

A GOAL-FREE EVALUATION OF ELEMENTARY PRACTICUM PROGRAMS

by

Gerhard Alfred Ameis

A THESIS

Submitted to the

University of Manitoba

in partial fulfillment of the requirements

for the degree of

MASTER OF EDUCATION

Department of Curriculum: Mathematics and Natural Science

December/1986



A GOAL-FREE EVALUATION OF ELEMENTARY PRACTICUM PROGRAMS

BY

GERHARD ALFRED AMEIS

A thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

MASTER OF EDUCATION

© 1987

Permission has been granted to the LIBRARY OF THE UNIVERSITY OF MANITOBA to lend or sell copies of this thesis, to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film, and UNIVERSITY MICROFILMS to publish an abstract of this thesis.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

ABSTRACT

A GOAL-FREE EVALUATION OF ELEMENTARY PRACTICUM PROGRAMS

The purpose of the study was to evaluate four elementary practica (Faculty of Education, University of Manitoba) by using a goal-free evaluation model. The practica were; (a) Faculty-Based, (b) School-Based, (c) St. Vital Project and (d) Redekopp (based at Shamrock School).

Two broad questions were addressed; (1) Can the theory of goal-free evaluation as the means of merit determination in the evaluation of practica be operationalized? and (2) Can the merit of the practica be determined by this means?

This model for evaluation led to conclusions and shows promise for other evaluation studies.

All practica provided organization and experience needed for fostering confidence and expertise in teaching and met some needs. There were some differences between practica. The Redekopp practicum may be better at preparing pre-service teachers for teaching.

It was recommended that future practica would benefit from incorporating (1) self-initiated planning of experiences, (2) self-evaluation of progress, (3) extensive contact with schools early in the school year and (4) immersion in the various aspects of schooling. The indications are that these features are important to the task of developing effective teachers.

ACKNOWLEDGEMENTS

The writer wishes to express his appreciation to all the people who assisted in this study.

Special thanks is extended to Dean John J. Stapleton for providing assistance in the form of a research grant.

Deepest gratitude is extended to Dr. Eric MacPherson, the writer's advisor, for suggesting this study as a possible project and for his many insights, and invaluable advice and guidance.

Appreciation is extended to the members of the committee; to Dr. Joanne Keselman for her careful reading and contribution to the statistical analysis and to Dr. Murray McPherson for his interest and commitment to improving practicum programs.

Finally, appreciation is extended to the fourth year students who participated in the study and to the principals and co-operating teachers of the host schools.

TABLE OF CONTENTS

	page
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
 Chapter	
1. INTRODUCTION	1
The Focus of this Study	1
Rationale for the Evaluation of Practica	1
Rationale for the Goal-Free Approach	3
Statement of the problem	4
2. THE NATURE OF GOAL-FREE EVALUATION	6
Introduction	6
Characteristics of GFE	8
Justification for Merit Determination in GFE .	10
Issues in GFE	10
Implied GFE Procedure	13
Examples of Goal-Free Evaluations	14
3. REVIEW OF THE LITERATURE	17
Overview	17
Evaluation	17
The Concept of Evaluation	17
Historical Perspective on Educational	
Evaluation	19
Theoretical Perspective on Evaluation	21

Models	23
Purpose of a Model	23
Properties of a Model	23
The Goal-Based Evaluation Model	25
Introduction	25
Characteristics of GBE	26
Issues	27
A Comparison of GBE and GFE	29
The Act of Teaching	31
Teacher Characteristics	31
Competency Measure	32
4. THE MODEL FOR THIS STUDY	35
Rationale for the Goal-Free Approach	35
Boundary of the Evaluation	37
The Components of the Evaluation Model	38
The Effective Teacher Model (ETM)	40
Role of the ETM	40
The Term 'Power' in the ETM	41
Terms and Variables of the ETM	42
The Testing of the ETM	44
Evaluation Instruments	45
Instrumentation Concerns	46
External Credibility	47
5. THE METHODOLOGY OF THE STUDY	49
Preparation for the Evaluation	49
Initial Contact with Practica	49
Construction of the ETM	50

Testing of the ETM	50
Selection of Participants	51
Construction and Field-testing of Instruments	51
Collection of Evaluation Data	52
Analysis of Evaluation Data	53
Equivalence of Practica Populations	53
Analysis from a Program Perspective	54
Analysis from a Pre-service Teacher Perspective	55
Analysis of Other Outcomes	57
6. RESULTS AND CONCLUSIONS	58
Overview	58
Section 1: Testing the ETM	59
Section 2: Equivalence of Practica Populations	67
Section 3: The Programs' Perspective	75
Section 4: The Participants' Perspective	82
Section 5: The Perspective from outside the ETM	93
Section 6: Effects of Participation in the Study	93
Section 7: Conclusions Concerning GFE	93
Section 8: Conclusions and Recommendations for Practica	94

BIBLIOGRAPHY	98
APPENDICES	101
A. Observation 1: Part 1 - Classroom Observation .	101
B. Observation 1: Part 2 - Pre-service Teacher	
Survey	102
C. Observation 2: Part 1 - Classroom Observation .	103
D. Observation 3: Part 1 - Classroom Observation .	106
E. Observation 3: Part 2 - Survey: Co-op Teacher .	107
F. Observation 3: Part 3 - Post Observation Survey:	
Pre-service Teacher ...	109
G. Observation 3: Part 4 - Interview: Pre-service	
Teacher	114
H. ETM Survey Form A	115
I. ETM Survey Form B	119
J. ETM Survey Form D	123
K. Observations from a Perspective outside the ETM	126

CHAPTER 1

INTRODUCTION

The Focus of this Study

The immediate purpose of this study is to provide information that would be useful to program managers when making decisions concerning the modification or elimination of existing practicum components of four teacher-education programs at the University of Manitoba.

There are other purposes. Further research is needed on evaluation of practica, on goal-free evaluation and on characteristics of effective teaching. This study explores these areas as well and, hopefully, increases understanding about them.

Rationale for the Evaluation of Practica

The teacher is at the center of the education delivery system. To be effective in the classroom a teacher must possess the skills and knowledge that are considered to be characteristic of effectiveness. An image of the effective teacher, drawn ultimately from an image of humanity, is essential for this.

Teacher-education institutions sharpen this image, and then design and deliver programs intended to equip teachers-to-be with the skills and knowledge inferred from that image. The effectiveness of their programs for that purpose is at least a major concern of these institutions

and the determination of this effectiveness is central to any evaluation of teacher-education programs.

But there are other concerns. A need for evaluations of practica is evident. A search of the literature yielded minimal information on such evaluations in general and no information on any evaluations of practica utilizing a goal-free model. As well, various researchers had attested to the paucity and inadequacy of past evaluations.

According to Kurachi and Irving (cited in Adams, 1983), the status and extent of evaluations of teacher preparation programs is largely unknown. Hord, Savage, and Bethel (1982) reported that existing evaluations of teacher education programs had focused primarily on follow-up studies of graduates; collecting data on their perceptions of and satisfaction with their programs; and on ratings of the graduates' performance by supervisors, other teachers and administrators.

They concluded that, despite internal needs and external demands, the state of the art of evaluation in this arena had improved very little over the past few decades.

Strathe (cited in Hord, Savage, Bethel, 1982) argued that evaluation of teacher education programs is beset with conflicting results, lack of clarity and agreement about evaluation designs, and lack of adequate focus.

At the root of these problems lies a lack of consensus (Gardner, 1982) as to what constitutes good teacher behaviour, how to measure it and, in fact, whether or not it is measurable. Jackson, in the same compendium, suggests that "there is no logical or empirical reason for believing that there must be a best way of teaching something to someone." (p. 37). Recently, Clarken (1983) reported that there is no consensus on operational definitions for teacher competency and effective performance.

This lack of consensus, coupled with the difficulty of evaluating according to even agreed-upon criteria, has limited the value of most of the systematic evaluations of graduates of teacher-education programs.

Rationale for the Goal-Free Approach

According to Evers (1980) educational evaluation has undergone substantive theoretical growth during the past three decades but many methods have not yet been put into practice. Smith (1982) also notes; "In recent years there has been increased interest in the development of new methods for conducting evaluations... As new approaches are identified, however, the problem arises of how to determine the usefulness of these new methods. This problem is exacerbated by the proliferation of new approaches." (p. 3).

The operation and utility of these evaluation methods, in particular Scriven's goal-free evaluation method, is

open to question. "There... still isn't, any standard protocol for doing goal-free evaluation. The early literature... was mainly philosophical in nature and contained little operational guidance." (Evers, 1980, p.10). Few examples [3] of goal-free evaluations are reported in the literature (Evers, 1980) and these do not adequately address questions concerning operation and utility. Consequently, there now seems to be a gap between theory and practice. The need to lessen this gap provides an important justification for utilizing the goal-free model in this study.

Statement of the Problem

The specific program evaluation needs of the Faculty of Education at the University of Manitoba coupled to broader issues concerning the evaluation of teacher-education programs in general and the value of exploring the goal-free evaluation technique led to the study of two broad questions:

1. Can the theory of goal-free evaluation as the means of merit determination in the evaluation of pre-service teacher-education practica be operationalized ?
2. Can the merit of four practicum programs be determined by this means ?

The more specific questions generated by these questions were:

1. How is a theory of goal-free evaluation made operative?
2. Do any differences exist between the current University of Manitoba practicum programs in producing an effective teacher?
3. Do any differences exist between the practicum programs in meeting the needs of pre-service teachers?
4. How useful is this realization of the goal-free model for the evaluation of practicum programs?
5. What are the interaction effects between the practicum programs, the faculty programs, the pre-service teachers and the host schools?
6. What is an effective teacher?

CHAPTER 2

THE NATURE OF GOAL-FREE EVALUATION (GFE)

Introduction

The central purpose of any evaluation is to determine merit. It must therefore begin with evaluation criteria. Basically, there are two possible sources of criteria; from an examination of program goals and by an adaption of some set of external and presumably more general goals to this particular context.

The dominant procedure has been, and remains, the derivation of criteria from the program being evaluated. In 1963, Cronbach discussed dangers associated with using objectives derived from program goals (Evers, 1980). His position was that the result was often dangerously narrow, and that a valid evaluation should examine outcomes and side-effects over the widest possible range.

The goal-free model of evaluation relies on external sources. A primitive form of goal-free evaluation is therefore being practiced whenever some external standard is used in the evaluation of a program. However, this in itself is not sufficient to characterize goal-free evaluation. The position is elaborated in Stake's and in then Scriven's writings.

Stake (Henderson, 1978) argued that a focus on program goals prevents evaluation from considering the complex and dynamic nature of education.

Scriven is mainly responsible for establishing this as an identifiable position. He built on his earlier insight concerning formative (developmental) and summative (final) evaluation. He stressed that traditional 'goals' are mainly useful in formative evaluation. Summative evaluation, he believes, should pay attention to all program outcomes, and hence to externally generated goals. There are several supports for this position.

First, goal attainment evaluation is pointless if the goals themselves have little merit. Evaluation must first include, as an equal partner with some measure of performance against goals, a process for evaluating the goals themselves. This way some assurance may be gained concerning the merit of particular goals. As well, the rhetoric of intent (goals) no longer may be used as evidence for success; thereby limiting practices of using program goals as evidence for program excellence (Scriven, 1974a).

Second, program goals are only a subset of possible outcomes, all of which are important to merit determination. A goal-based approach introduces a possible bias or tunnel vision into an evaluation as a consequence of restricting the evaluation process to a search for a specific set of realized effects.

Third, Scriven sees a separation of effects into secondary (unintended) and intended as an impediment to

effective evaluation. The language itself, implying lower status for unintended effects, tends to detract from what might be relevant or critical program achievements. As well, the implied lower status reduces the commitment to look for these effects.

Scriven formulated the goal-free evaluation model and offered it as a remedy for the perceived inadequacies of existing evaluation models. Interestingly, Scriven later suggested that an evaluation be conducted utilizing both the goal-free and the goal-based models. This is seen as the best of two worlds.

Characteristics of GFE

Scriven's position was that the evaluator's responsibility is to serve as an enlightened surrogate consumer concerned with the welfare of society as a whole. In 1977, he defined goal-free evaluation as follows;

"the evaluator ... enters into the evaluation with the purpose of finding out what the program is actually doing without detailed cueing as to what it is trying to do... Merit is determined by relating program achievements to the needs of the impacted population rather than to the program... goals." (cited in Evers, 1980, p.8)

As conceptualized by Scriven (1981), GFE must take into account the actual outcomes of a program, not merely the intended outcomes. The evaluator must make a deliberate effort not to be absorbed by program goals. They are appropriate considerations for program planning and development but not for program evaluation.

Data must be gathered over a spectrum of actual outcomes and these outcomes should be evaluated with respect to their meeting or not meeting extrinsic criteria. The program goals may be inferred from the outcome spectrum but they are irrelevant to the evaluation.

Scriven interprets 'objectivity' as a viewing from above and afar, in a way hopefully more relevant to discerning merit. Goal-free evaluation therefore rejects the assertion that a total specification of desirable events is adequate for determining truths about reality.

In summary, the goal-free evaluator tries to find out what a program is actually doing as opposed to what it is supposed to be doing. He/she is concerned with outcomes per se -- good, bad, anticipated or not. Evaluation becomes consequence-oriented, not objectives-oriented. Merit is determined by relating program outcomes to the needs of the impacted population and of society as a whole.

Justification for Merit Determination in GFE

Underlying merit determination is a question fundamental to evaluation itself. What justifies merit determination? That is to say, what principles justify evaluation and, consequently, by what principles should an evaluator be guided?

The viewpoints and insights that address these questions do not split into convenient categories but it may help clarify the matter to examine two polar positions.

One position reflects the needs of individuals. Something has merit when it fulfills a human need. The opposite position reflects the needs of a created social system. Something has merit when it is seen as contributing to a system functioning as it is planned to function.

These polar positions are at the heart of evaluation. Evaluation is needed because human needs or system requirements must be met. Any evaluator must be guided by these positions, whether in pure or blended form.

Goal-free evaluation incorporates both.

Issues in GFE

There are three clusters of issues involved in goal-free evaluation; a) those related to risks implicit in GFE, b) those related to formulation, and c) those related to methodology.

From the perspective of a manager or producer of a program, there is an obvious threat implicit in a GFE. He

or she loses a measure of control over an evaluation since the program goals are ignored by the evaluator. Second, a GFE can induce a degree of discomfort in individuals who are accustomed to working within predetermined frameworks. Third, the use of a goal-free model increases the risk of a program not being found meritorious (Scriven, 1981). The accompanying shock waves can surge up as far as the funding level.

There is the problem of contamination -- an evaluator inferring the goals of the project. One solution is to reduce the amount of data going to the evaluator, but the trade-off is an increase in the risk of missing some intended outcomes. It is one thing to avoid being captured by program goals. It is quite another thing to miss significant effects related to those goals.

How does a goal-free evaluator test for all possible outcomes? While Scriven stops short of providing clear operational guidance on this point, he emphasizes the importance of a thorough survey of experience, opinion and the literature.

