 75; Sec. 3.522, p. 78).</p> |
| <p>3) <u>Stufflebeam's CIPP Model.</u></p> <p>a) Provides data to decision makers but makes no decision on the worth of that data or on the worth of the product being evaluated.</p> | <p>a) Both the quality of data and the overall worth of the program are judged. (Sec. 3.32, p. 52, 3.60, p. 81)</p> |
| <p>4) <u>Stake's Model.</u></p> <p>a) It is not clear how the worth of the program goals is determined.</p> | <p>a) See 1b above and Section 3.521, p. 75.</p> |
| <p>5) <u>All Models.</u></p> <p>a) There is no clearly defined method of combining and weighting the data in order to arrive at an overall determination of worth for the program.</p> <hr/> <p>b) The models do not have a specific check on the moral acceptability of the program.</p> | <p>a) A clear, logical system of weighting the data of the criteria and combining these to determine the overall worth has been presented (Sections 3.522, p. 78, Sec. 3.60, p. 81.)</p> <hr/> <p>b) A check of the moral acceptability of the program is an integral part of the evaluation. If a program is immoral, then that alone is sufficient reason to make the program unacceptable. (Sec. 3.22, Sec. 3.60)</p> |

Scriven's recent checklist for evaluating products and proposals (Scriven 1974b) provided many of the checkpoints used in operationalizing the standard of Goal Achievement (Sec. 3.33). However, two changes were made in Scriven's Checklist to make it more congruent with Taylor's (1961) analysis of the process of evaluation. Firstly, the Checkpoint "Need" was reconceptualized and given separate status as a standard in the evaluation. As a standard it was reworded to "goal worth". It was reworded because "need" has a multiplicity of meanings (e.g., as an objective; a necessity, a deficiency, an obligation (Komisar, 1961)), and using any single meaning may be too restrictive. A goal, on the other hand, can be any intended outcome, the worth of which can be determined by applying the strategy outlined in Section 3.21. Goal worth was given separate status as a standard in the evaluation for two reasons: (1) To underline the importance of determining whether the goals of a program are worthwhile, (2) To clarify the contribution of "goal worth" in the overall merit rating of the program.

Secondly, in this thesis the Checkpoint of "Critical Comparisons" was deleted from Scriven's Checklist. Scriven explains "Critical Comparisons" in two ways. Firstly, it is used to determine the appropriateness of the hypothetical or ideal form of the program used in S-Grading; and secondly it is an actual comparison of how the program being evaluated compares with other, similar programs. That is, how it would be RANKED.

The second explanation of "Critical Comparisons" is a procedural technique, and, though it results in a judgement of the relative

value of a program, it does not contribute directly to the worth of the program. Therefore it should not be included in the Checklist. For example, a poor program, when compared to other similar programs, may be the best of the lot; but that judgement does not make it better than it was before it was compared to the other programs.

The first explanation of "Critical Comparison", that it is an indication of the appropriateness of the ideal form of the program used in S-Grading, is an evaluation of the evaluation. The ideal form of the program chosen for S-Grading will determine the criteria used in establishing the standards. If the criteria were so easy to attain that even a poor program could get high marks, then that fault reflects on the evaluation not on the program being evaluated. The realism of the criteria is reflected in the defensibility of the ideal to both the client and the educational community.

The parallel in statistics, to the preceeding paragraph, is a discussion of TYPE I and TYPE II errors. Suppose we worded the outcomes of an evaluation in the following statistical manner:

H_0 : The program is a poor program.

H_1 : The program is a good program.

If our criteria are so easy to attain that we reject H_0 in favour of H_1 when we should have accepted H_0 , then a kind of TYPE I error has been committed. If, on the other hand, we are so severe in our choice of criteria for the evaluation that we accept H_0 when we should have rejected it, then a kind of TYPE II error has been committed.

Therefore, as in statistics, the choice of criteria for an evaluation is an attempt to minimize both TYPE I and TYPE II errors. As such, the checkpoint of Critical Comparisons can be used as a method of scrutinizing the choice of criteria, but does not directly contribute to the worth of the program being evaluated. For this reason and the reason previously given, "Critical Comparisons" was omitted from the checklist outlined in Sec. 3.33.

The effect of the evaluator entering a program is an important concern which was not fully investigated in this thesis. However, it was apparent that some problems, which would have remained as internal grumblings and been accepted by the program participants without comment, rapidly gained more stature than they deserved, because there was an evaluator present who recorded them. The appearance of the evaluator tended to make many uncommitted participants of the program more critical, and many committed participants more defensive.

Although the methodology of evaluation developed by this thesis was only applied to a teacher training program, it would appear that it could easily be extended to the evaluation of other educational programs. The two standards used, goal worth and goal achievement, are standards which apply to all educational programs. The general explication of evaluation in Chapter Two can be applied to the evaluation of any product.

The form of the evaluation illustrated by this thesis was S-GRADING. However, as shown in Sec. 3.70, the extension to more comprehensive

evaluations in RANKING is logically apparent.

Finally, the methodology developed is seen as providing a link between the theoretical views of evaluation and their application to a particular evaluation study.

REFERENCES

- Aiken, W.M. 1942. The Story of the Eight Year Study. Harper & Brothers, N.Y., N.Y.
- Aiken, M., Klein, S., Fernstermacher, G. 1968. The Centers Changing Evaluation Model. Evaluation Comment, October, 1968.
- Baker, Eva. 1974. "Formative Evaluation of Instruction", Evaluation in Education. ed. W.J. Popham. McCutchan Publishing Corp. Berkeley, California. 1974. pp. 535-584.
- Bloom, B.S. 1969. "Some Theoretical Issues Relating to Educational Evaluation", Educational Evaluation: New Roles, New Means. ed. R. Tyler. The 68th Annual Yearbook of the NSSE. University of Chicago Press, Chicago, Ill. 1969. pp. 26-50.
- Broudy, Harry S. 1965. "Criteria for the Professional Preparation of Teachers." Journal of Teacher Education 16(4), 1965. pp. 408-415.
- Daniels, LeRoi B. 1971. The Justification of Curricula. A paper prepared for discussion at the AERA 1971 Annual meeting in New York City (session B 14).
- Evans, J.W. 1974. "Evaluating Educational Programs - Are We Getting Anywhere?" Educational Researcher, Vol. 3, No. 8, Sept., 1974. pp. 7-12.
- Frankena, W.K. 1973. Ethics. Prentice-Hall, Inc. Englewood Cliffs, New Jersey. 2nd Edition.
- Glass, G.V. 1969. The Growth of Evaluation Methodology. Laboratory of Educational Research, University of Colorado.
- Glass, G.V. 1975. "A Paradox about Excellence of Schools and the People in Them." Educational Researcher. March 1975, pp. 9-13.
- Gleadow, Norman E. 1975. "The Moberly Tots Summer School Program 1975" Research Report (in Print). Evaluation and Research, Education Services Group, Vancouver School Board, 1595 West 10th Ave. Vancouver, B.C. Canada
- Guba, Egon G. 1969. "The Failure of Educational Evaluation" Educational Technology, May 1969. pp. 29-38.

- Hall, 1966. "The Applied Sociologist and Organizational Sociology" in Sociology in Action. Ed. A.B. Shostak. Dorsey Press.
- Haller, E.J. 1974. "Cost Analysis for Educational Program Evaluation," Evaluation in Education. Ed. W.J. Popham. McCutchan Publishing Corporation, Berkeley, California. pp. 406-449.
- Hayes, Samuel P. (Jr.) 1969. Evaluating Development Projects. (5th Edition) United Nations Educational, Scientific and Cultural Organization. Place de Fontenay, Paris.
- Hemphill, John K. 1969. "The Relationship between Research and Evaluation Studies", Educational Evaluation, New Roles, New Means. Ed. R. Tyler. The 68th Annual Yearbook of the NSSE. University of Chicago Press, Chicago, Ill. pp. 189-220.
- House, Ernest R. 1973. School Evaluation: The Politics and Process. McCutchen Publishing Corporation. California, U.S.A.
- Kerlinger, Fred N. 1973. Foundations of Behavioral Research. 2nd Ed. Holt, Rinehart and Winston, Inc. Toronto, Ont.
- Komisar, B. Paul 1961. Needs and the Needs-Curriculum. Language and Concepts in Education. Smith, B.O. and Ennis, R.H. (Ed's) pp. 24-43. Rand McNally & Co., Chicago.
- Likert, R. 1932. "The method of Constructing an Attitude Scale" Archives of Psychology, 1402. pp. 44-53.
- Metfessel, N.S., Michael, W.B. 1967. "A Paradigm Involving Multiple Criterion Measures for Evaluating the Effectiveness of School Programs" Educational and Psychological Measurement. 27. p. 931-943.
- Peters, R.S. 1966. Ethics and Education. George Allen & Unwin Ltd. London.
- Popham, W.J. 1969. "Educational Needs Assessment." In Curriculum Theory Network Monograph Supplement Curriculum Evaluation: Potentiality and Reality Ed. Joel Weiss. The Ontario Institute for Studies in Education. 1972. p. 22-32.
- Popham, W.J. 1972. An Evaluation Guidebook. The Instructional Objectives Exchange, Los Angeles, California.
- Popham, W.J. 1974. Editor. Evaluation in Education: Current Applications. McCutchan Publishing Corp. Berkeley, California.