Concerning objectivity, House (1983) comments:

"Scriven employs the qualitative notion, in which objectivity is equated with freedom from bias... In Scriven's framework, it is possible for a single observer, unaided by any psychometric instrumentation, to be more

objective than a battalion of observers loaded with reliable instruments - if the single observer is looking for the right thing and is sufficiently protected from biases." (p. 55-56)

An analogy might be useful. Change in human weight is apparent in a number of ways; by touch, by sight, and by measurement. But if one is not looking for it then the finest weighing instrument will not be used. An observer who is sensitive to and uses less precise instruments (such as seeing) might be more likely to notice changes in weight.

Concerning methodology, the central problem remains that of finding meta-standards. This problem is not unique to GFE, but elsewhere it is implicit and may be finessed. Here it must be pressed directly. Scriven (1974a) rejects the substitution of a goal-free evaluator's own goals for the program goals. The standard must come from broadly based needs or goals. The difficulty lies in making explicit what these are.

Scriven (1974a) distinguishes between a goal and a standard. For him, goals and needs can become standards. However, the program's goals can never become the standard unless those goals have been validated by external judges. It is therefore perhaps a misnomer to call Scriven's approach 'goal-free'; a better descriptor might be

'goal-blind'. The point is that standards for merit determination are constructed without reference to the goals of the program being evaluated.

Implied GFE Procedure

Evers (1980) is the main commentator on the operationalization of GFE. He summarized and sometimes extended Scriven's writings on the matter.

Scriven sees GFE as requiring filters or buffers to keep the evaluator from gaining knowledge about program goals. The goal-free evaluator initially should review or examine general program information that has been divested of goal rhetoric.

The following reflective stage serves the purpose of generating a set of potential hypothetical statements about the nature of the outcomes prior to observing them. Two main techniques are described.

In the first, the evaluator relies on professional judgment, an approach that presents problems of validity and objectivity. However, an advantage is gained. Professional expertise and a large set of variables are exploited.

The second relies on an identification of the established needs of the target population or consumer of the program. It has the advantage of assessing merit in terms of a program's service to people. However, needs data

is often not available and any collecting of it can entail substantial resources.

The next stage of the evaluation is the collection of evaluation data, the actual outcomes of the program. Once this task has been completed, the initial hypotheses and data must be presented to the evaluatees and assessed as to importance (the analysis of the data). According to Scriven, a number of ways are possible.

Needs data (existing or forthcoming), the referenced program's goals or a panel of experts and others external to the program can be utilized to judge the effects. In the extreme case, if the goals of the program had been initially constructed utilizing data from a diligent needs assessment, then these goals can serve as criteria to assess the observed outcomes of the program.

Examples of Goal-Free Evaluations

Evers (1980) reported on the only examples of goal-free evaluations conducted to that date. Of the three examples described, one had been conducted by Scriven. The focus of the evaluations was on curriculum materials and the main thrust was to give the client information from a perspective that was outside the developmental process.

All three evaluations reported some means of limiting contamination by knowledge of program goals -- either limiting contact with evaluatees or using liaison persons.

The reported establishment of external credibility (perceived freedom from evaluator bias) varied. Two evaluators were used in the House/Hogben study (cited in Evers, 1980), an approach that has a minimal check for bias. Welch (cited in Evers, 1980) used a panel of five judges who rated the materials independently (moderate check). Scriven employed extensive checks against bias. He used a project manager to screen materials from the projects, a process for replacing any contaminated evaluators, a process for training and calibrating evaluators and a standard checklist for use in the observation of treatments. Scriven also used multiple site visits across time and the use of multiple evaluators per site.

The design of the evaluations varied, but not substantially. Scriven and Welch used a standard checklist as instrumentation for observations and interviews. House/Hogben gave no indication of the instrumentation used. Welch and Scriven did not utilize sampling. All sites and materials were reviewed. House/Hogben did not provide any information on sampling.

The goals of the curriculum projects were considered in different ways. In the House/Hogben study, the evaluators interviewed the project staff to obtain the goal priorities of the projects. In Scriven's study, the evaluators first wrote a goal-free report and then reviewed

the project's materials. In Welch's study, the use of project goals was not mentioned.

The studies reported different approaches to data analysis. House/Hogben used a narrative professional judgement. Scriven's study utilized raw reports that were analyzed by Scriven and independently by the evaluation project manager. These were then merged by the staff of the evaluation project. Welch's study converted the alphabetic judgements of four of the judges to numeric form and these were averaged by the fifth judge. Each study provided the clients with a written report.

In conclusion, it was apparent after a search of the literature, that there is no blueprint that clearly lays out all the necessary steps for making a theory of goal-free evaluation operative. As a consequence, the author of this study relied on what can be inferred from the above studies and from the theoretical literature on goal-free evaluation in designing this study.

CHAPTER 3

REVIEW OF THE LITERATURE

Overview

The review of some of the literature has been anticipated in chapters one and two. This chapter focuses on more specific matters of evaluation, models, goal-based evaluation, and on teacher characteristics insofar as they have a direct link with the problem at hand; the goal-free evaluation of practicum programs.

Evaluation

The Concept of Evaluation

People have made choices for centuries but formal evaluation is relatively young. There is consensus (Scriven, 1974b) that educational evaluation comprises one or more value judgements which assign worth to something.

For Scriven, values are derived from desires, needs and directed action. He described evaluation "as a process of compressing complex data in the vise of these contextual constraints so as to squeeze out the water and leave behind the meaty residue of directed information that is a value judgement" (Scriven, 1974b, p. 66).

While this definition leads naturally to the conclusion that evaluation is intrinsically subjective and relativistic, Scriven seems to have been the first to have dealt explicitly with the implications of that conclusion.

Scriven argues that "there can be no sharp line drawn between the search for truth and the search for merit since merit is one kind of truth" (Scriven, 1983b, p. 77).

For Scriven (1974b), value judgments are one form of scientific judgement and, in fact, science is not value-free. All science, he argues, is concerned with finding the best explanation for a phenomenon. It cannot be said that something is a good explanation of a phenomenon if a better explanation exists. Thus, any scientific statement is a statement of a measurement as well as a value statement.

From this point of view, evaluative terms and concepts are inherent in science. There can be no fact-value distinction. Scriven claimed that "evaluation is the only essential scientific activity" (Scriven, 1983b, p. 79). When value judgements are correctly supported they are as totally objective as any statement of science (Scriven, 1974b). Thus, the issue is the formulation of correct supports of and methodology for evaluation.

The major problem to be overcome or minimized in the development of correct supports of and methodology for evaluation in education is the multitude of variables which could have relevance in the collection and interpretation of data. They can rarely be controlled, as they can in the physical sciences, by such classic methods as randomization.

Since the subsequent questions of what to measure and how to measure are heavily conditioned by feasibility, past responses to these questions reflect the variety of ways that educators have deemed it possible to conceptualize evaluation.

Historical Perspective on Educational Evaluation

From the American perspective, Madaus, Stufflebeam and Scriven (1983) partitioned the history of educational evaluation into six periods;

1. During the Age of Reform (1800-1900), evaluations were loose and sporadic efforts. Undefined feelings were used as the basis for making judgements. Gradually, test scores became used as criterion measures and differential school results were used to establish the validity of some positions and to refute others.

2. The Age of Efficiency and Testing (1900-1930) accompanied the dominance of 'scientific' management in administrative theory and in industry. The emphasis was on systematization, standardization and efficiency. Educational evaluations of this era focused on school/teacher efficiency as measured by achievement, dropout rates, etc.. Norm-referenced tests were created and used to influence local practices.

3. The Tylerian Age (1930-1945) accompanied a renewal in education led by John Dewey and his contemporaries. Tyler had a considerable influence on education in general

and on evaluation in particular. His model of internal comparisons of objectives and outcomes was a major theoretical step.

4. The Age of Innocence (1946-1957) was characterized by extremes. It was a time to forget the war and enjoy the good life. Standardized testing had expanded enormously. The Tyler approach was simplified and 'objectives' came into vogue. Bloom's taxonomy was developed during this age. The spirit of technique dominated -- every problem can be solved by an appropriate technique. Nevertheless, there was little interest in holding educators accountable and thus little evaluation was actually conducted.

5. The Age of Expansion (1958-1972) was heralded by Sputnik and the accompanying threat to national security in the United States. Evaluation was reconceptualized, not as a horse race between competing programs, but as a process of gathering and reporting information that could help guide curriculum development. Standardized tests were recognized as being of little use in diagnosing needs and assessing gains of disadvantaged children.

6. The Age of Professionalism (1973 -) saw the crystallization of the field of evaluation as a distinct profession. Universities began to offer courses in evaluation methodology. Inevitably, polarization (positivistic/quantitative versus

phenomenological/qualitative) accompanied this growing status.

Popham's (1975, p. 1) comments, "Educational evaluation is in the air. Indeed, demands for educational evaluation are so prevalent that most educators must believe they are living in an evaluation generation.", clearly describe this period. It saw the emergence of the distinction between goal-free and goal-based evaluation.

The central problem in an evaluation remains the determination of merit whether the approach is goal-free or goal-based. However, they diverge as to what provides the basis for merit and how to proceed in the determination of it.

Theoretical Perspective on Evaluation

The basic question has to do with the source of values for merit determination. House (1983, p. 45) stated that "all evaluation models are based on variations in the assumptions of the liberal ideology, or if one prefers, the conceptions of liberal democracy". This ideology embraces the key ideas of individualism and empiricism.

From this viewpoint, values are based on principles derived from subjectivism and objectivism. Evaluation models are constructed with these principles in mind. A clearer picture of the nature of evaluation models emerges if the extreme positions of these principles are examined.

Where truth is seen as determined by human nature, values are derived from the concept of the maximization of happiness, an idea traceable to the epicurean philosophy, John Stuart Mill and Utilitarianism. Evaluation models based on this viewpoint try to base judgements on overall social utility. Scriven's goal-free model seems to belong in this camp.

Evaluation models based on objectivist principles, where reality must be quantitatively measurable in order to exist, perceive of evaluation as something of a watchdog. Accountability, efficiency and quality control rather than social utility are the themes. Values are ignored or are inferred from objective principles and processes. Goal-based models of evaluation seem to belong in this camp.

These justifications for value judgements reflect theoretical principles that are meant as guides. They also reflect different responses to a deep question, namely; what is the 'ideal' to draw from for guidance? As such, they reflect some sensing of what the 'ideal' may be and represent a way to attain it.

Models

Purpose of a Model

The purpose of a model is to specify variables, define them and focus on possible relationships between them. The nature of observations made depends on these agreements. That is to say, our models shape our perception of events.

In this study, we require some model which defines effective teaching and a way of searching for that outcome in evaluations of practicum programs. Although Willer (1967) discusses models in relation to theory development, his discussion has direct applicability here.

Properties of a Model

A model is neither real or fictional. It is a way of looking. "It is a wrong question to ask of any model its relative 'reality'; instead the model must be judged as relatively useful or useless." (Willer, 1967, p. 24)

The essential components of a model are the rationale, the definitions and the mechanism. The rationale is a point of view that provides a basis for creating nominal definitions. For definitions, the path proceeds from nominally defined concepts to operational terms. The mechanism is the interconnection or structure of the relationships assumed to exist between the categories of the model.

When operational definitions are substituted for their corresponding nominal terms then testable hypotheses are possible. Thus, indirectly, nominal definitions are testable. The degree to which nominally defined variables are in fact measured depends on the degree of correspondence between nominal and operational definitions.

Models are not true or false. "We may measure in reality, but we do not measure reality; we measure concepts nominally defined." (Willer, 1967, p. 87)

Willer proposes three types of models; (1) analogue models that borrow a mechanism from some other realm, (2) iconic models that attempt to simulate properties of empirical phenomena and (3) symbolic models that relate the concepts themselves. A symbolic model is used in this evaluation study.

The Goal-Based Evaluation Model (GBE)

Introduction

Measurement dominated evaluation prior to the Tylerian Age, with norm-referenced tests being the instruments. This approach can be viewed as a primitive form of goal-free evaluation, as norm-referenced tests can be considered to be a set of external standards.

During that age, evaluation essentially involved a horse race between competing programs. If the objectives of a program did not take the content of the testing instruments into account, the possibility of a favorable evaluation was lessened. It can therefore be argued that the items on the tests became de facto, the goals of these programs. In fact, if not in name, a form of goal-based evaluation was in vogue during that age.

Tyler (1949) subsequently refined this concept of evaluation. He argued that the purpose of evaluation is to make a check as to whether or not the plans for learning experiences (the objectives) actually produce the outcomes desired (objectives realized). In other words, evaluation is to be a process for determining the degree to which these planned-for changes in behaviour actually occur. The culmination of these refinements became goal-based evaluation.

Characteristics of GBE

The 'objectives' approach to evaluation can be dated from Mager's definition (1962) of the ideal behavioural objective (Henderson, 1978). Such an objective must communicate as explicitly and clearly as possible the instructional intents of a program.

It must describe the outcome desired, the conditions under which the outcome is to occur and the criterion for determining an acceptable outcome. This specification of objectives is intended to operationalize them to the point that all participants in a program -- instructors, trainees, observers, and evaluators -- can be clear about the goals of the program and the criteria for evaluation.

Five stages are involved in a goal-based evaluation;

1. The selection of the goals or objectives to be achieved
2. The formulation of behavioural objectives for these goals
3. The development of appraisal instruments to evaluate the desired behaviours
4. The examination of the collected data in light of norms to judge the adequacy of the observed behaviour
5. The making of a final decision as to the match between the observed behaviour and the intended behaviour

This evaluation model became dominant during the 60's. It was supported by the development of taxonomies of

educational objectives expressed in behavioural terms (Bloom -- cognitive domain, Krathwohl -- affective domain, and Simpson -- psycho-motor domain).

The determination of the match between the outcomes (observed behaviours) and the intents (objectives/goals) of a program became synonymous with the determination of the merit of that program. The use of this evaluation model was extended into many educational areas beyond its original focus (Henderson, 1978).

Despite the spirited advocacy of such individuals as Popham and Sullivan, this evaluation movement has waned. It became apparent that the specification of desirable behaviours (the goals/objectives) in behavioral terms and the measurement of the degree of their attainment did not capture the merits of a program in a satisfactory way.

These and other weaknesses of the goal-based approach to evaluation become more apparent when the theory and methodology of GBE are examined.

Issues

Henderson (1978) discusses some of the inherent weaknesses of GBE, particularly those related to behavioural objectives.

There is no agreement on who is to select the objectives, or which ones are to be selected from the array of possible choices. The evaluator must select the more valued goals from which to determine the objectives and

this involves more than simply establishing a congruence between specified goals and outcomes. At issue here is the nature of evaluator bias and how to minimize it.

Once the objectives have been selected and stated in operational terms, these objectives do not and can not determine the validity of the evaluation instruments. Other variables can account for behavioural changes. There is usually more than one test item possible per objective. As a consequence, the task of developing instruments is less certain and more contentious than is originally conceived in GBE.

In fact, the act of specifying in advance exactly what is to be counted as desirable runs contrary to many practices. We prefer to operate within looser frameworks that allow for surprises. Even in the objective disciplines of science, the pre-specification of goals is irrelevant. This rigid pre-specification, in fact, can be viewed as a barrier to promoting or noticing outcomes that result from creative and evolving programs.