- Provus, M. 1969. "Evaluation of Ongoing Programs in the Public School System". Educational Evaluation: New Roles, New Means. Ed. Paul Tyler Part II: The 68th Yearbook of the NSSE. University of Chicago Press, Chicago, Ill, pp. 242-283.
- Provus, M. 1970. "Evaluation or Research, Research or Evaluation?" Educational Technology, Aug. 1970. pp. 50-54.
- Provus, M. 1972. "Evaluation as Public Policy", Curriculum Theory Network. Monograph Supplement Curriculum Evaluation: Potentiality and Reality. Ed. Joel Weiss. The Ontario Institute for Studies in Education. 1972. Pp. 33-44.
- Rossi, P. 1966. "Boobytraps and Pitfalls in the Evaluation of Social Action Programs", Proceedings: Annual Meeting of the American Statistical Association, Washington, D.C. 1966.
- Scriven, M. 1966a. Value Claims in the Social Science. Boulder, Colorado, Social Science Education Consortium Publication No. 123.
- Scriven, M. 1966b. "Causes, Connections and Conditions in History," Philosophy of Educational Research. Ed. H.S. Broudy, R.H. Ennis, L.I. Krinerman. John Wiley & Sons, Inc. Toronto 1973. pp. 439-458.
- Scriven, M. 1967. "The Methodology of Evaluation". Perspectives of Curriculum Evaluation. AERA Monograph series on Curriculum Development No. 1, 1967. Chicago Ill. Radn McNally. pp. 39-83.
- Scriven, M. 1972. "Evaluation: Noble Profession and Pedestrian Practise." Curriculum Evaluation Potential. A Curriculum Theory Network Monograph. Toronto: Ontario Institute for Studies in Education. pp. 132-139.
- Scriven, M. 1973. "The Evaluation of Teaching at Berkeley" Draft paper (not a final version). 10/10/73.
- Scriven, M. 1974a. "The Evaluation of Educational Goals, Instructional Procedures and Outcomes, or, the Human Cometh", Program Development in Education, Ed. Blaney, Horsego and McIntosh. (Center for Continuing Education, University of B.C. 1974. Vancouver, B.C. Canada). pp. 134-162.
- Scriven, M. 1974b. "Evaluation Perspectives and Procedures", Evaluation in Education: Current Applications. Ed. W.J. Popham. McCutchan Publishing Corporation, Berkeley, California. pp. 1-94.

- Sjorgen, Douglas D. 1970. "Measurement Techniques in Evaluation" Review of Educational Research. 40. No. 2. April 1970. pp. 301-320.
- Stake, R.E. 1967. "The Contenance of Educational Evaluation" Teachers College Record. 68. pp. 523-540.
- Stake, R. 1971. An Evaluation of TCITY, the Twin City Institute for Talented Youth.
- Stake, R. 1972a. An Approach to the Evaluation of Instructional Programs. Paper delivered at AERA Annual Meeting April 4, 1972. in Chicago (ERIC ED064350).
- Stake, R.E. 1972b. "Responsive Evaluation". University of Illinois, unpublished manuscript.
- Stake, R.E. 1972c. "An Approach to the Evaluation of Instructional Programs" A paper delivered at the AERA Annual meeting, April 4, 1972. in Chicago. (ERIC access no: ED 064350)
- Stufflebeam, D.L. 1968. "Toward a Science of Educational Evaluation" Educational Technology, July 30, 1968. pp. 5-12.
- Taylor, P.W. 1961. Normative Discourse. Prentice-Hall, Inc. Englewood Cliffs, N.J.
- Taylor, P.A. and Maguire, T.O. 1966. "A Theoretical Evaluation Model" Manitoba Journal of Education Research. 1966, 1, pp. 12-17.
- Tyler, R. 1942. Basic Principles of Curriculum and Instruction. University of Chicago Press, Chicago, 1942.
- Tyler, R.W. 1969. Editor. Educational Evaluation: New Roles, New Means. Part II of the 68th Annual Yearbook of the NSSE. University of Chicago Press. Chicago, Ill.
- U.S. Dept. of Health, Education and Welfare. 1959. Accreditation in Higher Education. U.S. Government Printing Office, Washington, D.C.
- Welch, Wayne W. 1969. "Curriculum Evaluation". Review of Educational Research. Vol. 39, No. 4, Oct. 1969. pp. 429-443.
- Welch, Wayne W. 1972. "Evaluation of the PSNS course: I: Design and Implementation" Journal of Research in Science Teaching 9 pp. 139-145.

Worthen, B. 1968. "Toward a Taxonomy of Evaluation Designs". Educational Technology, Aug. 15, p. 3-9.

Worthen and Sanders. 1973. Editors. Educational Evaluation: Theory and Practice. Charles A. Jones Publishing Co. Worthington, Ohio.

Appendix A

Questionnaire for the Program Developers

Subject: Evaluation Priorities

Program B EVALUATION PRIORITIES

The persons involved in Program B should help determine the direction and emphasis of an internal evaluation. Therefore, you are asked to indicate on the scale, following each of the topics in the checklist, the importance and emphasis the topic should receive. This is not a test or attitude survey. The results will be used to help determine initially, which topics should be investigated, the depth of the investigation and the sequence of investigation.

1) MARKET

Is there an implementation plan for ensuring a market for the project?

Possible considerations should be:

1. Portability of the program.
2. Clarity, feasibility, ingenuity and economy of the implementation plan.
3. Size of the market for Program B.
4. Importance of the market for Program B.
5. Communications between Program B and potential consumers.

Importance of evaluating Market (check one)

- ☐ 4 Extremely important: a comprehensive check should be made.
☐ 3 Important: check as thoroughly as you can.
☐ 2 Definitely do some checking.
☐ 1 Check only if there is time.
☐ 0 Don't waste time on it.

Any comments?

2) TRUE FIELD TRIALS (performance)

How typical is the performance of the Program B program in school of the final, portable version.

For example:

1. Are the student-teachers in Program B typical of their group?
2. Are the U.B.C. staff members a representative sample?

3. Is the School representative?
4. Are any major changes seen for the final version?

Importance of evaluating True Field Trials: (check one)

- ☐ 4 Extremely important: a comprehensive check should be made.
☐ 3 Important: check as thoroughly as you can.
☐ 2 Definitely so some checking.
☐ 1 Check only if there is time.
☐ 0 Don't waste time on it.

Any comments?

3. TRUE CONSUMER (performance)

How do significant consumer groups gauge the performance of Program B.

Examples of significant consumers would be:

1. The student-teachers.
2. The U.B.C. staff.
3. The ~~Limestone~~ school teachers and principal.
4. The students in the school.
5. Vancouver School Board.

Importance of evaluating True Consumers: (check one)

- ☐ 4 Extremely important: a comprehensive check should be made.
☐ 3 Important: check as thoroughly as you can.
☐ 2 Definitely do some checking.
☐ 1 Check only if there is time.
☐ 0 Don't waste time on it.

Any comments?

4. CRUCIAL COMPARISONS (performance)

How does Program B perform as compared to other teacher training programs at U.B.C.? Is Program B achieving the same expectations as other teacher-training programs at U.B.C.?

A critical competitor might be the regular 3rd year teachers' training program.

Importance of evaluating crucial comparisons: (check one)

- ☐ 4 Extremely important: a comprehensive check should be made.
☒ 3 Important: check as thoroughly as you can.
☐ 2 Definitely do some checking.
☐ 1 Check only if there is time.
☐ 0 Don't waste time on it.

Any comments?

5. SIDE EFFECTS (performance)

Are there any unanticipated side effects to the program?

This checkpoint would look for (for example):

1. Any unanticipated benefits or problems to the student-teachers.
2. Any unanticipated benefits or problems to the students.
3. Any unanticipated benefits or problems to the school and its staff.
4. Any unanticipated benefits or problems to the University and the U.B.C. staff.

Importance of evaluating Side Effects: (check one)

- ☐ 4 Extremely important: a comprehensive check should be made.
☒ 3 Important: check as thoroughly as you can.
☐ 2 Definitely do some checking.
☐ 1 Check only if there is time.
☐ 0 Don't waste time on it.

Should this be done: during ____; at the end of ____; after ____; Program B?

Any comments?

6. PROCESS (performance)

Is the Program B program achieving what it set out to do? Is there any injustice; any unhappiness; any cruelty or their converses which are part of the program?