If evaluation is no more than relating outcomes to pre-specified desirable behaviours, then all intervening events become irrelevant. But these events are relevant because they are often attached to other goals. These unanticipated relevant side effects tend to be missed by the evaluation.

These weaknesses of GBE seriously reduce its usefulness. Information about the match between outcomes and intentions of any program gives an indication of only a narrow and possibly minor outcome of a program. The merit of this indication is then open to question.

There is another problem with GBE, one related to a different viewpoint on merit determination. The degree to which needs are met may also serve as an indicator of program success. GBE does not provide information on fulfillment of needs. It fails to incorporate an important perspective on merit determination and therefore misses valuable outcomes. This failure further reduces its usefulness.

These internal and external weaknesses of GBE limit the value of such an approach to evaluation. Once identified, they provided arguments that encouraged the development of alternate approaches to evaluation. One of these is the goal-free model.

A Comparison of GBE and GFE

There are important distinctions between goal-based and goal-free models of evaluation. Both use standards but not from the same sources nor in the same ways.

The goals of a teacher education program represent a vision of an effective or ideal teacher. They are meant to describe desirable teacher characteristics. In GBE, the

attempt is made to find out if these goals (the standard) have been attained by selecting a subset and describing these in behavioural terms. The match between intention and outcome is taken as an indication of program success.

In goal-free evaluation, another set of standards derived from some other source(s) is used. In a sense, the program's goals are replaced with goals broader in scope so as to facilitate a more inclusive search for program effects. GFE is open to wide perceptions of outcomes not readily translatable into some set of operationally defined behaviours. It is also open to varied ways of determining program success.

These distinctions make it clear that GBE is easier to apply than GFE. The trade-off is usefulness. Applying GBE is a little like using a fixed-sized dip net cast a few times into a lake to provide you with information pertinent to the habits and characteristics of the fish in that lake.

GFE has its own issues. A critical problem is the determination of the standard to be used for the evaluation. Needs assessments, subjective experiences and the literature can all be sources for this standard. The literature is a suitable starting point.

The Act of Teaching

Teacher Characteristics

Two sources were located which were helpful in the task of determining a standard. They provided information on teacher characteristics and competency measurement.

Smitheran, Anderson, Hillis and Loucks (1971) classified teacher characteristics according to four concepts;

1. In the craft concept, a straightforward and simple knowledge of subject-matter is sufficient. The difference between good and bad teachers depends on personality, not technique, and therefore common sense and experience are sufficient. In short, teaching is an art.

2. The omniscient concept is best portrayed by MacDonald's rush of adjectives. "Teachers are at once invited to refer themselves to an omniscient model, at once intelligent and affectionately warm, knowledgeable and tolerant, articulate and patient, committed and sympathetic, scholarly and practical, ... teachers are, or ought to be secular priests." (MacDonald, 1968, p.21). Here, teachers are comprehensive specialists -- ultimate human beings.

3. The functional expert concept is adapted from the systems approach where efficiency requires a specialist rather than a generalist. The teacher must possess a great

depth of knowledge and intensive skill in a narrow band of roles.

4. In the futuristic concept, the teacher requires skills related to assessing, diagnosing, evaluating, guiding, inspiring, and catalyzing self-discovery. The main function of a teacher will no longer be to transmit information; rather, it will be to act as a guide and resource.

These concepts provide a framework for defining quality in teaching. The Effective Teacher Model (ETM), which is the standard used in this evaluation, was developed within this framework. Most variables of the ETM were derived from the omniscient model with some derived from the futuristic model.

Competency Measurement

According to Weil and Joyce (1978), competency in teaching is expanded by increasing the range of teaching strategies that one can employ and by becoming increasingly skillful in their use. They presented four models of teaching from which characteristics of an effective teacher can be inferred and from which competency measures can be derived.

1. Social interaction models emphasize the relationships of the individual to society and to others. Competency is determined from the teacher's ability to relate to others and to foster pupil participation.

2. Information processing models emphasize the information processing capability of students. Competency is determined from the teacher's ability to involve the pupils in handling stimuli, organizing data, generating concepts and solutions to problems, and utilizing symbols.

3. Personal models emphasize the individual. Competency is determined from the teacher's ability to help students improve their emotional life and health.

4. Behaviour modification models emphasize the development of efficient systems for sequencing learning tasks and shaping behaviour. Competency is determined from the teacher's ability to promote measurable gains in desirable behaviours. Gains not specifiable in behavioural terms are ignored.

These models of teaching, while useful to defining quality in teaching, provide a viewpoint useful to measuring competency.

The variables of the ETM are derived from these models and are therefore measurable through them. For social interaction and information processing variables, competency measurement is possible directly from classroom observations. Observations carry enough information with them to enable low inference decision-making about competency.

This situation does not exist for personal and behaviour modification variables. It is difficult to assess

pupil status and gains only from classroom observations. Only high inference decision-making about competency is possible.

As most variables of the ETM are derived from the social interaction and information processing models, a low incidence of high inference measurements is included in the goal-free evaluation of practica.

CHAPTER 4

THE MODEL FOR THIS STUDY

Rationale for the Goal-Free Approach

There are two distinct justifications for using a goal-free approach instead of a goal-based approach in this study. One is linked to general concerns of evaluation and the other to conditions specific to the programs being evaluated.

With respect to general concerns, the GFE model has the potential advantage of a superior control for bias. It has the potential to expose new insights by uncovering outcomes that were not stated as goals. Program goals may be too vaguely or too narrowly stated limiting their usefulness for deriving standards.

Stephens' (1967) theory of primitive tendencies supports this case. He found that research studies had consistently reported failure to find significant differences in learning outcomes resulting from the application of different programs and curricula. This lack of significant differences led Stephens to postulate the existence of overlooked primitive tendencies. He hypothesized that these tendencies are really responsible for learning and that the typical input variables examined in studies were secondary if not irrelevant.

Goal-based evaluations, because of the highly specific nature of the criteria, may well be limited in this way,

and this limitation may account for the paucity of research findings concerning significant variations in the effects of practicum arrangements.

These problems may be avoided in GFE. The replacement standards for the goals of a program do not have to be reduced to narrow operationally defined objectives. The more vigorous search for outcomes of the program implicit in this model creates a more favourable climate for detecting effects due to any forces, including the hypothesized primitive tendencies of Stephens.

The justification specific to this study addresses the question of how programs with presumably different goals can be evaluated. The main issue focuses on determining standards for the evaluation.

The goals of programs cannot readily be used for determining these standards. Any selection made from those goals is subject to debate. At issue is the appropriateness of any such selection and the suitability of using goals from one program to evaluate another. The lack of consensus about teacher characteristics further contributes to this debate.

One way out of the morass might be to formulate and apply a goal-attainment coefficient. However, this hypothetical approach only further encumbers any attempt to use GBE for this study.

These objections to the use of a goal-based approach provide the rationale for using a goal-free approach in this evaluation. As a consequence, a single broadly-based external standard is applied to all programs.

Finally, as discussed in chapter one, there is an evident need to further develop and test GFE. It remains essentially an untried approach to evaluation which has potential in this area. As such, it merits further consideration and trials.

Boundary of the Evaluation

To use Scriven's terminology, two evaluation roles are possible -- formative and summative. These provide information at different points of a program or curriculum.

Formative evaluation occurs during the developmental stages. Its function is providing diagnostic information that assists in program or curriculum modification. Summative evaluation occurs at the end. Its function is providing judgemental information that determines merit.

The role of this evaluation is summative. The intent is to provide practica managers with goal-free information on program merit.

It is important to note that the ultimate product (effective practicum program) is the focus of the evaluation and not the intervening product (teaching effectiveness of pre-service teacher) even though the latter serves as the primary source of data. This

orientation towards the consumer of a program is an important aspect of goal-free evaluation.

The Components of the Evaluation Model

It became evident that a critical requirement of GFE is some suitable external standard replacing program goals. The importance of this standard is further reflected in its use in satisfying other requirements of GFE. Without this standard, the evaluation is open to very serious questions concerning its credibility and its conclusions.

The requirements for goal-free evaluation are not clearly outlined in any single literature source. It may, however, be inferred from Scriven and Evers (1980) that, the following components are required:

1. a suitable standard and process for determining merit
2. a mechanism to prevent or minimize the likelihood that the evaluator will be contaminated by knowledge of program goals
3. some way to assure managers that the evaluator(s) has strenuously avoided bias in the selection of the criteria for merit determination
4. a data collection process designed to enable the collection of the widest possible spectrum of outcomes
5. a focus on the needs of the consumer of the program

These five components were met by:

1. constructing a model of an effective teacher derived from the literature and from the evaluator's accumulated knowledge of teaching, and then by using this model as the basis for constructing instruments and analyzing outcomes
2. using a liaison person (evaluator's advisor) between the evaluator and the program managers and the co-operation of the managers in not divulging goal-related information
3. submitting the ETM to public scrutiny and modifying it accordingly
4. using instruments designed to be sensitive to as many as possible observable events in classrooms
5. collecting data pertaining to the perceived needs of pre-service teachers, and by interpreting other data from a perspective sensitive to needs

The Effective Teacher Model (ETM)

Role of the ETM

Reality is created by observers. Willer (1967), for example, said "We may measure in reality, but we do not measure reality" (p. 87). A model is implicit in any statement about reality, but it is therefore a conceptual filter. The problem is implicit in any evaluation of anything but in GFE it comes to the surface.

The problem is evaluator bias. The extent and nature of that bias dictates the outcomes that are captured and the acceptability of the conclusions that are presented. An evaluator can introduce bias at three different points of the evaluation; (a) at the formulation stage, (b) at the data collection stage and (c) at the analysis/conclusion stage.

The goal-replacement standard, in this case, the ETM, is the crucial element in minimizing evaluator bias. At the formulation stage, it provides some structure for designing the methodology. During data collection, it provides a basis for consistency of observation. During the third stage, it provides a standard for determining merit.

It cannot, however, be the sole window on the outcomes. Formal models and their derived instruments are not sensitive enough for capturing unknown, and hence, unspecifiable outcomes. This task is a vital component of GFE.

The Term 'Power' in the ETM

As mentioned in chapter one, the image of a teacher is drawn from an image of humanity. The prevalent icons or image determiners of this society imply certain values. In particular, individuals should obtain some power over their lives.

Paulo Freire reaffirms this perspective in another context; "Man's vocation is to be a subject who acts upon and transforms his world, and in so doing moves towards ever new possibilities of fuller and richer life individually and collectively." (Freire, 1974, p. 12-13). In order to act and transform, the individual requires the power to do so.

Education is perceived as a major vehicle for helping individuals attain such power, a power that is associated with status and ability. The educational system is perceived as set of institutions whose function is to educate and thereby to confer power. It thus becomes vital that the essential unit of the system, the teacher, be effective at enhancing the power position of a child, or, in Freire's terms, bestow upon children the means whereby to act and transform.

The teacher-to-be has power according to the extent he/she can function effectively in the classroom. In this study, 'power' consists of:

a) Cognitive aspects:

- teaching-related knowledge of participant
- teaching skills of participant
- decision-making ability

b) Affective aspects:

- attitude to self
- attitude to teaching

This power may be gained by acquiring pertinent skills, knowledge and attitudes during practica.

Terms and Variables of the ETM

Parents, administrators, teachers, students and citizens in general have varied perceptions of what these skills, knowledge and attitudes are. The difficulty lies in distilling their essential features. The literature and experience were major assists in this task.

Variables derived from omniscapable and futuristic concepts of teacher characteristics (Weil and Joyce, 1978) and subsumed in social and information processing models of teaching (Smitheran, Anderson, Hillis and Loucks, 1971), are grouped according to five terms:

- a) CLIMATE - the ease to which the classroom environment allows learning to occur
- b) CONTENT - the organized knowledge that is required by the teacher at the moment of instruction
- c) PURPOSE - the internalized reasons that constitute the basis for engaging and continuing in teaching

- d) MODE - the way the teaching act is performed
- e) IMAGE - the way the teacher is perceived by individuals outside the classroom

The variables (before the ETM was submitted to public scrutiny) associated with each term were:

a) CLIMATE:

1. Maintenance of classroom order
2. Maintenance of focus
3. maintenance of positive environment
4. Encouragement of pupil participation
5. Perceived honesty in relation to teaching

b) CONTENT:

1. Planning skills
2. Immediate subject matter competency

c) PURPOSE:

1. Liking children
2. Feeling positive about self
3. Feelin positive about task
4. Importance of teaching task
5. Reward associated with teaching task
6. Preparedness for teaching task

d) MODE:

1. Variability of lesson presentation
2. Suitability of teaching method
3. Organization of presentation
4. Clarity of presentation
5. Flexibility of presentation
6. Friendliness
7. Enthusiasm
8. Poise

e) IMAGE:

1. Appropriateness of appearance
2. Attitude towards superiors
3. Rapport with peers (in-service teachers)
4. Dependability
5. Overall perception of competence

The interplay of these terms provide the bonding for the model. The terms -- content, purpose, mode, climate and image -- are the determiners of power. The extent behaviours associated with these terms are observed or deduced determines the power gained by participants.

The Testing of the ETM

The initial ETM was submitted to public scrutiny as a way of determining whether its variables correspond to society's view of an effective teacher. This determination is useful for establishing external credibility as it addresses the problem of evaluator bias.

The testing of the ETM was accomplished by surveying members of the educational establishment, undergraduates and the public in general. Three variants (see appendix; H, I, J) of a questionnaire were used, consisting of items reflecting the variables of the ETM. The details are reported in chapters four and five.

As a consequence of this testing, it became necessary to remove one variable from the model -- the variable 'appropriateness of appearance' associated with the term 'image'. It did not receive enough support from those surveyed. This variable therefore was not used in instrument construction nor in merit determination.

Evaluation Instruments

Seven instruments (see appendix; A ... G) were constructed following a search of the literature. These instruments consist of; a) three classroom observations, b) three questionnaires and c) an interview.

Two of the classroom observation instruments are qualitative. By using a narrative design, they theoretically are sensitive to all outcomes. The third observation instrument is quantitative. It is designed to focus on a subset of variables, but it is not limited to this application. It is also open to other outcomes.

Two of the three questionnaires are targeted at pre-service teachers, with one being quantitative and the other being quantitative and qualitative. The third

questionnaire, targeted at co-operating teachers, is also quantitative and qualitative. The interview is with pre-service teachers.

Instrumentation Concerns

Whenever any instrument is used, concerns of validity and reliability are present; namely, the extent measures correspond to the nominal characteristics under investigation, and how reproducible the set of measures are.

In this study, it was not feasible to address these questions in a formal way. The questionnaires for testing the ETM and all evaluation instruments were developed without establishing formal validity or reliability. But it can be argued that the lack of formal attention to those concerns is not a serious impediment to the usefulness of the evaluation.

First, GFE requires the detection of unintended outcomes. This means that precise quantitative instruments have limited value since they depend on clear notions of what is being sought. The main instruments of GFE become observations and interviews. Such instruments do not readily admit of any formal measurement of validity.

Second, reliability refers to consistency of observation. Because GFE requires the evaluator being sensitive to all outcomes, reliability then implies a consistent sensitivity during observations, something that

is not readily testable. The best that can be hoped for is that the evaluator be aware of the need for consistent sensitivity.