Examples of this checkpoint are:

1. Are the students at the school being taught effectively by the student-teachers?
2. Are the student-teachers being taught effectively by the U.B.C. staff?
3. Is there any cruelty, unhappiness or injustice in the program?
4. Is there the converse of "3" in the program?
5. Are there any moral implications to consider?
6. Are observation and description congruent?

Importance of evaluating Process: (check one)

- ☐ 4 Extremely important: a comprehensive check should be made.
- ☐ 3 Important: check as thoroughly as you can.
- ☐ 2 Definitely do some checking.
- ☐ 1 Check only if there is time.
- ☐ 0 Don't waste time on it.

Any comments?

7. CAUSATION (performance)

Are the observed and reported outcomes attributable to Program B?

For example:

1. Have the classroom techniques used by the student-teachers been learned through Program B?
2. Are the behaviors identified under "Process" the result of the

Program B program?

Importance of evaluating Causation: (check one)

- ☐ 4 Extremely important: a comprehensive check should be made.
- ☐ 3 Important: check as thoroughly as you can.
- ☐ 2 Definitely so some checking.
- ☐ 1 Check only if there is time.
- ☐ 0 Don't waste time on it.

Any comments?

8. (a) The topic you are personally most interested in is number ____.

(b) The topic you consider most necessary to the evaluation of Program B is number ____.

Any comments?

Appendix B

Interview of School Administration

Questions asked:

- 1) What needs do you see Program B fulfilling?
- 2) Is Program B doing the things you expected it to do?
If Yes? What are those things?
If NO? What is being omitted?
- 3) Are you clear about your role in Program B?
- 4) Are there any unanticipated positive effects of the program so far? If yes, what are they?
- 5) Are there any unanticipated negative effects of the program so far? If yes, what are they?
- 6) What are your general impressions about the U.B.C. staff and their functioning in the program?
- 7) What are your general impressions about the student teachers?
- 8) Any other comments?

Appendix C

Likert Attitude Scale; to measure general attitudes
toward the teacher training process of the student teachers.

This is not a test. Consider each item carefully before answering.
Your answers will remain anonymous.

Record all your answers on the computer answer card (depicted below).

| IDENTIFICATION NUMBER | | | | | | | | | | ANSWER FIELD | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
| <div style="transform: rotate(-90deg); position: absolute; left: 50px; top: 50px; font-weight: bold;">LEAVE BLANK</div> | | | | | | | | | | <div style="position: absolute; top: 50px; left: 50px; font-weight: bold;">ANSWER FIELD</div> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <div style="transform: rotate(-90deg); position: absolute; left: 50px; top: 50px; font-weight: bold;">LEAVE BLANK</div> | | | | | | | | | | <div style="position: absolute; top: 50px; left: 50px; font-weight: bold;">ANSWER FIELD</div> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

UNIVERSITY OF BRITISH COLUMBIA

OMR MULTIPLE-CHOICE EXAMINATION ANSWER CARD

Do not put any marks in the section called "identification number", or in the section called "answer card number".

Use the answer field to record your answers. There are 50 columns for answers, but you will only use the first 18 columns. For each item in the following form there is a corresponding 5-choice column. Record your response to each item by blackening the appropriate bubble in the column corresponding to the item. See the following example:

EXAMPLE:

1. EDUCATION IS FUN.

1 - Strongly disagree

2 - Disagree

3 - Undecided

Answer → 4 - Agree

5 - Strongly agree

BLACKEN bubble 4 in column 1 as shown below:

USE PENCIL ONLY!

| | |
|------------|---|
| Item 1 | 1 |
| | 0 |
| | 1 |
| | 2 |
| Answer (4) | 3 |
| | 4 |

RESPONSE CODE

| | | |
|---|----|-------------------------------------|
| 1 | SD | <u>Strongly disagree</u> with item. |
| 2 | D | <u>Disagree</u> with item. |
| 3 | U | <u>Undecided.</u> |
| 4 | A | <u>Agree</u> with item. |
| 5 | SA | <u>Strongly agree</u> with item. |

PLEASE USE PENCIL ONLY:

Note: TEACHER TRAINING refers to your whole program of courses, classes, teaching practice, etc. i.e. it includes all aspects of your program. STUDENT TEACHING refers to your classroom teaching practice only.

THE STATEMENTS: PLEASE RESPOND TO THESE ON YOUR COMPUTER ANSWER CARD.

1. I HOPE TO KEEP IN TOUCH WITH MY PRESENT CLASSMATES IN THE FUTURE.
2. U.B.C. HAS MADE A SIGNIFICANT CONTRIBUTION TO TEACHING IN B.C.
3. STUDENT TEACHING ISN'T WORTH THE EFFORT.
4. UNIVERSITY IS AN ENJOYABLE EXPERIENCE.
5. I LOOK FORWARD TO MY STUDENT TEACHING EXPERIENCE.
6. GENERALLY SPEAKING, MY EDUCATION COURSES ARE USELESS.
7. MY U.B.C. INSTRUCTORS ARE FINE PEOPLE.
8. TEACHING IS AN ENJOYABLE EXPERIENCE.
9. MY CLASSMATES ARE A VALUABLE SOURCE OF IDEAS FOR TEACHING.
10. MY SPONSOR TEACHERS HAVE HELPED ME DEVELOP MANY USEFUL TEACHING SKILLS.
11. TEACHING SCHOOL IS A VERY IMPORTANT JOB.
12. TEACHER TRAINING IS BORING.
13. AS A WHOLE, MY EDUCATION COURSES ARE INTERESTING.
14. MY SPONSOR TEACHERS ARE GENERALLY ENTHUSIASTIC ABOUT TEACHING.
15. TEACHER TRAINING IS IMPORTANT.
16. MY EDUCATION COURSES ARE NOT HELPING ME TO BECOME A GOOD TEACHER.
17. TEACHER TRAINING IS AN ENJOYABLE EXPERIENCE.
18. MY SPONSOR TEACHERS HAVE A STRONG INFLUENCE ON MY TEACHING.

Appendix D
Reaction Form

REACTION FORM

Date _____

1. Circle the number below which best represents your feelings about Program B at this time.

| | | | | | | |
|------------------|---|---|---------|---|---|------------------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| very negative | | | neutral | | | very positive |

2. Briefly describe any incidents that you saw as positive, and any that you saw as negative relating to your current reaction to Program B.

(+)

(-)

3. Anything else you'd like to mention?

Appendix E

Evaluation of UBC instructors' teaching performance.

Form A.

Form ARESPONSE CODE

- | | | |
|----|----|-------------------------------------|
| 1 | SD | <u>Strongly disagree</u> with item. |
| 2. | D | <u>Disagree</u> with item. |
| 3. | U | <u>Undecided.</u> |
| 4. | A | <u>Agree</u> with item. |
| 5. | SA | <u>Strongly agree</u> with item. |

1. COURSE _____ INSTRUCTOR _____ IDENTIFICATION No. _____

2. COURSE _____ INSTRUCTOR _____ IDENTIFICATION No. _____

- 1) This instructor gives unsatisfactory answers to students' questions.
- 2) This instructor often misses the point of a question asked by a student.
- 3) This instructor seldom varies his method of instruction.
- 4) This instructor's presentations in class are poorly organized.
- 5) This instructor is very careful not to offend his students with his remarks.
- 6) This instructor is impatient with his students.
- 7) The level of difficulty of this course is about right.
- 8) The instructor seldom checks to see if students understand what he is teaching.
- 9) This instructor clearly explains difficult ideas.
- 10) This instructor uses humour effectively in his teaching.
- 11) This instructor seldom engages his students in class discussions.
- 12) The reading materials in this course (text, other sources) are of little use.
- 13) I developed my ability to marshal or identify main points or central issues in teaching from this course.
- 14) This course is of definite value for a person who wishes to become a teacher.

If you have any comments then please write them on the back of this page with the name of the instructor your comments pertain to.

OVERALL EVALUATION OF INSTRUCTORS

1. professor A 2. professor B 3. professor C 4. professor D
5. professor E 6. professor F.

| IDENTIFICATION NUMBER | | | | | | | | | | ANSWER FIELD | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------------|---|---|---|---|---|---|---|---|----|--------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|--|--|--|--|--|--|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

PLEASE RATE EACH INSTRUCTOR IN THE APPROPRIATE COLUMN (ILLUSTRATED ABOVE).

How good a job is the instructor doing in getting across to you the knowledge, understanding, skills and attitudes that appear to be the goals of the course? (In answering, try not to be unduly influenced by your liking or dislike for the instructor as a person; or by your feelings about the course content, if that wasn't something which the instructor could choose freely.)

Please use the following scale for your responses on the computer card:

- 1) Outstandingly bad
- 2) Below average
- 3) Average
- 4) Above average
- 5) Outstandingly good

Appendix F

Questionnaire to student-teachers on their alternate placement.

PROGRAM B EVALUATION

This questionnaire deals with your most recent experience in alternate placements.