Third, all of these observation instruments were field-tested. Besides providing an indication of usefulness for the evaluation, this field-testing probably addresses, at least subjectively, concerns for validity and reliability.

Fourth, the number [7] and the variety [4] of the evaluation instruments may help in reducing concerns of validity and reliability by providing different viewing portals of outcomes.

The questionnaires used in testing the ETM are the least defensible from this point of view. As they are quantitative instruments, it would have been possible to address these questions in a formal way. It was, however, not feasible because of time constraints.

External Credibility

There are two problems to be overcome in establishing the external credibility of the evaluation. One, evaluator bias, has been discussed above and has been accommodated by constructing and testing the ETM.

The second rises out of an inadequate control of the samples. Complete randomization was not possible as the populations of the practica were already established prior to the evaluation. While participants were randomly

selected from each of the practica for purposes of the evaluation, the participants were not randomly placed in the practica to begin with.

This condition may limit the generalization of conclusions concerning merit. The possibility exists that any differences of outcome may be attributable to differences of practica populations rather than to differences of practica experiences.

The methodology of this study attempted to overcome this limitation. An attempt was made to determine whether or not the practica populations were equivalent initially.

CHAPTER 5

THE METHODOLOGY OF THE STUDY

The evaluation consisted of three phases. In order of occurrence, these were; (1) the preparation for the evaluation, (2) the collection of evaluation data and (3) the analysis of the data.

Preparation for the Evaluation

There were five stages; (1) initial contact with practica, (2) construction of the ETM, (3) testing of the ETM, (4) selection of participants and (5) construction and field-testing of evaluation instruments.

Initial Contact with Practica

The evaluator's advisor provided preliminary information regarding practica. Subsequent contact was made with practica managers with information on time frames, locations and populations being obtained. No information concerning goals was solicited from managers or volunteered by them.

Construction of the ETM

The ETM was constructed in September/85. The terms and variables of this model are described in chapter four.

Testing of the ETM

The ETM was tested over a two-week period and provided information for modifying the model.

Data was collected for measuring the 'importance of a variable' by using a questionnaire employing a four point Likert Scale. This scale was considered sufficient for choice discrimination and forces a positive or negative choice as there is no neutral point on it. Respondents were given the opportunity to propose additional variables.

Three variants (see appendix; H, I, J) of the questionnaire were constructed and distributed to randomly selected members of targeted populations. These populations were; (a) undergraduate students in the Faculty of Education (Form A), (b) the educational establishment, represented by faculty members, principals and teachers (Form B) and (c) the public in general, represented by senior citizens, young adults and a mixed age group (Form D). The sample size for each population was 30.

A fourth population (pupils in the classroom) was considered originally but it was not feasible to include them in the study.

The 't' statistic was used for testing 'the importance of a variable' at the .025 level of significance. Details are provided in chapter six.

Selection of Participants

Four practica were evaluated; (a) Faculty-Based, (b) St. Vital Project (a subgroup of the Faculty-Based), (c) School-Based and (d) Redekopp group at Shamrock School (a subgroup of the School-Based). All participants were fourth year students at the Faculty of Education.

For each practica, sufficient host schools were randomly selected to provide eight participants. Participants were then randomly selected from these schools. A total of 32 participants became subjects of this study.

Construction and Field-Testing of Instruments

Evaluation instruments were constructed as described in chapters three and four. They are provided in appendix A through G.

The three observation instruments were field-tested by using them to observe three in-service teachers. Information was obtained from these concerning the correspondence between evaluator-recorded observations and teacher-perceived events. This field-testing served to improve recording procedures and consistency of observations. The information obtained at this stage led to minor modification of the instruments.

Collection of Evaluation Data

Evaluation data was collected during the Fall and Spring practica. There were three stages of data collection; (1) early in the practica, (2) in the middle of the practica and (3) at the conclusion of the practica.

The order of observations regarding specific programs and participants was randomly determined within limits imposed by feasibility. No participants were lost from the study.

In the first stage, data was collected by using a narrative instrument to record classroom events. This was followed by a questionnaire constructed by using a three point Likert Scale and intended for gathering data about participants' needs.

Participants were observed once near the start of the fall practicum. Participants were not informed of the time of observations. All participants agreed to this method of observation.

The second stage used a quantitative instrument for recording frequency data for five variables. Participants were observed once near the start of the spring practicum. Participants were again not informed of the time of observations.

Four instruments were used for the third stage of data collection. These are, in order of application; (1) a questionnaire for co-operating teachers, (2) a narrative

observation, (3) a questionnaire for participants, and (4) an interview with participants.

This time, participants were informed of the time of observation so as to reduce problems of scheduling.

Analysis of Evaluation Data

No data was analyzed until after all the data had been collected. This delaying of analysis was intentional so as to maintain consistency of observation. Ignorance of results was intended to promote consistent open viewing, with no 'flagging' of information.

The analysis employed a range of qualitative and quantitative methods. GFE legitimizes this approach. By using varied methods of analysis, the exploration of GFE may be enhanced.

Equivalence of Practica Populations

The practica populations were established prior to the evaluation. The participants chose their programs. It therefore cannot be assumed that the populations are equivalent; a fact that hinders the usefulness of any conclusions. Testing for equivalence served to provide a measure of initial conditions (pre), thereby facilitating pre-post comparisons and hence any conclusions concerning gains and differences in gains.

Two dimensions of this equivalence were considered; (1) participant's needs and (2) participant's teaching expertise. The CHI-SQUARED statistic was used for testing equivalence. Observed frequency was compared with an estimated expected frequency at the .025 level of significance.

The retention of the omnibus null hypothesis would indicate that the practica populations were equivalent on at least this measure derived from the ETM. For purposes of this evaluation, it could then be concluded that the practica populations were equivalent. Details are provided in chapter six.

Analysis from a Program Perspective

The desired outcome from a program perspective is an effective teacher. The two dimensions that were examined were; (1) gain in effectiveness and (2) differences in gain. The corresponding questions were; (1) Does teaching effectiveness improve as a result of practica? and (2) Are there any differences between practica concerning this outcome?

A qualitative approach was used to address the first question. A subjective judgement of gain was made based on all the collected data.

A quantitative approach was employed to address the second question. For each term of the ETM, a single factor ANOVA was used for comparing final teaching effectiveness

(post) of practica participants. Ratings of participants were used for the ANOVA's.

By considering these results together with results from the testing for equivalence, conclusions concerning any differences in gains between practica were possible. This approach, while not technically appropriate (MANOVA needed), is satisfactory, in this case, because of the exploratory nature of study.

The omnibus null hypothesis was tested at the .01 level of significance. The retention of the null hypothesis would indicate no differences in final effectiveness. If the results of the F-tests showed any differences, a t-test(s) would be employed for identifying specific program(s). The details are provided in chapter six.

Analysis from a Pre-service Teacher Perspective

The needs of pre-service teachers were considered from two viewpoints; (1) self-expressed needs and (2) the implicit need to improve teaching expertise. The latter also was a program concern and has already been examined from that perspective by employing ratings. A different approach is employed in this section.

Two dimensions of self-expressed needs were examined; (1) meeting of needs and (2) differences between the practica. Quantitative approaches were employed for the analysis.

Descriptive statistics were used to examine the first dimension by tabulating the frequency of 'satisfied with' that were obtained from the pre and post self-reports of needs (see appendix; B, F). No attempt was made to test for statistical significance of any changes in frequency as conflicting interpretations may be made concerning these.

The second dimension was examined by using the CHI-SQUARED statistic for comparing the observed frequency of responses on a three point Likert Scale with the estimated expected frequency. The observed frequencies were obtained from a post self-report of needs.

The omnibus null hypothesis was tested at the .025 level of significance. Retention of the null hypothesis would indicate that there were no differences in meeting needs. The details are provided in chapter six.

Two dimensions of the implicit need to improve expertise were examined; (1) improvement in expertise and (2) differences in improvement. Quantitative methods were employed for the analysis.

The first dimension was examined with descriptive statistics by tabulating the frequency of negative events. These frequencies were derived from observations one and three (see appendix; A, D).

The second dimension was examined by using the CHI-SQUARED statistic for comparing the observed frequency of negative events with the estimated expected frequency.

The observed frequencies were derived from observation three. By considering these results together with results from the testing for equivalence, conclusions concerning any differences in improvement between practica were possible.

The omnibus null hypothesis was tested at the .025 level of significance. Retention of the null hypothesis would indicate that there were no differences in final expertise. Details are provided in chapter six.

Analysis of Other Outcomes

GFE implies capturing all outcomes, even those not accounted for by the replacement standard for program goals; in this study, the ETM.

Locating and analyzing these outcomes was accomplished by examining the data from a perspective outside of the ETM. The results are provided in appendix K.

CHAPTER 6

RESULTS AND CONCLUSIONS

Overview

The results and conclusions fall into eight sections; (1) the testing of the ETM, (2) the equivalence of the practica populations, (3) the programs' perspective, (4) the participants' perspective, (5) the perspective from outside the ETM, (6) the effects concerning study participation, (7) the usefulness of GFE and (8) the overall conclusions and recommendations. These are provided in the above order. Each section provides any required details of analysis and results, the pertinent tables and conclusions.

Section 1: Testing the ETM

The Effective Teacher Model was modified slightly after the data from the three questionnaires was analyzed. The percentages of returned questionnaires were; 100% for the public in general, 77% for the educational establishment and 53% for the faculty undergraduates.

The following tables (1, 2, 3) provide the raw data concerning responses to questionnaire items and show the response frequencies for the four Likert Scale categories. These categories were; Strongly Disagree, Disagree, Agree and Strongly Agree and they are signified respectively by; SD, D, A, and SA. An '*' indicates that the item representing the variable was not included in that variant of the questionnaire.

Each questionnaire item represented one variable of the ETM. Henceforth, 'variable of the ETM' and 'item of the questionnaire' are equivalent.

TABLE 1: Undergraduate Students (Form A)
Response Frequency per Likert Scale Category

	SD	D	A	SA	Total
POWER:					
knowledge	0	0	12	4	16
skills	0	1	11	4	16
+ self-image	0	0	9	7	16
+ learning att.	1	1	5	9	16
decision making	0	3	4	9	16
CLIMATE:					
order	0	1	6	9	16
focus	0	3	8	5	16
+ learning envr.	0	0	2	14	16
pupil particip.	0	2	3	11	16
honest in task	1	1	3	10	15
CONTENT:					
planning skills	0	1	4	10	15
subject matter	1	0	5	9	15
PURPOSE:					
likes children	0	2	7	7	16
self-image	0	0	5	11	16
+ to task	0	1	5	10	16
task important	1	0	6	9	16
task rewarding	1	0	8	7	16
ready for task	1	1	6	8	16
MODE:					
variability	1	0	1	13	15
suitability	0	0	2	13	15
organization	1	1	4	10	16
clarity	0	1	3	12	16
flexibility	0	2	5	9	16
friendliness	0	3	10	3	16
enthusiasm	0	0	4	12	16
poise	0	6	7	3	16
IMAGE:					
appearance	0	4	11	1	16
attitude-super.	0	1	9	6	16
peer rapport	0	0	8	8	16
dependability	0	0	7	9	16
competence	0	1	7	8	16

TABLE 2: Educational Establishment (Form B)
Response Frequency per Likert Scale Category

variable	SD	D	A	SA	Total
POWER:					
knowledge	0	1	17	5	23
skills	0	0	15	7	22
+ self-image	1	2	11	8	22
+ learning att.	0	1	14	8	23
decision making	0	0	12	11	23
CLIMATE:					
order	0	2	13	8	23
focus	0	2	14	7	23
+ learning envr.	0	0	3	20	23
pupil particip.	0	0	9	13	22
honest in task	0	0	5	15	20
CONTENT:					
planning skills	0	1	11	8	20
subject matter	0	1	7	12	20
PURPOSE:					
likes children	0	1	11	10	22
self-image	0	0	12	10	22
+ to task	0	1	11	10	22
task important	0	4	7	11	22
task rewarding	0	2	12	8	22
ready for task	0	0	9	13	22
MODE:					
variability	0	0	10	10	20
suitability	0	0	5	15	20
organization	0	1	11	10	22
clarity	0	1	8	13	22
flexibility	0	1	7	14	22
friendliness	0	2	8	11	21
enthusiasm	0	0	9	13	22
poise	0	1	15	6	22
IMAGE:					
appearance	2	8	11	0	21
attitude-super.	0	3	17	2	22
peer rapport	0	1	16	6	23
dependability	0	2	14	7	23
competence	0	1	15	7	23

TABLE 3: Public in General (Form D)
Response Frequency per Likert Scale Category

variable	SD	D	A	SA	Total
POWER:					
knowledge	5	3	17	5	30
skills	2	4	18	6	30
+ self-image	0	2	10	18	30
+ learning att.	0	1	12	17	30
decision making	0	1	18	11	30
CLIMATE:					
order	1	4	11	14	30
focus	2	12	10	6	30
+ learning envr.	0	0	10	20	30
pupil particip.	0	2	9	19	30
honest in task	0	2	11	17	30
CONTENT:					
planning skills	0	1	11	18	30
subject matter	1	1	13	15	30
PURPOSE:					
likes children	2	6	15	6	29
MODE:					
variability	0	2	19	9	30
suitability	0	4	20	6	30
organization	0	3	16	11	30
clarity	0	0	11	19	30
flexibility	0	8	17	5	30
friendliness	0	3	17	10	30
enthusiasm	0	2	11	17	30
poise	0	1	14	14	29

This information was tested by employing a one-tailed 't' test on the response mean at the .025 level of significance. For this purpose, the four categories of the Likert Scale (SD, D, A, SA) were respectively assigned the numerical equivalents; 1, 2, 3, and 4.

The null hypothesis was that the response mean for the variable is less than or equal to 2.5 (midpoint of the Likert Scale). The alternative hypothesis was that the response mean for the variable is greater than 2.5.

The rejection of the null hypothesis was taken as evidence that the surveyed population holds that the variable is important and that it should remain in the ETM.

It was possible for a variable to be acceptable to some but not all populations. This did occur. It was accommodated by treating the data for that variable as originating from a larger population by combining the samples. The same 't' test and hypotheses were used for determining whether the variable remains in the ETM.

Table 4 shows the response mean and standard deviation for the variables.