1. During your alternate placement were you given the placement you requested?

YES _____ NO _____

2. Which of the following was your placement?

OPEN AREA _____

FREE SCHOOL _____

FAMILY GROUPING _____

OTHER (PLEASE SPECIFY) _____

3. Was the placement more demanding or less demanding of your time and energy than you had anticipated it would be?

MORE DEMANDING _____

LESS DEMANDING _____

WHAT I EXPECTED _____

UNDECIDED _____

COMMENTS:

4. Did you find your experience beneficial? YES _____ No _____

UNDECIDED _____

Please comment:

5. Were your sponsor teachers expecting you? YES _____ NO _____

UNDECIDED _____

Comments:

6. Did you have a clear idea of your teaching duties before your alternate experience began? YES _____ NO _____

Comments:

7. Did your sponsor teacher have a clear idea of your teaching duties when you arrived: YES _____ NO _____

Comments:

8. Generally speaking, did you actually do what you had expected to do in your alternate experience? YES _____ NO _____
(If "NO" please comment)

Comments:

9. Did you feel you were supervised often enough by:

a. YOUR SPONSOR TEACHER YES _____ NO _____ UNDECIDED _____

b. YOUR U.B.C. ADVISOR YES _____ NO _____ UNDECIDED _____

Comments:

10. Do you think this alternate experience should be continued in future Program B's?

YES, WITH CHANGES _____

YES, WITHOUT CHANGES _____

NO, IT SHOULD BE DISCONTINUED _____

10. (Continued).

If you checked "YES, WITH CHANGES", what changes would you suggest?

If you checked "NO, IT SHOULD BE DISCONTINUED", would you please indicate why you feel that way?

11. Considering everything, what is your overall evaluation of the alternate placement?

_____ EXTREMELY VALUABLE

_____ VALUABLE

_____ HAD SOME VALUE

_____ HAD LITTLE VALUE

_____ A COMPLETE WASTE OF TIME

12. Any other comments?

Appendix G

Evaluation of UBC Instructors'
teaching performance
Form B

FORM B

- 1 SD Strongly disagree with item.
2 D Disagree with item.
3 U Undecided.
4 A Agree with item.
5 SA Strongly agree with item.

COURSE: _____ INSTRUCTOR: _____ IDENTIFICATION NO. ____
COURSE: _____ INSTRUCTOR: _____ IDENTIFICATION NO. ____

- 1) The instructor appears interested in teaching this course.
- 2) Overall, this is a good course.
- 3) The plan for this course is clearly evident.
- 4) The instructor allows a few students to monopolize class discussions.
- 5) The instructor is seldom available for assistance outside of class time.
- 6) The instructor often stimulates students to undertake additional work in the course on their own.
- 7) The instructor seldom asks questions which stimulate discussion.
- 8) The instructor seems competent in the subject matter of this course.
- 9) The instructor frequently points out how different parts of the course are related.
- 10) I can recognize good and bad teaching techniques because of this course.
- 11) This course increased my awareness of issues related to the subject matter.
- 12) This course is of definite value for a person who wishes to become a teacher.

If you have any comments then please write them on the back of this page with the name of the instructor your comments pertain to.

Overall Evaluation of Instructors (see Appendix E)

| | | | | | |
|-------------|-------------|-------------|-------------|-------------|------|
| 1 Professor | 2 Professor | 3 Professor | 4 Professor | 5 Professor | etc. |
| A | B | C | D | E | |

PLEASE RATE EACH INSTRUCTOR IN THE APPROPRIATE COLUMN (ILLUSTRATED ABOVE)

How good a job is the instructor doing in getting across to you the knowledge, understanding, skills and attitudes that appear to be the goals of the course? (In answering, try not to be unduly influenced by your liking or dislike for the instructor as a person; or by your feelings about the course content, if that wasn't something which the instructor could choose freely.)

Please use the following scale for your responses on the computer card:

1. Outstandingly bad.
2. Below average.
3. Average.
4. Above average.
5. Outstandingly good.

Appendix H

Interview of student teachers

Interview Questions

- 1) Did you have any serious problems:
 - a) In finding a place to live?
 - b) With transportation to the school?
 - c) With the pupils you taught?
 - d) With the UBC professors?
 - e) With the teachers
 - f) With your fellow student-teachers?
 - g) Anything else?
- 2) Was Program B more costly to you than a 3rd year regular program at the University would have been?
- 3) Would it matter to you where the school was located in terms of cost? Transportation? Type of students?
- 4) How did you come to know about Program B?
- 5) Why did you apply to Program B?
- 6) Why do you want to be a teacher?
- 7) Did Program B meet your expectations?
- 8) Were the following physical facilities at the school adequate:
 - a) Washrooms?
 - b) Lunchrooms, staffrooms?
 - c) Areas to relax, smoke, talk?
 - d) Work and preparation areas?
- 9) Did you have sufficient access to books and supplies for:
 - a) Your university course work?
 - b) Your teaching?
- 10) Did you feel that the work load (courses, classroom teaching, supervision, etc.) was excessive?
- 11) Do you think that having methods courses at the school is an advantage over having them at the university? Why?

- 12) Overall, were the method courses useful?
 - How were they useful or why were they not useful?
- 13) What do the goals of Program B seem to be?
 - Do you think it has succeeded in reaching these goals?
 - On what evidence are you basing your opinion?
- 14) Would you recommend Program B as it is to other potential student-teachers? Why or why not?
- 15) Would you take Program B again if you were back in September of 1974? Why or why not?
- 16) Is Program B especially suited for certain types of student teachers? What type?
- 17) Do you think the year following this will be:
 - valuable?
 - interesting?
- 18) Did any individuals have a large negative or positive effect on the direction the program took?
- 19) Was this school ideal for Program B? Why or why not?
- 20) Could there have been more student-teacher input into the program? If so, when or where? If not, are you satisfied with the amount of input you had?

Appendix I

"Final" Questionnaire for Student Teachers

Program B Evaluation

1. Did Program B link the methods courses to actual teaching?

frequently _____ never

2. Did the program relate the study of developmental psychology to immediate experiences?

frequently _____ never

3. Did the foundation course focus on school based issues and problems?

frequently _____ never

4. Did the psychology component increase your sensitivity to the needs and requirements of your students?

Yes _____ No _____ Undecided _____

5. Did the foundations component enable you to make clear, well-founded decisions as to what goals a teacher ought to pursue?

Yes _____ No _____ Undecided _____

6. Did the foundations component enable you to promote the moral development of your students?

Yes _____ No _____ Undecided _____

7. Did the foundations component give you a clear understanding of the nature of the task of developing the ability to think?

Yes _____ No _____ Undecided _____

8. Did you have sufficient opportunity to try out your own ideas in class?

Yes _____ No _____ Undecided _____

9. Do you feel the U.B.C. professors are enthusiastic about Program B's approach to teacher-training?

Yes _____ No _____ Undecided _____

10. Do you feel the teachers at (school name) are enthusiastic about Program B's approach to teacher-training?

Yes _____ No _____ Undecided _____

11. Do you think program B has been flexible?

Yes _____ No _____ Undecided _____

12. Do you think Program B has been varied?

Yes _____ No _____ Undecided _____

13. Were you able to make as much use of the U.B.C. library as you wanted to?

Yes _____ No _____ Undecided _____

14. Considering everything, what is your overall evaluation of Program B?

EXCELLENT _____ GOOD _____ O.K. _____ POOR _____ A WASTE OF TIME _____

Appendix J

Discipline techniques

DISCIPLINE TECHNIQUES

1. Simple control
 - a. A look or a frown.
 - b. Shake of head or other mild physical gesture.
 - c. Mild verbal reproof.
 - d. Movement to trouble area.
 - e. Shouting at student (s).
2. Individual conference with pupil.
3. Paying for, rebuilding or replacing damaged item.
4. Loss of some privilege.
5. Detention after school.
6. Dismissal from class.
 - a. Sent to office.
 - b. Sent to hallway.
 - c. Other area.
7. Isolation in the classroom, i.e. sent to a corner, etc.
8. Punishing a whole group or class.
9. Extra work or tasks.
10. Enforced apologies.
11. Physical contact -- shaking, striking, forcefully directing.
12. Threat of any of the above.
13. Other.

Appendix K

Side Effects of Program B (Teachers' Questionnaire)

Program B's Evaluation

Dear Teacher,

During the past few months of Program B an internal evaluation, as you know, has been carried out. Knowing how loaded down with paper-work and forms you are, I have not bothered you with any questionnaires or opionnaires -- until now. This is the first and the last questionnaire that I will give to you. Its purpose is to attempt to establish if there have been any unanticipated benefits, and/or problems which have resulted from the operation of Program B in your school. These unanticipated aspects of the program (or side-effects) are extremely important.

Will you please put your responses to this questionnaire in the envelope provided for them in the office. Please do not put your name on your response.