TABLE 4: Response Mean and Standard Deviation (S)
For each Variable of the ETM

POPULATION	Educat. Establ.		Undergrads		General Populace	
	mean	S	mean	S	mean	S
POWER:						
knowledge	3.17	.49	3.25	.45	2.73	.94
skills	3.32	.48	3.19	.54	2.93	.78
+ self-image	3.18	.80	3.44	.51	3.53	.63
+ learning att.	3.30	.56	3.38	.89	3.53	.57
decision making	3.47	.51	3.38	.81	3.33	.55
CLIMATE:						
order	3.26	.62	3.50	.63	3.27	.83
focus	3.22	.60	3.13	.72	2.67	.88
+ learning envr.	3.87	.34	3.88	.34	3.67	.48
pupil particip.	3.59	.50	3.56	.73	3.57	.63
honest in task	3.75	.44	3.47	.92	3.50	.63
CONTENT						
planning skills	3.35	.59	3.60	.63	3.57	.57
subject matter	3.55	.60	3.47	.83	3.40	.72
PURPOSE:						
likes children	3.41	.59	3.31	.70	2.86	.83
self-image	3.45	.51	3.69	.48	*	*
+ to task	3.41	.59	3.56	.63	*	*
task important	3.32	.78	3.44	.81	*	*
task rewarding	3.27	.63	3.31	.79	*	*
ready for task	3.59	.50	3.31	.86	*	*
MODE:						
variability	3.50	.51	3.73	.79	3.23	.57
suitability	3.75	.44	3.87	.35	3.07	.58
organization	3.41	.59	3.44	.88	3.27	.64
clarity	3.55	.60	3.69	.60	3.63	.49
flexibility	3.59	.59	3.44	.72	2.90	.66
friendliness	3.43	.68	3.00	.63	3.23	.63
enthusiasm	3.59	.50	3.75	.45	3.50	.63
poise	3.22	.53	2.81	.75	3.45	.57
IMAGE:						
appearance	2.43	.68	2.81	.54	*	*
attitude-super.	2.95	.49	3.31	.60	*	*
peer rapport	3.22	.52	3.50	.52	*	*
dependability	3.22	.60	3.56	.52	*	*
competence	3.26	.54	3.44	.63	*	*

The null hypothesis was not rejected for the following variables and corresponding populations:

1. POWER - knowledge: public in general
2. CLIMATE - focus: public in general
3. IMAGE - appearance: undergraduates
4. MODE - poise: undergraduates

These results implied that the particular population did not think that the variable was important enough to include in the ETM.

These four variables were tested again by combining the populations. The results were:

1. POWER - knowledge: mean = 3.00 S = .75
2. CLIMATE - focus: mean = 2.96 S = .79
3. IMAGE - appearance: mean = 2.53 S = .76
4. MODE - poise: mean = 3.22 S = .65

The null hypothesis was not rejected only for the variable, 'appearance', and it therefore was removed from the ETM. The other three variables were retained in the ETM. No variables were added as few respondents included information useful for this purpose.

These results indicated that there was strong agreement with the ETM. Evidently, sufficient respondents agreed or strongly agreed with statements concerning the importance of the proposed variables of the ETM. It therefore could be used as a standard.

This meant that criteria for effectiveness of teaching could be derived from it and used for the evaluation.

Some summary information was also determined. The grand mean across all variables for each population was; (a) undergraduates 3.42, (b) educational establishment 3.37 and (c) general population 3.27. These values indicate that the typical response to the items of the questionnaire was somewhere between 'Agree' and 'Strongly Agree'.

Section 2: Equivalence of Practica Populations

The first dimension of this equivalence, participants' needs, was examined by analyzing needs data that was obtained from a self-report of needs (see appendix B) that was applied early in the practicum. The instrument employed a three point Likert Scale. Four terms - CLIMATE, MODE, CONTENT AND PURPOSE - were considered.

This data was organized into frequency tables and the Karl Pearson Chi-Square test of homogeneity of distributions (Marascuilo and McSweeney, 1977) was employed to analyze it.

The same procedure was used for all four terms. The observed frequency for a cell of the table was determined by summing the responses for a scale category for all the variables associated with the term. The estimated expected frequency was determined by multiplying the relative frequency for a scale category by the program sample size.

The null hypothesis was that the practica have the same frequency distribution of responses. The alternative hypothesis was that the omnibus null hypothesis was false. The rejection of the null hypothesis would indicate that practica populations were not equivalent on needs.

The test was conducted at the .025 level of significance with the critical value of CHI-SQUARED being 14.45 for six degrees of freedom.

Tables 5 through 12 show the observed and the estimated expected frequencies, and the CHI-SQUARED values.

The four practica are denoted by:

FB - the Faculty-Based program

SV - the St. Vital Project

SB - the School-Based program

RD - the Redekopp group at Shamrock School

L.S. denotes the Likert Scale category with '1' signifying 'satisfied with', '2' signifying 'some improvement needed' and '3' signifying 'much improvement needed'.

Term: CLIMATE (5 variables)

TABLE 5: Observed and Relative Frequency per Category of the Likert Scale for Practica

L.S.	FB	SV	RD	SB	total	Relative Fq.
1	25	22	24	19	90	$90/160 = .563$
2	14	17	13	18	62	$62/160 = .388$
3	1	1	3	3	8	$8/160 = .050$
Total	40	40	40	40	160	1.001

TABLE 6: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L.S.	FB	SV	RD	SB	Total CHI-SQ
1	22.52	22.52	22.52	22.52	
2	15.52	15.52	15.52	15.52	
3	2.00	2.00	2.00	2.00	
CHI-S	.922	.653	1.006	1.446	4.03

The null hypothesis was not rejected.

Term: CONTENT (2 variables)

TABLE 7: Observed and Relative Frequency per Category of the Likert Scale for Practica

L.S.	FB	SV	RD	SB	total	Relative Fq.
1	5	7	6	5	23	$23/64 = .359$
2	11	7	8	11	37	$37/64 = .578$
3	0	2	2	0	4	$4/64 = .063$
Total	16	16	16	16	64	1.000

TABLE 8: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L.S.	FB	SV	RD	SB	Total CHI-SQ
1	5.74	5.74	5.74	5.74	
2	9.25	9.25	9.25	9.25	
3	1.01	1.01	1.01	1.01	
CHI-S	1.436	1.794	1.151	1.436	5.82

The null hypothesis was not rejected.

Term: PURPOSE (5 variables)

TABLE 9: Observed and Relative Frequency per Category of the Likert Scale for Practica

L.S.	FB	SV	RD	SB	total	Relative Fq.
1	28	29	32	28	117	117/160= .731
2	12	11	8	12	43	43/160= .269
3	0	0	0	0	0	0/160= .000
Total	40	40	40	40	160	1.000

TABLE 10: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L.S.	FB	SV	RD	SB	Total CHI-SQ
1	29.24	29.24	29.24	29.24	
2	10.76	10.76	10.76	10.76	
3	0	0	0	0	
CHI-S	.196	.007	.969	.193	1.37

With an expected frequency of 0, category 3 was not amenable to treatment with CHI-SQUARED, and therefore this category was eliminated. The critical value of CHI-SQUARED is 9.35 for three degrees of freedom.

The null hypothesis was not rejected.

Term: MODE (8 variables)

TABLE 11: Observed and Relative Frequency per Category of the Likert Scale for Practica

L. S.	FB	SV	RD	SB	total	Relative Fq.
1	28	24	28	23	103	$103/256 = .402$
2	33	36	32	34	135	$135/256 = .527$
3	3	4	4	7	18	$18/256 = .070$
Total	64	64	64	64	256	.999

TABLE 12: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L. S.	FB	SV	RD	SB	Total CHI-SQ
1	25.73	25.73	25.73	25.73	
2	33.33	33.33	33.33	33.33	
3	4.48	4.48	4.48	4.48	
CHI-S	.705	.320	.340	1.710	3.08

The null hypothesis was not rejected.

For this dimension, participants' needs, the null hypothesis was not rejected for any term. It was therefore concluded that the needs of the practica populations were equivalent prior to the evaluation.

The second dimension, participants' teaching expertise, was examined by comparing the observed frequency of 'negative events' with the expected frequency. A negative event was defined as a participant's action or lack of action that could be interpreted as being negative from a perspective based on the ETM.

There are limitations in this approach to quantification. The events may be weighted differently according to severity of impact -- impossible to measure. The data was not rich enough to encompass all the terms of the ETM. Only two terms (MODE and CLIMATE) were considered.

Data on teaching expertise was obtained from an observation (see appendix A) that was applied early in the practicum. The instrument employs a narrative recording of classroom events.

For each term, the observed and expected frequency were determined. The observed frequency was the sum of the occurrences of negative events for the variables associated with a term. The expected frequency was the mean of the observed frequencies across the practica for a given term.

The null hypothesis was that the practica have the same frequency of negative events. The alternative hypothesis was that the omnibus null hypothesis was false. The rejection of the null hypothesis would indicate non-equivalence of teaching expertise.

The critical value of CHI-SQUARED is 9.35 for three degrees of freedom at the .025 level of significance.

Tables 13 and 14 show the observed and the expected frequencies, and the CHI-SQUARED values.

Term: CLIMATE

TABLE 13: Observed and Expected Frequency of Negative Events and CHI-SQUARED for Practica

	FB	SV	SB	RD	Total	Expect.
obs.	10	21	38	34	103	25.8
CHI-S	9.68	.89	5.77	2.61	18.95	

The null hypothesis was rejected. Cochran's theorem (Marascuilo & McSweeney, 1977) for partitioning cannot be used for identifying the specific program(s) causing the indicated statistical significance. However, from the table, it was evident that the FB practicum contributed most to the CHI-SQUARED statistic and that it had the lowest occurrence of negative events.

Term: MODE

TABLE 14: Observed and Expected Frequency of Negative Events and CHI-SQUARED for Practica

	FB	SV	SB	RD	Total	Expect.
obs.	19	19	23	29	90	22.5
CHI-S	.544	.544	.011	1.88	2.98	

The null hypothesis was not rejected.

For the second dimension, participants' teaching expertise, the null hypothesis was rejected for the term CLIMATE. It therefore could be concluded that the populations were not equivalent for this term. This implied that the practica populations were of differing ability concerning the ease to which the classroom environment allowed learning to occur.

While it was not statistically possible to isolate the specific program(s) in this case, a tentative conclusion was formed. From the data, it could be inferred that the Faculty-Based group had the most teaching expertise with respect to the term CLIMATE prior to the practicum.

Section 3: The Programs' Perspective

The first dimension, gain in teaching expertise, was examined qualitatively. Based on subjective judgement, it was evident that there was a marked change from the beginning to the end of practica. Gains in teaching expertise occurred for the participants.

The atmosphere established in the classroom, as measured by maintenance of order and focus and by encouragement and involvement of students, improved for all participants. The quality of lesson presentation, as measured by variability, suitability, organization, clarity, flexibility, enthusiasm, friendliness and poise, improved for all participants. The inner drive or motivation, as measured by self-confidence and attitude to teaching,

increased for most participants. The knowledge base, as measured by planning skills and subject matter competency increased for all participants.

The conclusion is that the practica increased the POWER of participants. That is to say, they acquired knowledge and skills, became more positive towards self and teaching, and improved their decision-making. These gains improved their teaching effectiveness.

The second dimension, difference of gains, was examined by applying an ANOVA to each of five terms and interpreting these results (post) in conjunction with results obtained from the testing for equivalence (pre). The terms were; CLIMATE, CONTENT, PURPOSE, MODE and IMAGE.

The ANOVA's tested for final differences between practica. The tested variable was 'participant rating', derived from the final observation with all of its parts (see appendix; D, E, F, G).

Participants were rated on a variable of the ETM by using a seven point scale with '1' signifying 'low'. The mean of these ratings was used as the participant's rating for a term.

The null hypothesis was that the practica have equal ratings. The alternative hypothesis was that the omnibus null hypothesis was false. The rejection of the null hypothesis would indicate that there was a final difference in ratings for the term examined.

The null hypothesis was tested at the .01 level of significance. The critical F-value is 4.57 for three degrees of freedom in the numerator and 28 in the denominator.

Tables 15, 16, 17, 18 and 19 show the participants' ratings, and mean program ratings and variances.

TABLE 15: Ratings of Practica Participants for CLIMATE

Program	FB	SV	SB	RD
rating n = 8	6.2	5.5	6.4	6.2
	6.6	4.7	5.2	5.4
	5.2	5.7	5.8	5.2
	6.4	4.3	4.4	5.0
	5.8	4.4	4.6	6.2
	3.8	5.1	5.0	6.4
	3.8	5.7	5.4	5.0
	4.6	4.7	4.2	6.2
mean rating	5.3	5.0	5.1	5.7
variance	1.28	.31	.55	.37

The calculated F-value was 1.15. The null hypothesis was not rejected.

TABLE 16: Ratings of Practica Participants for CONTENT

Program	FB	SV	SB	RD
rating n = 8	5.5	5.1	5.0	5.5
	5.0	4.6	5.5	5.5
	5.5	5.6	4.5	5.0
	5.0	5.1	5.5	5.0
	5.5	4.1	5.5	5.0
	5.0	5.1	5.5	5.0
	4.0	5.1	5.0	5.5
	4.0	5.1	4.5	4.5
mean rating	4.9	5.0	5.1	5.1
variance	.39	.20	.20	.13

The calculated F-value was .315. The null hypothesis was not rejected.

TABLE 17: Ratings of Practica Participants for PURPOSE

Program	FB	SV	SB	RD
rating n = 8	5.8	5.9	5.3	6.3
	6.0	4.8	4.8	5.5
	5.0	4.6	6.2	6.0
	6.2	5.9	6.0	5.8
	5.2	4.8	4.8	6.3
	4.5	5.1	5.2	5.8
	5.7	5.4	4.6	5.2
	5.5	5.3	5.8	5.8
mean rating	5.5	5.2	5.3	5.8
variance	.31	.25	.36	.14

The calculated F-value was 2.15. The null hypothesis was not rejected.

TABLE 18: Ratings of Practica Participants for MODE

Program	FB	SV	SB	RD
rating n = 8	5.9	5.0	6.0	5.9
	6.0	4.1	4.6	5.1
	4.6	6.0	5.5	4.1
	6.5	4.4	4.6	5.3
	5.1	4.7	4.4	6.3
	3.6	5.1	5.3	6.1
	4.1	5.0	4.7	5.5
	4.9	4.1	4.0	6.0
mean rating	5.1	4.8	4.9	5.6
variance	.99	.39	.43	.33

The calculated F-value was 1.91. The null hypothesis was not rejected.

TABLE 19: Ratings of Practica Participants for IMAGE

Program	FB	SV	SB	RD
rating n = 8	6.0	4.0	6.5	5.8
	7.0	5.0	5.3	5.8
	5.8	4.8	6.8	5.8
	5.8	4.5	5.3	5.0
	5.8	5.5	5.5	6.8
	4.0	5.0	4.5	4.8
	5.8	5.0	5.8	4.5
	6.0	5.8	5.5	6.8
mean rating	5.8	5.0	5.7	5.7
variance	.68	.31	.53	.74

The calculated F-value was 2.02. The null hypothesis was not rejected.

While there are no statistically significant results in this cluster of measures, some patterns of results warrant their identification for further study. For the terms -- CLIMATE, PURPOSE and MODE -- the FB, SV and SB means are closely grouped and the RD mean seems higher. This observation may have implications for further study. It might be productive to study in greater detail what it is about the RD protocol that accounts for this possible superiority.

For the term -- IMAGE -- the FB, SB and RD means are closely grouped and the SV mean seems lower. This observation may have implications for orientation sessions with the SV co-operating teachers and student teachers. They may be confusing ideals with perceived performances.

The ANOVAs did not reveal any final differences. The FB group initially was superior in teaching expertise for the term -- CLIMATE. This advantage was not maintained. It therefore could be concluded, assuming linearity, that the Faculty-Based practicum was not as effective in developing the skills pertinent to this term, those related to creating an environment which facilitated learning.