Thank you,

Norman Gleadow

SIDE EFFECTS OF PROGRAM B

This form is to determine if there are any unanticipated effects of Program B. These effects are the benefits and/or the problems which were not foreseen when the program was initiated. Please do not put your name on this form so as to maintain your anonymity.

1. Are you aware of any UNANTICIPATED BENEFITS or any UNANTICIPATED PROBLEMS of Program B to the STUDENT-TEACHERS?

UNANTICIPATED BENEFITS

UNANTICIPATED PROBLEMS

2. Are you aware of any UNANTICIPATED BENEFITS or UNANTICIPATED PROBLEMS of Program B to the SCHOOL'S STUDENTS?

UNANTICIPATED BENEFITS

UNANTICIPATED PROBLEMS

If you have any additional comments please put them on the back.

3. Are you aware of any UNANTICIPATED BENEFITS or any UNANTICIPATED PROBLEMS of Program B to the SCHOOL'S TEACHERS?

UNANTICIPATED BENEFITS

UNANTICIPATED PROBLEMS

4. Are you aware of any UNANTICIPATED BENEFITS or any UNANTICIPATED PROBLEMS of Program B to the U.B.C. PROFESSORS?

UNANTICIPATED BENEFITS

UNANTICIPATED PROBLEMS

5. Are you aware of anything you would consider unjust, immoral, or causing serious unhappiness as a result of Program B?

Appendix L

Interview with Program B's Director

Questions Asked.

- 1) How did the idea of Program B originate? When? What was the impetus for Program B.
- 2) What did Program B hope to do that wasn't being done by the third year regular course at U.B.C.?
- 3) Why were the third year transfer students chosen?
- 4) How were perspective student-teachers notified of Program B?
- 5) How were the student-teachers finally selected for Program B? How many were turned down? Why?
- 6) How did the UBC staff for Program B get chosen?
- 7) How was the host school chosen?
- 8) What connection has the school board with the program?
- 9) Has there been any unsolicited, written feed-back on the student-teachers' performance?
- 10) Do you think the average third year student could operate well in Program B?
- 11) What do you think constitutes the ideal student-teacher for Program B?
- 12) Do you know of any research results on programs similar to Program B?
- 13) Do you think it would be feasible to have all teacher-training done via Program B? Why or why not?
- 14) What is the most expensive part of Program B as compared to the regular third year Program?
- 15) What are some other expenses which are associated with Program B?
- 16) What do you think are the greatest strengths and weaknesses of Program B?

Appendix M

Questionnaire for School Pupils.

Your Grade _____

QUESTIONNAIRE FOR STUDENTS

We would like you to answer some questions about the student-teachers who have taught you over the past school year. Please do NOT put your name on this paper. If you wish to make any written comments, please do so in the spaces under each question.

1. Did the student-teachers teach too fast? YES _____ NO _____
2. Did the student-teachers teach too slow? YES _____ NO _____
3. Did they give you fair tests and exams? YES _____ NO _____
4. Did you understand what they were trying to teach you in class?
YES _____ NO _____
5. Did any of them ever help you with school work outside of class?
YES _____ NO _____
6. Could the student-teachers control the kids who didn't behave?
YES _____ NO _____
7. Were they ever unfair to you? YES _____ NO _____
If you answered yes to this question, would you please tell us
how the student-teachers were unfair?
8. Did you find them interesting to listen to?
YES _____ NO _____
9. Did you find it easy to learn things from them?
YES _____ NO _____
10. Were the student-teachers' lessons interesting?
YES _____ NO _____

11. Did they give you good answers to your questions?

YES _____ NO _____

12. Did you find it easy to get the student-teachers' attention in class?

YES _____ NO _____

13. Did they learn your name? YES _____ NO _____

14. Did you find them easy to talk to? YES _____ NO _____

15. Did you ever talk to any of the student-teachers in the classroom?

YES _____ NO _____

16. Did you ever talk to any of the student-teachers outside of the classroom?

YES _____ NO _____

17. Did you find switching back-and-forth from your regular teacher to the student-teachers confusing?

YES _____ NO _____

18. Place a check-mark beside those subjects which a student-teacher taught you

- _____ ART
- _____ ENGLISH OR LANGUAGE ARTS (reading, spelling, writing)
- _____ MATH or ARITHMETIC
- _____ SCIENCE
- _____ P.E.
- _____ SOCIAL STUDIES
- _____ MUSIC
- _____ OTHER (WHAT IS IT? _____)

19. Which of the above subjects do you think the student-teachers taught the best? _____.

20. Which of the above subjects do you think the student-teachers taught the worst? _____.

21. Did the student-teachers do a good job of teaching you?

YES _____ NO _____ NOT SURE _____

Appendix N

Interview with UBC Administrators.

Questions Asked.

- 1) What need do you see Program B fulfilling?
- 2) Have you had any specific feedback on Program B's participants?
If so, what?
- 3) What do you see as the strong and weak points of Program B?
- 4) What is your own judgment as to the overall performance of Program B?
- 5) Would you describe the U.B.C. professors involved in Program B as an average group, a poor group or an above average group?

Appendix 0

Student Teachers' Questionnaire (Side Effects)

SIDE EFFECTS - STUDENTS

1. Please indicate your overall evaluation of the registration procedures.

Excellent, couldn't be smoother

☐

Good, hardly any difficulties

☐

Acceptable, there were a few problems

☐

Not good, there were many problems

☐

Terrible, complete confusion

☐

Are there any good and/or bad points about the registration procedure you would like to comment on? If so, please use the space below.

2. (a) Please indicate your method of transportation to School from your place of residence.

Walking (or running)

☐

Bicycle

☐

Motorcycle

☐

Your own car

☐

In a car pool

☐

Bus

☐

Other (Please specify) _____

- (b) About how long does it take you to get to School from your place of residence?

Less than 15 minutes

☐

Between 15 and 30 minutes

☐

Between 30 and 45 minutes

☐

Between 45 and 60 minutes

☐

More than an hour

☐

- (c) Are you experiencing any problems with your present transportation arrangements from your place of residence to School?

Yes

☐

No

☐

If "yes" please indicate what those problems are.

3. Were you given an adequate orientation to the school?

No

☐

Yes

☐

Not Sure

☐

- (a) If "No" please indicate how it was inadequate.
 (b) If "Not Sure" please indicate, in general terms, why you feel uneasy about the orientation.

4. Do you have adequate work space in School?

Yes ☐

No ☐

Comments _____

5. Are the washroom facilities adequate?

Yes ☐

No ☐

Comments _____

6. Are arrangements for eating lunch, drinking coffee, smoking, etc. acceptable to you?

Yes ☐

No ☐

Comments _____

7. Are there any problems or irritations which are "in the bud" and should be dealt with now before they get worse?

Yes ☐

No ☐

If "Yes" please outline those problems.

8. Is there anything else you would like to comment on?
Please do in the space provided below.

Appendix P

U.B.C. Staff Questionnaire on Courses

Instructors' Questionnaire A

Course: _____ Instructor _____

Course will be given in Fall Term ☐ Spring Term ☐

Hours of instruction per week _____ hrs.

1. Recommended textbook (s) if any _____

2. Is (are) the textbook(s) available at the University Bookstore?

Yes ☐ No ☐ Comments: _____

3. Which part(s) of the textbook(s) will you be emphasizing?

All ☐; or _____

4. How will you be measuring the student's progress in your course (please check all that apply)?

Comments:

Final exam (written) ☐ _____
 Final exam (oral) ☐ _____
 Mid term exam ☐ _____
 Other exams (or tests) ☐ (Please give approximate frequency _____)
 Observation ☐ (Please indicate the type of observational setting you will be arranging).

Short written assignments (one page or less) ☐
 Longer written assignments (short essays) ☐
 Major projects ☐

Other (please specify) _____

Any other comments? _____

Appendix Q

Student-teachers' Questionnaire on Evaluation
of U.B.C. Instructors' Teaching Performance.

EVALUATION OF UBC STAFF TEACHING PERFORMANCE

In this particular questionnaire, you are being asked to rate the UBC staff, as a group, on a number of teaching criteria. If there are any individual staff members who you feel do not fit into a rating for the rest of the group then please indicate, under the heading of exceptions, who those individuals are and why they do not fit into your overall rating.

1. How frequently do UBC staff members come late to class without an adequate explanation?

☐
☐
☐
☐
☐

very
frequently

sometimes

never

exceptions? No ☐ or Yes ☐: please indicate

who the exceptions are and why they are exceptions.

2. How well is the staff sticking to what you were led to expect in their courses?

☐

very well

☐☐sometimes
deviate☐☐very
poorly

exceptions? No ☐ or Yes ☐ : please indicate

who the exceptions are and why they are exceptions.

3. How much opportunity do you get to ask questions in class?

☐ample
opportunity☐☐occasional
opportunity☐☐insufficient
opportunity

exceptions? No ☐ or Yes ☐ : please indicate

who the exceptions are and why they are exceptions.

4. How well organized are the UBC staff member's presentations in class.

☐

very well
organized

☐☐

O.K.

☐☐

very poorly
organized

exceptions? No ☐ or

Yes ☐ : please indicate

who the exceptions are, and why they are exceptions.

5. Do the UBC staff members know their subject material?

Yes ☐

No ☐

Don't Know ☐

exceptions? No ☐ or

Yes ☐ : please indicate

who the exceptions are, and why they are exceptions.

6. How useful are the assignments you have had so far?

☐
☐
☐
☐
☐

very useful

useless

exceptions? No ☐ or

Yes ☐ : please indicate

what the exceptions are, and why they are exceptions.

7. How worthwhile have you found the textbooks you bought and are presently using? (Please take into consideration the cost of the books, the assignments (readings, problems, etc.) you have had using the textbook, the reading level, the content, etc.)

☐
☐
☐
☐
☐

very
worthwhile

O.K.

worthless

exceptions? No ☐ or

Yes ☐ : please indicate

what the exceptions are, and why they are exceptions.

8. Are the UBC staff member readily available when you need them?

Yes ☐

No ☐

Don't Know ☐

exceptions? No ☐ or

Yes ☐ : please indicate

who the exceptions are, and why they are exceptions.

9. What is your overall evaluation of the job the UBC staff members are doing in getting across to you the knowledge, understanding, skills and attitudes that appear to be the goals of the courses? (In answering, try not to be unduly influenced by your liking or disliking for the instructors as people; or by your feelings about the courses' content if that isn't something the instructors could choose freely).

Outstandingly good ☐

Above average ☐

Average ☐

Below Average ☐

Outstandingly bad ☐

exceptions? No ☐ or

Yes ☐ : please indicate

who the exceptions are, and why they are exceptions.

10. If you have any other comments, please do so:

Appendix R
Teachers' Questionnaire

Dear Teacher,

This questionnaire asks a number of questions about how ^{Program 8} has been operating ⁱⁿ your school. The questionnaire looks more formidable than it actually is. The questions require just a simple checkmark; however, if you wish to add to your answers, please do so in the "comments" box under each question.

When you complete this questionnaire, please put it in the mail slot in your office. It would be appreciated if you could complete them within a week.

Thank you for your time.

Yours truly,

Norman Gleadow -
Evaluator

egr

TEACHER'S QUESTIONNAIRE

PLEASE INDICATE YOUR ANSWER BY PUTTING A CHECKMARK IN THE APPROPRIATE BOX. PLEASE DO NOT PUT YOUR NAME ON THIS FORM. IF YOU WISH TO MAKE ANY COMMENTS, PLEASE DO SO IN THE BOX PROVIDED.

1. The student-teachers have been taking a number of University courses at your school. Please place checkmarks beside those university courses for which you have received course OUTLINES:

| | | |
|---------------|-------------------------|------------------|
| Music _____ | Math _____ | Psychology _____ |
| Science _____ | Social Studies _____ | Other _____ |
| P.E. _____ | Language Arts _____ | (Please specify) |

Comments:

2. Do you find the material you receive about the courses being taught the student teachers of any value?

Not applicable ☐, Yes ☐, No ☐

Haven't had time to read it ☐

Comments:

3. Are you being kept aware of the student-teachers' course ASSIGNMENTS?

Yes, for all courses ☐

Yes, for most courses ☐

Yes, for about half the courses ☐

Yes, but for less than half the courses ☐

No, have not received any information of this type ☐

Comments:

4. In your opinion, have the U.B.C. professors adequately consulted with you with respect to:

- (a) Placement of student-teachers with sponsor teachers?

Yes ☐ No ☐ Not applicable ☐

Comments:

- (b) Teaching time for the student-teacher working with you?

Yes ☐ No ☐ Not applicable ☐

Comments:

- (c) Appropriate teaching strategies (methods, classroom management, material to cover, etc.) for your student teacher?

Yes ☐ No ☐ Not applicable ☐

Comments:

- (d) The evaluation of your student-teachers' teaching effectiveness.

Yes ☐ No ☐ Not applicable ☐

Comments:

5. Are you clear about your role in the program?
(Please check one of the following:)

No, I'm totally confused about what I'm expected to do. ☐

I think I know what I'm expected to do but there are still some areas of confusion. ☐

Yes, I have a very clear idea of what I am expected to do. ☐

Comments:

6. Generally, how would you rate the communication between the teachers and the U.B.C. staff.

☐ ☐ ☐ ☐ ☐
Excellent Good Fair Poor Terrible

Comments:

7. Are the student-teachers courteous.

☐ ☐ ☐ ☐ ☐
Always Sometimes Never

Comments:

8. Are the student-teachers appropriately dressed?

☐ ☐ ☐ ☐ ☐
Always Sometimes Never

Comments:

9. Do you feel the student-teachers are separating themselves from the operation of the school (i.e. staying in their room too much, not mixing sufficiently with the teachers, etc.)

Yes ☐

No ☐

Undecided ☐

Comments:

10. Do you approve of the student-teachers' behavior in all their contacts with the pupils in your school?

approve ☐

disapprove ☐

undecided ☐

If you disapprove of their behavior, would you please indicate your reasons.

11. Are you satisfied with the student-teachers' preparation for teaching?

Yes ☐

No ☐

undecided ☐

If you are not satisfied with their preparation, please indicate how they could prepare better.

12. Generally, how would you rate the overall rapport between the student-teachers and the teachers?

☐ excellent ☐ good ☐ fair ☐ poor ☐ non-existent

Comments:

13. Do you feel that your pupils are getting behind in their work because of the program?

Yes ☐

No ☐

Undecided ☐

If you answered "yes" would you please indicate the severity of the problem.

14. Have you found Program B too tiring?

Yes ☐

No ☐

Not Sure ☐

If yes, in what way has it been too tiring?

15. Do you have adequate time to do your own lesson planning, marking, etc.

Yes ☐

No ☐

Not Sure ☐

Comments:

16. Did you find the Communications Workshop useful?

Yes ☐

No ☐

Didn't attend ☐

Undecided ☐

Comments:

17. What is your overall evaluation of the program as it has operated in your school for the past month?

Excellent ☐

Good ☐

O.K. ☐

Poor ☐

Terrible ☐

18. Are there any minor or major problems you feel should be taken care of now before they get any worse?

Yes ☐

No ☐

Appendix S

Interview with Director of Elementary Education
at School Board Offices.

Questions Asked.

- 1) Have you received any information on beneficial effects of the program from: Parents? Students? Teachers? Principals? Administrators?
- 2) Have you received any information on harmful effects of the program from: Parents? Students? Teachers? Principals? Administrators?
- 3) Does the program cost the school District money? time? manpower? Is it worth the cost (if any)?
- 4) Has it increased your workload unacceptably?
- 5) Do you see any need for this approach to teacher-training? Why or why not?
- 6) What aspect of the program are you most interested in?
- 7) Do you have any plans for evaluating the program? If so, what? If not, why not?
- 8) What is your overall impression of Program B as it has operated to date?
- 9) Any other comments?

Appendix T

Evaluation of the Student Teachers Effectiveness
in Program B (a Proposal submitted to and accepted
by the teachers hosting Program B)

Proposal for the Evaluation of the Student Teachers
Effectiveness in the Program

Research has not yet established usable connections between observed teacher behaviors in the classroom and educational gains by the pupils. Classroom observational instruments are legitimate tools for obtaining an accurate account of what is taking place in the classroom but they do not necessarily identify variables which can be used to judge effective teachers. At best descriptions of a teacher's classroom technique are only secondary indications of teaching effectiveness. For example, one teacher may run a very "tight" classroom with well developed systems of management and regulation, while another teacher may have a very "loose" classroom. The two teachers would have different profiles on any observational instrument which described the teaching act, yet both might be described as being "good" teachers.

If we are to evaluate teaching effectiveness, then we must go beyond descriptions of the teaching act and search for primary indicators of "good" teaching. This is not meant to suggest that there are no secondary or descriptive indicators of good teaching. Some weight in the evaluation can be given to a small number of secondary indicators chosen on a common sensical or moral basis, and which can be expected to at least weakly correlate with good teaching.

Therefore, a major feature of the evaluation plan is extreme parsimony in the data-gathering, combined with extreme care in identifying the use to which the data will be put. Parsimony is advocated for the sake of validity. Only data which can be demonstrated to be connected with teaching merit should be collected and then only when there is a legitimate use for it.

The major uses of the evaluation of student-teaching include:

1. Evaluation for self-improvement.
2. Evaluation to assist the sponsor teacher and U.B.C. adviser with the process of improving the student teacher's teaching performance.
3. Evaluation to advise other possible consumers, such as pupils desiring teacher assistance in other school activities.
4. Evaluation for personnel decisions by school and U.B.C. staffs.
5. Evaluation by external audiences interested in, e.g., success of the Project.