The Redekopp group had a higher rating than the rest for the terms -- CLIMATE, PURPOSE and MODE. This provides some indication that this practicum may have a more positive influence on those aspects of teaching related to; (a) creating an environment which facilitated learning, (b) fostering positive attitudes in teachers and (c) improving lesson presentation.

The St. Vital group had a lower rating than the rest for the term -- IMAGE. This provides some indication that the teachers from the host schools were more critical and perhaps had higher expectations of the participants.

Section 4: The Participants' Perspective

Viewpoint 1: Self-expressed needs

The first dimension, meeting of needs, was examined by using frequency information. The frequency of responses for the category 'satisfied with' were summed across the variables associated with the term for each practicum. The final value was obtained by summing these practicum frequencies. This was done for the terms; CLIMATE, CONTENT, PURPOSE and MODE. These frequencies were obtained from the pre and post instruments.

Table 20 shows the total pre and post response frequencies for the category, satisfied with, and the changes.

TABLE 20: Total Frequency of Satisfied With for Practica

Term	PRE	POST	CHANGE	
CLIMATE	90	116	+ 26	29%
CONTENT	23	40	+ 17	74%
PURPOSE	117	130	+ 13	11%
MODE	103	160	+ 57	55%

CONTENT was affected most by practica but there was an increase in the frequency of satisfied with for all terms.

Based on the increase in the number of participants expressing satisfaction with their teaching expertise, it could be concluded that gains occurred for participants concerning their perceived needs throughout the practica. Practica seem to provide positive experiences which result in an increased confidence concerning their teaching skills.

The second dimension, differences between practica, was examined by analyzing needs data that was obtained from a self-report of needs (see appendix B) that was applied towards the end of practica. The instrument employed a three point Likert Scale. Four terms - CLIMATE, MODE, CONTENT AND PURPOSE - were considered.

This data was organized into frequency tables and the Karl Pearson Chi-Square test of homogeneity of distributions (Marascuilo and McSweeney, 1977) was employed to analyze it.

The same procedure was used for all four terms. The observed frequency for a cell of the table was determined by summing the responses for a scale category for all the variables associated with the term. The estimated expected frequency was determined by multiplying the relative frequency for a scale category by the program sample size.

The null hypothesis was that the practica have the same frequency distribution of responses. The alternative hypothesis was that the omnibus null hypothesis was false.

The rejection of the null hypothesis would indicate that differences existed between practica concerning meeting of needs.

The test was conducted at the .025 level of significance with the critical value of CHI-SQUARED being 14.45 for six degrees of freedom.

Tables 21 through 28 show the observed and estimated expected frequencies, and the CHI-SQUARED values.

L.S. denotes the Likert Scale category with '1' signifying 'satisfied with', '2' signifying 'some improvement needed' and '3' signifying 'much improvement needed'.

Term: CLIMATE (5 variables)

TABLE 21: Observed and Relative Frequency per Category of the Likert Scale for Practica

L.S.	FB	SV	RD	SB	total	Relative Fq.
1	29	24	32	31	116	$116/160 = .725$
2	10	15	7	9	41	$41/160 = .256$
3	1	1	1	0	3	$3/160 = .019$
Total	40	40	40	40	160	1.000

TABLE 22: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L.S.	FB	SV	RD	SB	Total CHI-SQ
1	29.00	29.00	29.00	29.00	
2	10.24	10.24	10.24	10.24	
3	.76	.76	.76	.76	
CHI-S	.082	3.151	1.411	1.048	5.69

The null hypothesis was not rejected.

Term: CONTENT (2 variables)

TABLE 23: Observed and Relative Frequency per Category of the Likert Scale for Practica

L.S.	FB	SV	RD	SB	total	Relative Fq.
1	12	9	11	8	40	$40/64 = .625$
2	4	6	5	8	23	$23/64 = .359$
3	0	1	0	0	1	$1/64 = .016$
Total	16	16	16	16	64	1.000

TABLE 24: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L.S.	FB	SV	RD	SB	Total CHI-SQ
1	10.00	10.00	10.00	10.00	
2	5.74	5.74	5.74	5.74	
3	.26	.26	.26	.26	
CHI-S	1.187	2.218	.455	1.550	5.41

The null hypothesis was not rejected.

Term: PURPOSE (5 variables)

TABLE 25: Observed and Relative Frequency per Category of the Likert Scale for Practica

L.S.	FB	SV	RD	SB	total	Relative Fq.
1	30	29	35	36	130	130/160 = .813
2	10	11	4	4	29	29/160 = .181
3	0	0	1	0	1	1/160 = .006
Total	40	40	40	40	160	1.000

TABLE 26: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L.S.	FB	SV	RD	SB	Total CHI-SQ
1	32.52	32.52	32.52	32.52	
2	7.24	7.24	7.24	7.24	
3	.24	.24	.24	.24	
CHI-S	1.487	2.573	4.049	2.062	10.17

The null hypothesis was not rejected.

Term: MODE (8 variables)

TABLE 27: Observed and Relative Frequency per Category of the Likert Scale for Practica

L.S.	FB	SV	RD	SB	total	Relative Fq.
1	42	30	51	37	160	160/256 = .625
2	20	31	13	27	91	91/256 = .355
3	2	3	0	0	5	5/256 = .020
Total	64	64	64	64	256	1.000

TABLE 28: Estimated Expected Frequency per Category of the Likert Scale and CHI-SQUARED for Practica

L.S.	FB	SV	RD	SB	Total CHI-SQ
1	40	40	40	40	
2	22.72	22.72	22.72	22.72	
3	1.28	1.28	1.28	1.28	
CHI-S	.831	7.829	8.463	2.311	19.43

The null hypothesis was rejected.

The null hypotheses were retained for the terms; CLIMATE, CONTENT and PURPOSE. There is therefore no evidence of differences between the practica concerning the meeting of participants' needs for these terms.

The null hypothesis was rejected for the term MODE. It is concluded that a difference exists here. A post hoc multiple comparison technique for the I x K test for homogeneity (Marascuilo & McSweeney, 1977) was used to identify the specific program(s) causing the rejection of the null hypothesis. The level of significance used for these comparisons was .025. The corresponding critical value is the square root of 14.45 (3.801).

The following contrasts (differences) were tested. The results are;

1) SV - SB:	1.23
2) SV - FB:	2.17
3) SV - RD:	4.1
4) RD - FB:	2.7
5) RD - FB:	3.31
6) SV - (FB + SB + RD):	2.9
7) RD - (SV + SB + FB):	3.74

The only statistically significant finding is that there is a significant difference between the SV (St. Vital) and the RD (Redekopp) programs. It is concluded that, for the aspect of teaching related to presenting of lessons, the

Redekopp practicum seems to meet the needs of participants more than the St. Vital practicum.

Viewpoint 2: The implicit need to improve expertise

The first dimension, change in expertise, was examined by comparing the frequency of negative events from pre and post observations. The definition for negative event was as before.

For each term, the observed frequency was the sum of all occurrences of negative events for the variables associated with it. The terms examined were CLIMATE and MODE.

Table 29 shows the number of negative events occurring during the pre and post observations and the changes.

TABLE 29: Pre and Post Frequency of Negative Events for Practica

Term	OB 1	OB 3	CHANGE	
CLIMATE	103	37	- 66	64%
MODE	90	52	- 38	42%

Based on this reduction of negative events, it may be concluded that gains occurred throughout all practica. Practica cause an improvement in teaching expertise. The implicit need of participants to improve their expertise therefore was met.

This conclusion is consistent with that obtained in the qualitative approach employed when considering improvement in expertise from the program perspective.

The second dimension, differences of improvement between practica, was examined by applying the CHI-SQUARED statistic to frequency information, and interpreting these results (post) in conjunction with results from the testing for equivalence (pre). CHI-SQUARED was used for comparing observed and estimated expected frequencies of negative events. The definition of negative event was as before.

The observed frequencies were derived from the post observation. Only two terms (MODE, CLIMATE) were considered. For each practicum, the observed frequency was obtained by summing the occurrences of negative events for the variables associated with the term. The estimated expected frequency was the mean of the observed frequencies across the practica.

The null hypothesis was that the practica have an equal frequency distribution of negative events. The alternative hypothesis was that the omnibus null hypothesis was false.

The rejection of the null hypothesis would indicate that there were final differences between practica.

The critical value of CHI-SQUARED is 9.35 for three degrees of freedom at the .025 level of significance.

Tables 30 and 31 show the observed and expected frequencies, and CHI-SQUARED.

TABLE 30: Observed and Expected Frequency of Negative Events for Practica for CLIMATE

FB	SV	SB	RD	mean	CHI-SQUARED
7	10	13	7	9.25	2.68

The null hypothesis was not rejected.

TABLE 31: Observed and Expected Frequency of Negative Events for Practica for MODE

FB	SV	SB	RD	mean	CHI-SQUARED
18	11	15	8	13.0	4.46

The null hypothesis was not rejected for any term, providing no evidence of final differences.

The FB group initially was superior in teaching expertise for the term -- CLIMATE. This advantage was not maintained. It therefore could be concluded, assuming linearity, that the Faculty-Based practicum is not as

effective in meeting the implicit need to improve expertise for those skills related to creating an environment which facilitated learning.

It could equally be argued that all of these programs aim towards a fixed paradigm, and tend to bring students to it regardless of their initial skills and perceptions.

Section 5: The Perspective from outside the ETM

The observations and conclusions are provided in appendix K.

Section 6: Effects of Participation in the Study

There appeared to be little interference with or influence on participants related to their participation in the study. Only three indicated any.

One participant stated that the presence of the evaluator tended to force more awareness of classroom events. The other two indicated some nervousness resulting from the presence of the evaluator.

Most participants expressed approval of this evaluation and suggested that some personal benefit had resulted from participating. It was not made clear just what these benefits might be, and the question was not pursued for fear of compounding results. The interview -- the final instrument -- was often taken as an opportunity to unload with someone listening to them. It may have provided a forum for self-analysis.

Section 7: Conclusions Concerning GFE

There remains the question concerning the usefulness of GFE as an approach to evaluation.

One method for making a theory of GFE operative has been illustrated in the model and methodology of this evaluation. The need for an external standard, and the range of instruments and the methods of analysis made legitimate by a theory of GFE made it useful for evaluating practica concurrently. It was also useful in locating unanticipated outcomes.

As this method led to conclusions concerning the merit of the practica, it could be concluded that this realization of the GFE model was appropriate and worthwhile for this domain. The model shows promise for other evaluation studies.

Section 8: Conclusions and Recommendations for Practica

The salient conclusion concerned the marked and uniform gains made by all participants during practica. From both the program's and the participant's perspective, strong gains are apparent.

The practica seem to have provided the participants with the organization and experience needed to foster confidence and expertise and thereby met some needs. This conclusion is further supported by the participants' expressions of satisfaction concerning their growth in teaching expertise.

The practica increased the POWER of participants. That is to say, they acquired knowledge and skills concerning lesson presentation and the establishment of a positive classroom atmosphere. They became more positive towards self and teaching, and improved their decision-making. These gains increased their teaching effectiveness.

There seem to be differences between the practica concerning gains in teaching expertise and the meeting of needs. These less strong conclusions are primarily derived from the analysis in this chapter, an analysis that was based on the ETM.

The Redekopp practicum may be better at preparing pre-service teachers for teaching. The indication is that this practicum had a somewhat more positive influence on those aspects of teaching related to; (a) creating an environment which facilitated learning, (b) fostering positive attitudes in teachers and (c) improving lesson presentation.

There are two other conclusions concerning gains in teaching expertise. There is an indication that the co-operating teachers of the St. Vital Project have higher expectations and that the disparity between these expectations and actual performances was sufficient to have a negative effect on pre-service teachers.

Second, the Faculty-Based practicum may not be as effective as the others in fostering those skills related to

the creation of an environment that facilitated learning. This program did not seem to meet the perceived needs of participants concerning these skills either.

The Redekopp practicum seems to be most effective and the St. Vital practicum least effective at meeting those needs related to the development of lesson presentation skills.

Some subjective conclusions outside the ETM are now presented but these can only be regarded as rough indicators or suggestions for further study.

The Redekopp practicum seemed to be appreciated the most and the Faculty-Based practicum the least by both co-operating and pre-service teachers. The Faculty-Based practicum was perceived to have no notable qualities. The special features of the Redekopp and St. Vital practica were strongly supported by pre and in-service teachers.

A good number of students complained that the faculty assignments given to be completed during the practicum seriously hindered their ability to devote full attention to the practicum. Both tasks suffered as a result of this additional stress.

Based on a composite of the above conclusions, it is recommended that future practica would benefit from incorporating the following features;

- 1) self-initiated planning of experiences
- 2) self-evaluation of progress
- 3) extensive contact with schools early in the school year
- 4) immersion in the various aspects of schooling

The indications are that these features are important to the task of developing effective teachers.

It is also recommended that any assignments that are concurrent with practica should either be minimal in substance and quantity or should be highly pertinent to the practicum experience. This should allow for more energy being directed towards this experience, thereby increasing the likelihood of greater gains in teaching expertise by pre-service teachers.

BIBLIOGRAPHY

- Adams, R. D., & Craig, J. R. (1983). A status report of teacher education program evaluation. Journal of Teacher Education, 34 (2), 33-36.
- Bradly, J. V. (1968). Distribution free statistical tests. New Jersey: Prentice-hall Inc.
- Buber, M. (1955). Between man and man (R. G. Smith, Trans.). Boston: Beacon Press. (Original work published in 1947)
- Clarcken, R. (1983). Evaluating teacher education graduates and programs. Michigan: University of Detroit. (ERIC Document Reproduction Service No. ED 230 510)
- Collier, R. O., Jr., & Hummel, T. J. (Eds.). (1977). Experimental design and interpretation. California: McCutchan Publ. Corp.
- Evers, J. W. (1980). A field study of goal-based and goal-free evaluation techniques. (Unpublished doctoral dissertation), Western Michigan University, Kalamazoo.
- Ferguson, G. A. (1971). Statistical analysis in psychology & education. New York: McGraw-Hill.
- Freire, P. (1974). The pedagogy of the oppressed. New York: The Seabury Press.
- Gardner, W. E. (1982). NCATE Accreditation: problems, issues, and needed research. In S. M. Hord, T. V. Savage, & L. J. Bethel (Eds.), Toward usable strategies for teacher education program evaluation. Teacher education networks. Austin, Texas: Research and Development Centre for Teacher Education. (ERIC Document Reproduction Service No. ED 229 370)
- Hays, W. L. (1973). Statistics for the social sciences (2nd ed.). New York: Holt, Rinehart and Winston Inc.
- Henderson, E. S. (1978). The evaluation of in-service teacher training. London: Croom-Helm.