Description of the Evaluation Instruments (These instruments are attached to these proposal)

I. Primary Indicators

(1) Overall Evaluation of Teaching.

This is a simple rating on a five point scale, ranging from outstandingly good to outstandingly bad of the student teacher's teaching ability. This single rating has two main advantages. Firstly, it cannot be easily misinterpreted, and secondly, the chances that a rating is not related to an indication of teaching ability are extremely remote.

Ideally this rating shall be given by the pupils being taught by the student teacher. However, in an elementary school, this presents special difficulties, particularly in the lower grades where the pupils may not fully understand what is expected of them. In this situation, we may have to rely on ratings given the student teacher by the sponsor teachers and the U.B.C. advisors. Since these people are not primary consumers of the teaching of the student teacher, then the overall evaluative rating must be based on as many classroom visits as possible - certainly not on the basis of a single visit.

(2) Pupil Gains

A teacher's main function is to meet the educational needs of his/her pupils. In other words, as a result of the teaching act, the pupils should have gained some combination of knowledge, skill, disposition or attitude. It is inconceivable to call someone a good teacher who's pupils do not "learn". Therefore, it is proposed that the student teachers use some method of determining whether the pupils have learned anything while exposed to the student teacher's teaching. To give this some experimental validity a simple pre-test, post-test procedure is suggested. That is, in consultation with the sponsor teacher, a test will be made up on a particular unit the student teacher will be teaching.

(Dr. Ed. 311 course will provide the techniques needed for test construction.) This test will be given to the pupils before the unit is taught and after the unit is taught. Using simple statistical techniques it can be determined whether or not there was pupil gain as measured by the test.

II. Secondary Indicators

1) Minimal teaching obligations

Minimal teaching obligations are duties or minimum requirements for teachers. Fulfilling these duties does not normally count for the teacher, but not fulfilling them does account against the teacher. Typically these minimal obligations are being on time for the class he/she is to teach (except when ill etc.), testing and marking pupils fairly, knowing the subject matter of what he/she is teaching, etc. The best judges of whether the student teacher is fulfilling these minimal obligations are the primary consumers of his teaching, the pupils. However, as for the overall evaluation, the responsibility for determining if the minimal obligations

are being fulfilled may have to fall on the sponsor teachers and U.B.C. advisors.

2) Out-of-Class Contributions

In the program, the student teachers are part of the school. The school is not simply a geographical location, different from that of the University, where classes are held. Therefore, the student-teachers should do more than teach and attend classes, and they should teach other than by being in a classroom. Their public and professional service to the school, staff and pupils through extra-curricular activities should be acknowledged. The school administration, the sponsor teacher and the U.B.C. advisors would probably be the best judges of the degree of out-of-class contributions.

3) Description of Teaching

"Description of Teaching" is perhaps a misnomer for this section. The teaching act is not to be described; instead, there is a check made for undesirable extremes in the student teachers' teaching style. These extremes are common sensical. For example, it is unlikely that a good teacher would constantly use oral or written language which was too difficult for the pupils to understand; or it is unlikely that a good teacher would be consistently disorganized. As before, it would be desirable to have the pupils respond to this checklist; however, if that is not possible, then the sponsor teachers, and U.B.C. advisors should complete the checklist on the basis of many classroom visits.

4) Performance in University Courses.

The student teachers are also students of the University. Therefore, they have certain standards to achieve in their university courses. Whether or not a student teacher has met the standards for a particular course will

be determined by the professor teaching that course. These marks are not indicators of the student teachers teaching effectiveness, but they are necessary for establishing an overall evaluation of the student teachers' performance in the program.

The Evaluation Procedures

Each sponsor teacher and each U.B.C. advisor will receive a booklet for each student teacher under his/her supervision. The booklet will be organized as follows:

COVER - Student teachers' name, grade he/she is teaching,
Sponsor Teachers' or U.B.C. Advisor's name.

Page 6 - Overall evaluation of teaching rating scale.

Page 5 - Minimal teaching obligations checklist

Page 4 - Out-of-Class Contributions

Page 1, 2 and 3 - Description of teaching checklist.

The measurement of pupil gains will be handled by the student teachers themselves. The grades obtained on the U.B.C. courses will be collected by the evaluator when those marks are available. The various ratings will be made on the basis of as many classroom observations as possible by the Sponsor Teacher and the U.B.C. advisor.

All of this data will be entered on the student teacher's teaching evaluation form. This teaching evaluation form (T-form) will be an averaging and condensing of the above information from the sponsor teacher, U.B.C. advisor and (if possible) the pupils.

| Inputs to T-Form | Method of Obtaining Inputs |
|---|--|
| <ol style="list-style-type: none"> 1. Pupil estimate of teaching merit if possible. Otherwise, sponsor teacher and U.B.C. advisor over-all evaluation of teaching. 2. Estimates of classroom justice, e.g., appropriate quizzes and materials, fair hearing for complaints, etc.* 3. Non-classroom contribution to teaching. 4. Learning gains due to student teacher. 5. Course-work in university courses related to teaching. | <ol style="list-style-type: none"> 1. Single question calling for an evaluation of the teaching skill on a simple A-F scale (special consideration for K-3). 2. Program of adequate random visits to yield reliable judgments, plus detailed study of materials, quizzes, etc. Visits by school staff and U.B.C. staff. 3. School staff, peers, and possibly pupils. 4. Controlled study using pre- and post-tests etc., done by student teacher with advice of evaluation specialist. 5. U.B.C. staff. |

*If it is really necessary to do so, school and U.B.C. staff could estimate merit of teaching style of student teacher.

Each of these inputs contributes something in the direction of good teaching.

The optimal model for weighting these inputs is equal weighting. This means that in a summative evaluation of the student teacher each input has equal weight.

Any evaluation of a student teacher by the pupils (if performed) should be made available to the student teacher. Anonymity of the pupils should be assured.

Hopefully, in the years to come, a better way to evaluate teaching will come about. But, meanwhile, let's not use something worse.

Note: This is the final revised form 164.

EVALUATION OF STUDENT TEACHERS

STUDENT TEACHER NAME _____

DATE _____

(please put your name in the appropriate space)

SPONSOR TEACHER _____

U.B.C. ADVISOR _____

(please indicate when the student teacher was observed, and what subject and grade he/she was teaching)

| Date of observation | Subject being taught | grade |
|---------------------|----------------------|-------|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

DESCRIPTION OF STUDENT TEACHING

Please rate each of the following descriptions of the student teacher's teaching on the scale beside each question.

SA = strongly agree (exceptionally good performance)

A = agree (acceptable performance)

U = undecided

D = disagree (performance not at an acceptable level)

SD = strongly disagree (exceptionally poor performance)

1. The student teacher presented the content of the lesson at an appropriate pace.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

2. The student teacher's lessons are presented in a well organized manner.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

3. The student teacher's oral and written English are appropriate for the grade level he/she is teaching.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

4. The student teacher has an effective and fair system of classroom control (discipline).

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

5. The student teacher teaches the content of the lesson at an appropriate level of difficulty for the grade level.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

6. The student teacher uses audio-visual material when appropriate.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

7. The student teacher accepts and acts on constructive criticism.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

8. The pupils react positively to the student teacher.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

comments: _____

9. The pupils appear to be learning the course material being taught them by the student teacher.

| | | | | |
|----|---|---|---|----|
| | | | | |
| SA | A | U | D | SD |

Will you please give the evidence you used to determine the rating given in no. 9.

10. Please comment on any other traits which you think are contributing to, or decreasing the student teacher's teaching effectiveness.

OUT-OF-CLASS CONTRIBUTION

1. Has the student teacher contributed to the organization and/or operation of school sports?

☐

never

☐

seldom

☐

sometimes

☐

frequently

☐very
frequently☐don't
know

2. Has the student teacher contributed to or participated on school committees?

☐

never

☐

seldom

☐

sometimes

☐

frequently

☐very
frequently☐don't
know

3. Has the student teacher spent out-of-class time (for example, lunch time or after school) helping pupils in their school work?

☐

never

☐

seldom

☐

sometimes

☐

frequently

☐very
frequently☐don't
know

4. Has the student teacher contributed any time or effort to organized social events in the school, such as plays, skits, assemblies, pageants, etc.?

☐

never

☐

seldom

☐

sometimes

☐

frequently

☐very
frequently☐don't
know

5. Has the student teacher volunteered for or participated in any extra-curricular activities not covered by the above four questions? If so please briefly note the nature of the activity, and the amount of participation.