- Hord, S. M., Savage, T. V., & Bethel, L. J. (Eds.). (1982). Toward usable strategies for teacher education program evaluation. Teacher education networks. Austin, Texas: Research and Development Centre for Teacher Education. (ERIC Document Reproduction Service No. ED 229 370)
- House, E. R. (1983). Assumptions underlying evaluation models. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models Viewpoints on educational and human services evaluation (pp.45-64). Boston: Kluwer-Nijhoff Publishers.
- Howey K. R., & Gardner, W. E. (Eds.). (1983). The education of teachers A look ahead. New York: Longman Inc.
- Jackson, P.W. (no date available). Is there a best way of teaching Harold Bateman?. Winnipeg: University of Manitoba (Prof. R. T. Gray)
- Jaspers, Karl. (1971). Philosophy of existence (R. F. Grabau, Trans.). Oxford: Basil Blackwell. (Original work published in 1941)
- Keppel, G., & Saufley, W. H. Jr. (1980). Introduction into design and analysis. New York: W. H. Freeman and Company.
- MacDonald, J. (1968). The discerning teacher. Ottawa, Ontario: Canadian Teacher's Federation.
- Madaus, G. F., Scriven, M. S., & Stufflebeam, D. L. (Eds.). (1983). Evaluation models Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhoff Publishing.
- Marascuilo, L. A., & McSweeney, M. (1977). Nonparametric and distribution-free methods for the social sciences. Monterey, California: Brooks/Cole Publishing Co.
- Popham, W. J. (1975). Educational evaluation. Englewood: Prentice-Hall.
- Scriven, M. (1974a). Evaluation perspectives and procedures. In W. J. Popham (Ed.), Evaluation in education (pp. 3-93). Berkeley, California: McCutchan Publ. Corp.

- Scriven, M. (1974b). The concept of evaluation. In M. W. Apple, M. J. Subkoviak, & H. S. Lufler, Jr. (Eds.), Educational Evaluation Analysis and responsibility (pp. 55-82). Bereley, California: McCutchan Publ. Corp.
- Scriven, M. (1981). Evaluation thesaurus (3rd ed.). California: Edgepress.
- Scriven, M. (1983a). Evaluation ideologies. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), Evaluation Models Viewpoints on educational and human services evaluation (pp. 229-260). Boston: Kluwer-Nijhoff Publ.
- Scriven, M. (1983b). The evaluation taboo. In E. R. House, Philosophy of evaluation, New Directions for Program Evaluation, 19, 75-82.
- Smith, N. L. (1982). Introductory guidelines for designing evaluations of teacher education programs. In S. M. Hord, T. V. Savage, & L. J. Bethel (Eds.), Toward usable strategies for teacher education program evaluation. Teacher education networks. Austin, Texas: Research and Development Centre for Teacher education. (ERIC Document Reproduction Service No. ED 229 370)
- Smitheran, V., Anderson, D., Hillis, E., & Loucks, L. (1971). Teacher education: Perseverance or Professionalism. Prince Edward Island: Committee on Teacher Education.
- Stephens, J. M. (1967). The process of schooling A psychological examination. New York: Holt Rhinehart & Winston Inc.
- Tyler, R. W. (1949). Basic principles of curriculum and instruction. Chicago: University of Chicago Press.
- Weil, M., & Joyce, B. (1978). Social models of teaching Expanding your teaching repertoire. New Jersey: Prentice-Hall Inc.
- Willer, D. (1967). Scientific sociology Theory and method. New Jersey: Prentice-Hall Inc.
- Wolf, R. M. (1984). Foundations of competency assessment and program review. New York: Praeger Ltd.

APPENDIX A

OBSERVATION 1: PART 1 - CLASSROOM OBSERVATION

TIME	EVENTS	COMMENTS

APPENDIX B

OBSERVATION 1: PART 2 - PRE-SERVICE TEACHER SURVEY

The following list denotes some aspects of teaching. Rate yourself as to the degree of improvement that you think you require at this point in time.

- 1 - GENERALLY SATISFIED WITH
- 2 - SOME IMPROVEMENT NEEDED
- 3 - MUCH IMPROVEMENT NEEDED

maintaining classroom order:	1	2	3
maintaining lesson focus	1	2	3
encouraging pupils	1	2	3
involving pupils	1	2	3
being sincere while teaching	1	2	3
planning lessons	1	2	3
subject matter competency	1	2	3
liking children	1	2	3
feeling positive about self	1	2	3
feeling positive about teaching	1	2	3
personal importance of teaching	1	2	3
degree of preparedness for teaching	1	2	3
variability of lesson presentations	1	2	3
knowledge of teaching methods	1	2	3
organization during lesson	1	2	3
clarity of lesson presentation	1	2	3
flexibility during lesson	1	2	3
friendliness towards pupils	1	2	3
enthusiasm during lesson	1	2	3
poise during lesson	1	2	3

APPENDIX C

OBSERVATION 2 : PART 1 - CLASSROOM OBSERVATION

VISITATION CODE: _____ DATE: _____

GRADE LEVEL: _____

Variable: PUPIL PARTICIPATION

Operational definition: The frequency of questions asked by the pre-service teacher.

YES/NO	
RECALL	
RELATED INFORMATION	
OPINION	
FUTURE DIRECTION	

Variable: POSITIVE LEARNING ENVIRONMENT

Operational definition: the frequency of reassuring statement/actions made by the pre-service teacher.

FACIAL	
VERBAL	
PHYSICAL	

Variable: ENTHUSIASM

Operational definition: the number of "changes of state" of the pre-service teacher.

VOICE INFLECTION	
GESTURES	
HUMOUR	
PHYSICAL MOVEMENT	

Variable: VARIABILITY

Operational definition: the frequency of different teaching methods used along the 'pure telling' to 'pure discovery' continuum.

LECTURE	
QUESTION/ANSWER	
PUPIL EXERCISES	
OPEN DISCUSSION	
EXPERIMENTS	

Variable: ORGANIZATION

Operational definition: the frequency of confusion (teacher-originated) in the lesson.

GIVING DIRECTIONS	
BLACKBOARD PRESENTATION	
MATERIALS HANDOUT	
MEDIA UTILIZATION	

APPENDIX D

OBSERVATION 3: PART 1 - CLASSROOM OBSERVATION

TIME	EVENTS	COMMENTS

APPENDIX E

OBSERVATION 3: PART 2 - SURVEY: CO-OP TEACHER

CODE _____

Please respond to the following items keeping in mind that they refer solely to the practice teacher who is working with you now.

PART A. Please indicate your rating of the practice teacher for the following aspects of teaching where:

1 - VERY POOR 2 - POOR 3 - ADEQUATE 4 - GOOD
5 - EXCEPTIONAL

1. maintenance of classroom order during lessons

1 ____ 2 ____ 3 ____ 4 ____ 5 ____

2. lesson planning skills

1 ____ 2 ____ 3 ____ 4 ____ 5 ____

3. allowance for pupils' individual differences during lessons

1 ____ 2 ____ 3 ____ 4 ____ 5 ____

4. working relationship with immediate superiors
(principal, supervisors, you - average value)

1 ____ 2 ____ 3 ____ 4 ____ 5 ____

5. rapport with the other teachers in the school

1 ____ 2 ____ 3 ____ 4 ____ 5 ____

6. dependability in meeting the requirements of the practicum program

1 ____ 2 ____ 3 ____ 4 ____ 5 ____

7. innovativeness of lessons

1 _____ 2 _____ 3 _____ 4 _____ 5 _____

PART B.

1. How highly would you rate him/her on a scale of 1 to 7, where 7 is high, as to overall competence in teaching.

2. Briefly describe the aspect of teaching for which he/she improved most significantly during the practicum experience.

3. Briefly describe the aspect of teaching for which the practice teacher:

a) requires the most improvement

b) requires the least improvement

4. Comments/recommendations concerning the practicum program (use the reverse side of the paper if necessary) ?

APPENDIX F

OBSERVATION 3 : PART 3. POST-OBSERVATION SURVEY:
PRE-SERVICE TEACHER

VISITATION CODE: _____ DATE _____

GRADE LEVEL: _____

DIRECTIONS

Please complete the questionnaire by responding to the provided questions or statements.

1. Did you make a written lesson plan?
YES _____ NO _____

RESPOND TO #2 ONLY IF THE PREVIOUS ANSWER WAS YES

2. Did your lesson plan include objectives/learning outcomes?
YES _____ NO _____

3. How much time did you spend on planning for this lesson?

LESS THAN 1/2 HR. _____
1/2 TO 3/4 HR. _____
3/4 TO 1 HR. _____
1 - 1 1/4 HR. _____
MORE THAN 1 1/4 HR. _____

4. How would you rate the quality of the effort that you put into planning for this lesson.

VERY POOR _____ POOR _____ ADEQUATE _____ GOOD _____
EXCELLENT _____

5. In your opinion, what percent of the success of the lesson is attributable to the extent and adequacy of your planning for it.

10% _____ 25% _____ 50% _____ 75% _____ 90% _____

6. Give yourself an overall rating as to the success of the lesson just presented.

VERY POOR _____ POOR _____ ADEQUATE _____ HIGH _____
VERY HIGH _____

7. During the lesson planning stage, how much thought did you give to a potential straying from the prepared plan if the actual classroom situation suggested such an action?

NONE _____ VERY LITTLE _____ SOME _____ ENOUGH _____
A LOT _____

8. How closely did you match the lesson to the pupils' abilities?

NOT AT ALL _____ VERY LITTLE _____ SOME _____
ENOUGH _____ A LOT _____

9. How closely did you match the lesson to the pupils' interests?

NOT AT ALL _____ VERY LITTLE _____ SOME _____
ENOUGH _____ A LOT _____

10. During the presentation of the lesson, how aware were you of other events occurring in the classroom.

NOT AT ALL _____ VERY LITTLE _____ ENOUGH _____
USUALLY _____ ALWAYS _____

11. How deep was your understanding of the subject matter for this lesson?

VERY POOR _____ POOR _____ ADEQUATE _____
GOOD _____ EXCELLENT _____

12. Rate yourself as to your satisfaction with the level of your subject matter competency required for this lesson.

VERY LOW _____ LOW _____ AVERAGE _____ GOOD _____
EXCELLENT _____

13. Rate yourself as to the "strength of your liking" of the pupils in the class just taught.

VERY LOW _____ LOW _____ NORMAL _____ HIGH _____
VERY HIGH _____

14. What percent of the success of your lesson is attributable to your liking the pupils?

10% _____ 25% _____ 50% _____ 75% _____ 90% _____

15. What percent of the success of the lesson is attributable to the children liking you?

10% _____ 25% _____ 50% _____ 75% _____ 90% _____

16. How do you feel about yourself right now?
 VERY DOWN _____ DOWN _____ NEUTRAL _____ UP _____
 VERY UP _____

EXAMINE THE NEXT 3 QUESTIONS FIRST BEFORE RESPONDING TO #17

17. What percent of this feeling is attributable to the just completed lesson presentation?

18. What percent of this feeling is attributable to your practice teaching experience in general?

19. What percent of this feeling is attributable to other circumstances in your life?

20. What impact did the presentation of the lesson have on your feeling of self-worth?

GREATLY LOWERED IT _____
 LOWERED IT _____
 NONE _____
 RAISED IT _____
 GREATLY RAISED IT _____

21. What priority does teaching have in your life?
 VERY LOW _____ LOW _____ AVERAGE _____ HIGH _____
 VERY HIGH _____

22. Rate the importance of the lesson just presented with regard to your professional development.
 VERY LOW _____ LOW _____ AVERAGE _____ HIGH _____
 VERY HIGH _____

23. Rate the importance of the lesson just presented with regard to your personal growth?
 VERY LOW _____ LOW _____ AVERAGE _____ HIGH _____
 VERY HIGH _____

24. How prepared did you feel you were for the lesson?
 NOT AT ALL _____ A LITTLE _____ ADEQUATELY _____
 HIGHLY _____ VERY HIGHLY _____

25. Was the presenting of the lesson rewarding for you in any way?

YES _____ NO _____

RESPOND TO ITEM #26 ONLY IF THE PREVIOUS RESPONSE WAS 'YES'

26. Rank the following list in order of importance of reward for you where ;

1 - most important, 5 - least important.

FEEDBACK RECEIVED FROM PUPILS	_____
KNOWLEDGE GAINED BY PUPILS	_____
SKILLS GAINED BY PUPILS	_____
EXPERTISE GAINED BY YOU	_____
CONFIDENCE GAINED BY YOU	_____

27. The following list denotes some aspects of teaching. Rate yourself as to the degree of improvement that you think you require at this point in time.

- 1 - GENERALLY SATISFIED WITH
 2 - SOME IMPROVEMENT NEEDED
 3 - MUCH IMPROVEMENT NEEDED

maintaining classroom order:	1	2	3
maintaining lesson focus	1	2	3
encouraging pupils	1	2	3
involving pupils	1	2	3
being sincere while teaching	1	2	3
planning lessons	1	2	3
subject matter competency	1	2	3
liking children	1	2	3
feeling positive about self	1	2	3
feeling positive about teaching	1	2	3
personal importance of teaching	1	2	3
degree of preparedness for teaching	1	2	3
variability of lesson presentations	1	2	3
knowledge of teaching methods	1	2	3
organization during lesson	1	2	3
clarity of lesson presentation	1	2	3
flexibility during lesson	1	2	3
friendliness towards pupils	1	2	3
enthusiasm during lesson	1	2	3
poise during lesson	1	2	3

APPENDIX G

OBSERVATION 3: PART 4 - INTERVIEW: PRE-SERVICE TEACHER

CODE _____ DATE _____

1. How has the practicum experience made you:
 - i) more aware of your strengths and weaknesses in relation to being an effective teacher
 - ii) more capable of being an effective teacher

2. What negative effects or impact (if any) did
 - i) the practicum experience have on you?
 - ii) the concurrent interaction of the practicum program and the faculty have on you?

3. How strong was the influence (positive and negative) of the co-operating teacher with respect to you mimicing him/her and with respect to you being dominated by him/her?

4. Did the practicum experience help you to resolve the issue (if it existed) of theory versus real world realities of teaching? Explain.

5. How faithfully did you follow the recommendations, advice etc. of the individuals directly connected with the faculty's practicum program? Explain.

6. What recommendations/comments have you in regard to
 - i) the faculty's practicum program?
 - ii) the faculty programs in general?

7. Was your involvement with this evaluation study a helpful, a neutral or a hindering experience for you? Explain.

APPENDIX H

SURVEY FORM A

DIRECTIONS:

Please complete all the items in SECTION A by selecting one of: STRONGLY AGREE, AGREE, DISAGREE, STRONGLY DISAGREE. Section B, at the end of this questionnaire, need only be completed if you wish to contribute additional information. The choices denote:

STRONGLY AGREE (SA)	- Agree, with little uncertainty
AGREE (A)	- Agree, but with some hesitancy
DISAGREE (D)	- Disagree, but with some hesitancy
STRONGLY DISAGREE (SD)	- Disagree, with little uncertainty

SECTION A

1. Knowledge gained during schooling increases a person's ability to function in our society.
SA ___ A ___ D ___ SD ___
2. Skills gained during schooling increase a person's ability to function in our society.
SA ___ A ___ D ___ SD ___
3. Attitudes to oneself acquired through schooling have an impact (positively correlated) on a person's ability to function in our society.
SA ___ A ___ D ___ SD ___
4. Attitudes to learning acquired through schooling have an impact (positively correlated) on a person's ability to function in our society.
SA ___ A ___ D ___ SD ___
5. Increasing a person's ability to make decisions through schooling increases his/her ability to function in our society.
SA ___ A ___ D ___ SD ___

6. Maintaining classroom order is important for effective teaching.

SA ___ A ___ D ___ SD ___

7. Maintaining the focus during a lesson is important for effective teaching.

SA ___ A ___ D ___ SD ___

8. Establishing a learning environment that allows a student to feel welcome in it is important for effective teaching.

SA ___ A ___ D ___ SD ___

9. Establishing a learning environment that allows for student participation in a non-trivial way is important for effective teaching.

SA ___ A ___ D ___ SD ___

10. A teacher that is 'honest' (perceived by the pupils as believing in the subject matter and teaching task) is more effective.

SA ___ A ___ D ___ SD ___

11. Well planned lessons are important for effective teaching.

SA ___ A ___ D ___ SD ___

12. Knowing the subject matter in some depth is important for effective teaching.

SA ___ A ___ D ___ SD ___

13. Varying the presentation style of lessons is important for effective teaching.

SA ___ A ___ D ___ SD ___

14. Matching the teaching method with needs of the pupils is important for effective teaching.
SA ___ A ___ D ___ SD ___
15. A well organized lesson is important for effective teaching.
SA ___ A ___ D ___ SD ___
16. A clearly presented lesson is important for effective teaching.
SA ___ A ___ D ___ SD ___
17. A teacher who is able to change the lesson in midstream if deemed appropriate is more effective.
SA ___ A ___ D ___ SD ___
18. A teacher who is friendly is more effective.
SA ___ A ___ D ___ SD ___
19. A teacher who is enthusiastic is more effective.
SA ___ A ___ D ___ SD ___
20. A teacher who is poised is more effective.
SA ___ A ___ D ___ SD ___
21. A teacher who likes his/her pupils is more effective.
SA ___ A ___ D ___ SD ___
22. It is important for a teacher to have a positive self-image in order to be effective.
SA ___ A ___ D ___ SD ___
23. It is important for a teacher to feel comfortable with the teaching task in order to be effective.
SA ___ A ___ D ___ SD ___
24. It is important for a teacher to perceive the teaching task as important in order to be effective.
SA ___ A ___ D ___ SD ___
25. It is important for the teacher to feel that the teaching task is rewarding in order to be effective.
SA ___ A ___ D ___ SD ___

26. It is important for the teacher to feel prepared for the teaching task in order to be effective.

SA ___ A ___ D ___ SD ___

27. It is important that a teacher present an appropriate appearance as determined by immediate superiors.

SA ___ A ___ D ___ SD ___

28. It is important to have a positive attitude towards superiors.

SA ___ A ___ D ___ SD ___

29. It is important to establish a good rapport with fellow teachers.

SA ___ A ___ D ___ SD ___

30. It is important to be perceived as being dependable by the administrators of a school.

SA ___ A ___ D ___ SD ___

31. It is important to be perceived as being competent by the administrators of a school.

SA ___ A ___ D ___ SD ___

SECTION B

If you have any additional thoughts that pertain to the conditions of effective teaching and attributes of effective teachers please include them in the space provided.

APPENDIX I

SURVEY FORM B

DIRECTIONS:

Please complete all the items in SECTION A by selecting one of: STRONGLY AGREE, AGREE, DISAGREE, STRONGLY DISAGREE. Section B, at the end of this questionnaire, need only be completed if you wish to contribute additional information. The choices denote:

STRONGLY AGREE (SA)	- Agree, with little uncertainty
AGREE (A)	- Agree, but with some hesitancy
DISAGREE (D)	- Disagree, but with some hesitancy
STRONGLY DISAGREE (SD)	- Disagree, with little uncertainty

SECTION A

1. Knowledge gained during schooling increases a person's ability to function in our society.

SA ___ A ___ D ___ SD ___

2. Skills gained during schooling increase a person's ability to function in our society.

SA ___ A ___ D ___ SD ___

3. Attitudes to oneself acquired through schooling have an impact (positively correlated) on a person's ability to function in our society.

SA ___ A ___ D ___ SD ___

4. Attitudes to learning acquired through schooling have an impact (positively correlated) on a person's ability to function in our society.

SA ___ A ___ D ___ SD ___

5. Increasing a person's ability to make decisions through schooling increases his/her ability to function in our society.

SA ___ A ___ D ___ SD ___

6. Maintaining classroom order is important for effective teaching.
SA ___ A ___ D ___ SD ___
7. Maintaining the focus during a lesson is important for effective teaching.
SA ___ A ___ D ___ SD ___
8. Establishing a learning environment that allows a student to feel welcome in it is important for effective teaching.
SA ___ A ___ D ___ SD ___
9. Establishing a learning environment that allows for student participation in a non-trivial way is important for effective teaching.
SA ___ A ___ D ___ SD ___
10. A teacher that is 'honest' (perceived by the pupils as believing in the subject matter and teaching task) is more effective.
SA ___ A ___ D ___ SD ___
11. Well planned lessons are important for effective teaching.
SA ___ A ___ D ___ SD ___
12. Knowing the subject matter in some depth is important for effective teaching.
SA ___ A ___ D ___ SD ___
13. Varying the presentation style of lessons is important for effective teaching.
SA ___ A ___ D ___ SD ___
14. Matching the teaching method with needs of the pupils is important for effective teaching.
SA ___ A ___ D ___ SD ___
15. A well organized lesson is important for effective teaching.
SA ___ A ___ D ___ SD ___
16. A clearly presented lesson is important for effective teaching.
SA ___ A ___ D ___ SD ___
17. A teacher who is able to change the lesson in midstream if deemed appropriate is more effective.
SA ___ A ___ D ___ SD ___

18. A teacher who is friendly is more effective.

SA ___ A ___ D ___ SD ___

19. A teacher who is enthusiastic is more effective.

SA ___ A ___ D ___ SD ___

20. A teacher who is poised is more effective.

SA ___ A ___ D ___ SD ___

21. A teacher who likes his/her pupils is more effective.

SA ___ A ___ D ___ SD ___

22. It is important for a teacher to have a positive self-image in order to be effective.

SA ___ A ___ D ___ SD ___

23. It is important for a teacher to feel comfortable with the teaching task in order to be effective.

SA ___ A ___ D ___ SD ___

24. It is important for a teacher to perceive the teaching task as important in order to be effective.

SA ___ A ___ D ___ SD ___

25. It is important for the teacher to feel that the teaching task is rewarding in order to be effective.

SA ___ A ___ D ___ SD ___

26. It is important for the teacher to feel prepared for the teaching task in order to be effective.

SA ___ A ___ D ___ SD ___

27. It is important that a teacher present an appropriate appearance as determined by immediate superiors.

SA ___ A ___ D ___ SD ___

28. It is important that a teacher have a positive attitude towards superiors.

SA ___ A ___ D ___ SD ___

29. It is important that a teacher establish a good rapport with fellow teachers.

SA ___ A ___ D ___ SD ___

30. It is important that a teacher be perceived as being dependable by the administrators of a school.

SA ___ A ___ D ___ SD ___

31. It is important that a teacher be perceived as being competent by the administrators of a school.

SA ___ A ___ D ___ SD ___

SECTION B

If you have any additional thoughts that pertain to the conditions of effective teaching and attributes of effective teachers please include them in the space provided.

APPENDIX J

SURVEY FORM D

DIRECTIONS:

Please complete all the items in SECTION A by selecting one of: STRONGLY AGREE, AGREE, DISAGREE, STRONGLY DISAGREE. The choices mean:

STRONGLY AGREE (SA)	- Agree almost totally
AGREE (A)	- Mildly agree with
DISAGREE (D)	- Mildly disagree with
STRONGLY DISAGREE (SD)	- Disagree almost totally

SECTION A

1. The things that you learned about at school helped you to take care of yourself as an adult.
SA ___ A ___ D ___ DA ___
2. The things that you learned to do at school helped you to take care of yourself as an adult.
SA ___ A ___ D ___ DA ___
3. Success in life increases if your schooling helps you to develop a positive attitude to yourself.
SA ___ A ___ D ___ DA ___
4. Success in life increases if your schooling helps you to develop a positive attitude to learning.
SA ___ A ___ D ___ DA ___
5. Success in life increases if your schooling helps you to learn to make decisions.
SA ___ A ___ D ___ DA ___

6. More learning occurred when the teacher kept good order in the classroom.

SA ___ A ___ D ___ DA ___

7. More learning occurred when the teacher kept on topic during a lesson.

SA ___ A ___ D ___ DA ___

8. More learning occurred when the teacher got you to feel comfortable with learning.

SA ___ A ___ D ___ DA ___

9. More learning occurred when the teacher let you take part in decisions about learning.

SA ___ A ___ D ___ DA ___

10. More learning occurred when the teacher seemed to believe in what he/she was teaching.

SA ___ A ___ D ___ DA ___

11. You learned better when the teacher knew what he/she was doing during the lesson.

SA ___ A ___ D ___ DA ___

12. You learned better when the teacher knew the topic well.

SA ___ A ___ D ___ DA ___

13. When the ways of giving lessons were varied more learning occurred.

SA ___ A ___ D ___ DA ___

14. It was important that lessons were given in a way that suited how you learned things.

SA ___ A ___ D ___ DA ___

15. You learned better when the lesson seemed to fit together.

SA ___ A ___ D ___ DA ___

16. It was important for a lesson to have been presented clearly.

SA ___ A ___ D ___ DA ___

17. It was important for the teacher to be able to change a lesson in midstream when necessary.

SA ___ A ___ D ___ DA ___

18. More learning occurred when the teacher was friendly.

SA ___ A ___ D ___ DA ___

19. More learning occurred when the teacher was enthusiastic.

SA ___ A ___ D ___ DA ___

20. More learning occurred when the teacher seemed confident.

SA ___ A ___ D ___ DA ___

21. More learning occurred when the teacher liked you.

SA ___ A ___ D ___ DA ___

SECTION B

If you have anything to add that has to do with what things should be like in a classroom in order for you to learn better, please write them in the space below.

APPENDIX K

Observations from a Perspective Outside the ETM

This appendix is a compendium of observations and represents an attempt to organize the large amount of data that lies outside the domain of the ETM. No attempt was made to establish any statistical significance. These observations therefore should be perceived as rough indicators that may be useful for revealing trends.

Two perspectives were considered; (1) host schools represented by co-operating teachers and (2) practica participants. The observations were derived from comments of co-operating teachers included in questionnaires and from interviews with participants.

The co-operating teachers had differing attitudes towards the practica. The Redekopp practicum appeared to receive the most enthusiastic support with seven expressing high regard for this program with such comments as "excellent program for the student teacher" and "feel more involved in the program". The aspects of self-initiation and self-evaluation were favourably regarded and were seen as beneficial for developing teaching expertise. No negative feelings were expressed.

For the St. Vital practicum, five co-operating teachers expressed high regard for the program with such comments as "best program available" and "the student

teachers are more ready to teach". The aspect of early school contact was seen as beneficial for developing teaching-readiness, especially for those tasks of the 'book-keeping' type.

There however were concerns. One faculty advisor was perceived to be out of touch with current educational thought. Several teachers felt that there should be more frequent communication between the Faculty and host schools. Two felt that Friday was not the best day for ongoing weekly contact time.

For the School-Based practicum, three co-operating teachers expressed high regard with such comments as "completely satisfied, no flaws" and "step in the right direction". Two (both from the same school) expressed strong negative views. They found the workload too demanding and the presence of pre-service teachers intrusive. Two preferred keeping pre-service teachers all week and three felt that pre-service teachers should learn 'book-keeping' tasks. One felt that there should be more contact with the faculty advisor.

The Faculty-Based practicum appeared to be the least regarded of the four. Two co-operating teachers expressed high regard with such comments as "impressed this year, not last year" and "program went very well". This group, in general, did not make many comments concerning this program.

Two opinions were expressed that were common to all practica. First, the juxtaposition of faculty-based assignments and practica was seen as an impediment for producing effective teachers by placing an additional stress on pre-service teachers and by reducing focus on and energy for practica (total of six). Second, the time spent in practica should be increased (total of nine).

The pre-service teachers also had differing attitudes towards the practica. The same order concerning enthusiasm for practica was noted.

For the Redekopp practicum, seven participants expressed high regard and support for its continuance with such comments as "would not have grown without it" and "can see benefits compared to other programs". The aspects of self-initiation and self-evaluation were strongly supported. No negative feelings were evident except for one negative experience with a co-operating teacher. Only one participant desired increased time in practica.

Six participants of the St. Vital Project expressed enthusiasm for this program with such comments as "good program that should continue" and "everyone should be in this program". The aspect of early contact with schools was strongly supported. There were no indications of a desire for longer practica.

There were concerns. Friday was seen as not a good day for the weekly visit (indicated by four). One felt like a

teacher-aide during these Friday visits. Two indicated poor communication with the faculty advisor. One reported friction between St. Vital participants and participants of other practica concerning the advertised 'specialness' of this project. Two participants reported problems with co-operating teachers concerning excessive dominance.

Four participants of the School-Based practicum expressed enthusiasm for this program with such comments as "good blend of what is needed" and "should be continued". Four indicated a desire for longer practica.

There also were concerns. One indicated a need for more contact with the faculty advisor. A need for more feedback and feedback from different observers was indicated by two participants. One wanted Friday visits to the Faculty be removed as this disrupted the continuity of practica.

The Faculty-Based practicum received the fewest expressions of enthusiasm with only one participant providing any. Three indicated a desire for longer practica.

There were many concerns; (a) more frequent contact with faculty advisors (two), (b) more feedback (two), (c) better prepared co-operating teachers (three), (d) an unclear evaluation process (four) and (e) more information on the host schools (six). Two participants indicated

problems with co-operating teachers concerning excessive criticism.

A common concern was noted. The faculty assignments that were concurrent with the practica placed an additional burden on participants; one that interfered by reducing commitment and energy devoted to practica. A sense of being pulled in conflicting directions was evident. The number of participants expressing this is; (a) St. Vital project (eight), (b) Faculty-Based (five), (c) School-Based (four) and Redekopp practicum (two). As this concern was also expressed by some co-operating teachers, it may be advisable for program managers to reconsider the extent and nature of faculty assignments that are concurrent with practica.

In summary, the Redekopp and St. Vital practica seemed to have the strongest support with the School and Faculty-Based practica having the least support from co-operating and pre-service teachers. The special features of the Redekopp and St. Vital practica were applauded by those concerned. This may indicate that these features are useful for improving teaching expertise. Their presence or absence may be the critical factor that was responsible for the attitudes observed and hence they should be incorporated into any practicum program. This may also explain why, for the less favoured practica, participants felt that they required a longer practicum experience.

The relationship between practica and faculty programs, and the nature of the latter were broad concerns for the participants.

They had varying opinions concerning the correspondence between practica and faculty programs; a correspondence that concerned the extent that faculty programs prepared participants for the situations encountered in practica. Six from the St. Vital group felt that this correspondence was inadequate. For each of the Faculty and School-Based groups, three expressed this view. There was one from the Redekopp group that felt the correspondence was inadequate.

Participants offered many suggestions for improving the faculty programs as they related to practica. These were too numerous to itemize but the most frequent belonged in the following categories; (a) connecting theory to application in a better way, (b) restructuring the four year faculty program and increasing the time span, (c) placing less emphasis on the 'ideal' classroom and more on the realities of the classroom.