MINIMAL STUDENT TEACHER OBLIGATIONS

1. The student teacher spends an adequate amount of time consulting with the sponsor teacher.

☐ ☐
yes no

☐
undecided
(to teach)

2. The student teacher failed to appear on occasions without adequate explanations; or was frequently very late.

☐ ☐
yes no

☐
undecided

3. The student teacher gives the pupils enough opportunities to ask questions.

☐ ☐
yes no

☐
undecided

4. The student teacher gives satisfactory responses to the pupils' questions.

☐ ☐
yes no

☐
undecided

5. The student teacher seems seriously lacking in knowledge of the subject matter he/she is teaching.

☐ ☐
yes no

☐
undecided

6. The student teacher is available to the pupils, when needed, for out-of-class questions or extra help.

☐ ☐
yes no

☐
undecided

7. The tests and marking of the student teacher seem fair.

☐ ☐
yes no

☐
undecided

8. Further comments on the above ratings, or further information? _____

OVERALL EVALUATION OF TEACHING

Will you please give your overall rating of the job the student teacher is doing in communicating the knowledge, understanding, skills, and attitudes that seem to be appropriate for these grade levels. In answering try not to be influenced by your liking or disliking for the student teacher as a person.

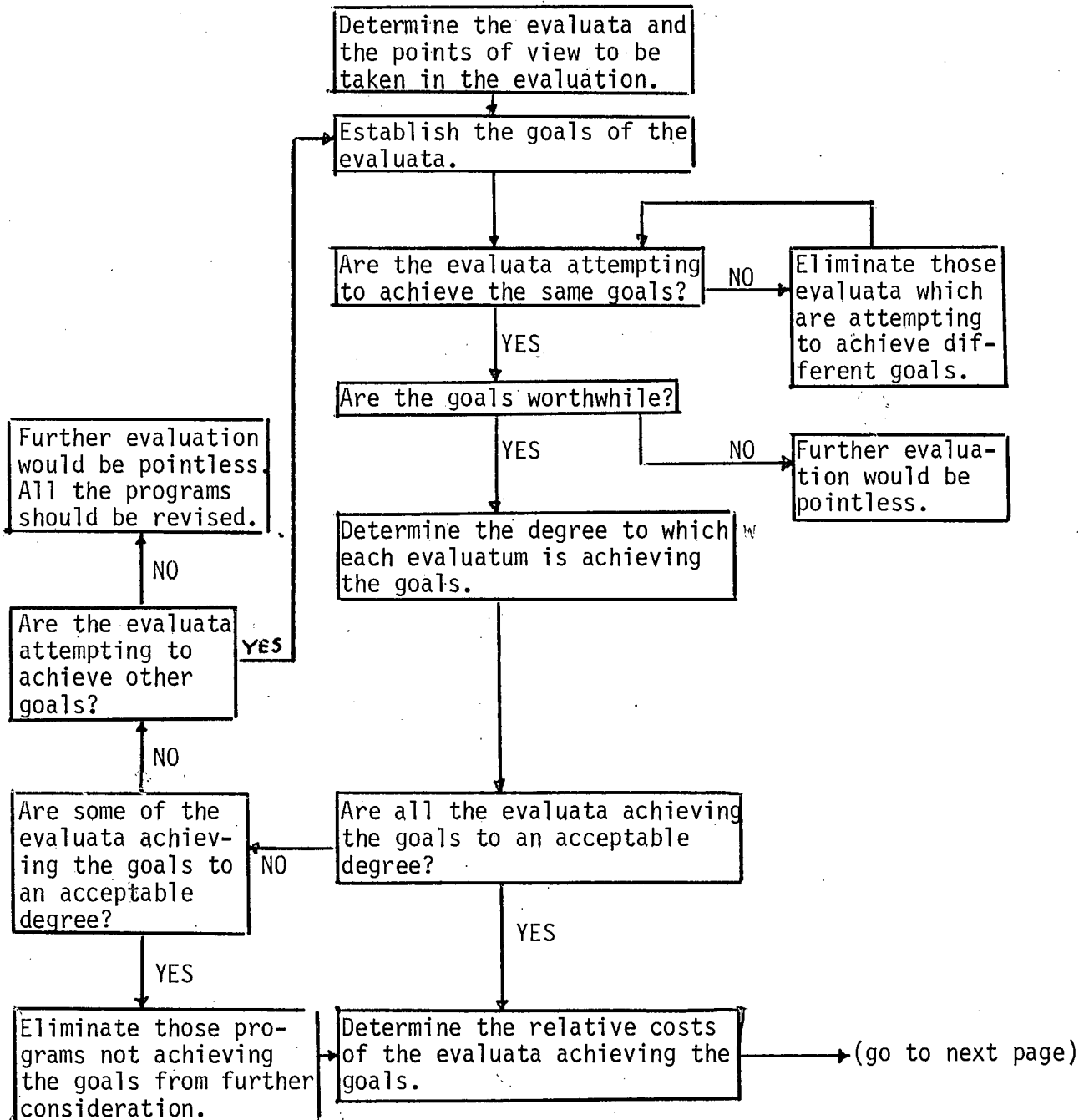
| | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Outstandingly Good | Above Average | Average | Below Average | Outstandingly Bad |

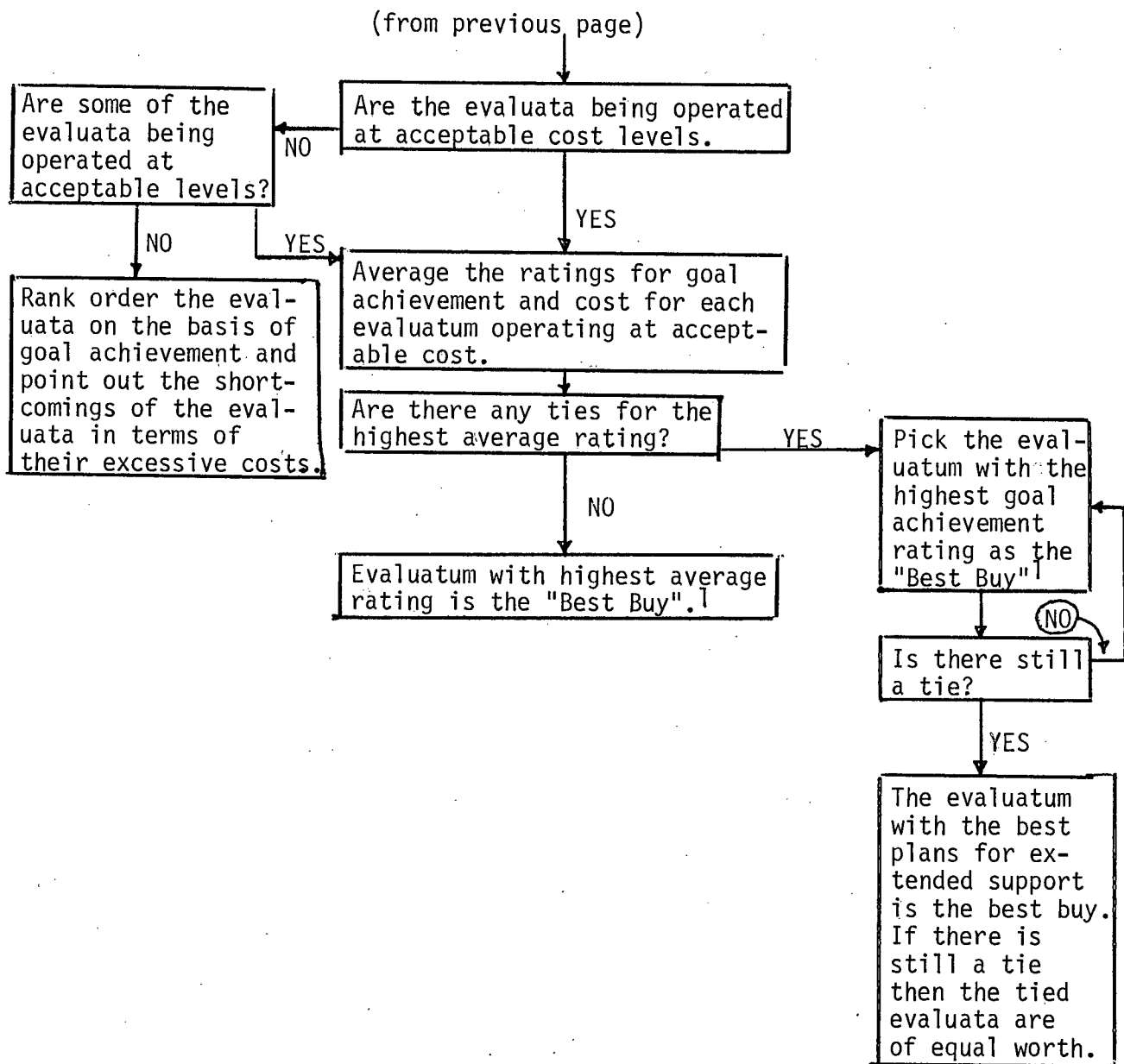
Comments to clarify the above rating? _____

Appendix U

Summary of Evaluation Steps in RANKING.

Steps in RANKING. (See Sec. 3.70 of this thesis)





Note: ¹ "Best Buy" means that the evaluatum, X, offers the best performance for less cost than the evaluata to which it is being compared.