AN INVESTIGATION OF EVALUATION STRATEGIES OF LANGUAGE ARTS TEACHERS

SUSAN ANN BERNADETTE TILLEY

AN INVESTIGATION OF EVALUATION STRATEGIES OF LANGUAGE ARTS TEACHERS

by

Susan Ann Bernadette Tilley, B.A., B.Ed.

A thesis submitted to the School of Graduate Studies
in partial fulfillment of the requirements
for the degree of
Master of Education

Department of Curriculum and Instruction
Memorial University of Newfoundland
September 1986

St. John's Newfoundland

ISBN 0-315-36988-4

ACKNOWLEDGEMENTS

Sincere appreciation is extended to Dr. Frank Wolfe who supervised this thesis. I am indebted to him for his invaluable assistance and encouragement.

A special thank you is extended to my family and friends who tolerated the constant chatter concerning my thesis. The end is in sight.

ABSTRACT

The writer, while employed as either a classroom teacher or as coordinator, was made aware of the confusion existing in the area of evaluation. In her mind, many questions needed answering, including those concerned with how teachers evaluate their students and why they choose the strategies they do. The purpose of this study was to examine questions such as these but also to progress one step further by proposing a plan to help alleviate problems in the area.

In order to assess practice at a classroom level, the survey presented in Chapter III was carried out. Data on grade four language arts teachers and the strategies they used for student evaluation were collected. The chapter includes a description and analysis of this data.

In Chapter II the writer reviewed the literature pertaining to the area of evaluation. The chapter follows the logical pattern which emerged during the research of the material.

The first section examines the roots of evaluation while expounding on the test and measurement movement's influence on its development.

The second section examines research concerning program evaluation. The writer suggests much information can be utilized from this research and applied to student evaluation in the classroom.

The third section examines student evaluation. A link is established between yesterday's theory and today's practice.

This chapter attempts to provide a picture of what research is suggesting should be occurring in student evaluation at the classroom level.

Chapter II presents "what should be" occurring in practice, while Chapter III provides a description of "what is". The final two chapters attempt to provide the means by which the "what is" can be converted into the "what should be".

Chapter IV presents a rationale for the use of inservice to effect this suggested change. The complexity of the change process is described in detail. The necessity of giving this process serious consideration when developing inservice is emphasized. Also, the chapter includes a discussion of what constitutes effective inservice and the importance of agreed upon time-lines for achieving goals.

Chapter V is an inservice plan for grade four language arts teachers, principals and professional school board personnel. The plan is considered an initial step in attempting to effect change in teachers' methods of evaluating students.

TABLE OF CONTENTS

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

> Evaluation can be a powerful force; it is essential to improvement in schools. It can also be dangerous if misused; so glib acceptance of any activity as evaluative (with the thought of at worst it will be harmless) must be avoided (Hayman and Napier, 1975, p. 129).

Statement of the Problem

Teachers presently working in the school system are assigned the responsibility of evaluating their students. However, little opportunity seems available for these teachers to gain the knowledge and skills necessary to accomplish this task.

Teacher training lacks sufficient emphasis in the area of student evaluation, and experience in the field is limited in its value to teachers in their attempts to become proficient evaluators.

The concept of evaluation is discussed, in great detail, in the literature. Programs being offered as part of the school curriculum refer to evaluation; especially in connection with course objectives. However, little of what the literature has to offer as advice to the educator has apparently been converted into classroom practice.

> Despite a half century of research and the development of several sophisticated theories, the teacher's classroom activities have been relatively unaffected by what the learning theorist has to say (Jackson, 1968, p. 148).

This statement can be extended to include the thoughts of any theorist. Practice in evaluation is grounded in test and measurement theory which advances the importance of testing as the means of evaluating students. Jackson (1968) suggests that "Tests are as indigenous to the school environment

as are textbooks or pieces of chalk" (p. 19). In many instances, evaluation and testing are believed to be synonymous.

There needs to be a greater understanding of evaluation developed among those who are involved daily in the process.

As Hayman and Napier (1975) state:

> Evaluation is carried out by people doing their jobs - by teachers seeking to improve learning in their classrooms, by administrators seeking to make more efficient decisions, and by program developers seeking to be more certain of the value of their products. All need a basic knowledge of how (and why) evaluation is planned and conducted ... (p. 13).

For those teachers already in the workforce, gaining this "knowledge" referred to by Hayman and Napier becomes a matter of concern for district school boards. Inservice presents itself as the most viable option for teachers who wish to develop a greater understanding of evaluation and its role in the classroom.

In the past, many have viewed inservice as an inadequate tool for affecting change. Days alloted for inservice are limited by the Department of Education, and finances for inservice use are at an all-time low. As well, school board personnel often lack adequate training in delivery of effective inservice and are without an understanding of the complexity of the change process.

The problem stated simply is that teachers are being asked to perform a task which they may well be ill-prepared to accomplish, and little is being done to alleviate the situation.

Purpose of the Study

Oftentimes educators tend to bemoan the problems in teaching and to

offer numerous suggestions as solutions to these problems. As Fullan (1982) remarks:

> There is no shortage of recommendations about how the ills of education should be rectified. But the remedies remain pie in the sky as long as competing "shoulds" fight it out without an understanding of what is (p. 39).

This writer attempts to gain a piece of this "pie" by first examining the "what is" and then offering a possible remedy.

The purpose of the study is twofold: first, it is to find out how grade four language arts teachers are evaluating their students; and second, it is to propose, where necessary, a means by which school board personnel can attempt to effect change in practice at a classroom level.

Chapter II, the review of the literature, provides information necessary to understand the concept of evaluation. It presents an historical perspective of the development of evaluation. Three aspects are examined closely: the test and measurement movement; program evaluation; and student evaluation.

Chapter III presents the "what is". It is a description of a survey conducted concerning the evaluation strategies of grade four language arts teachers. Description of the survey and analysis of the data are supplied.

Chapter IV provides a rationale for the creation of an inservice plan to promote the development of teachers as proficient evaluators. A discussion of the difficulty of affecting change in teachers' practices is included.

Chapter V is a presentation of a plan for inservice, in the area of evaluation, for grade four language arts teachers.

Scope and Limitations

One limitation of the study is its narrow focus on grade four language arts teachers. There is a need for research with a much broader scope in

order to promote an understanding of the evaluation strategies of teachers at all levels.

The data on grade four language arts teachers and their evaluation strategies have been collected through the use of a questionnaire. Readers may consider this a limitation in light of the possibility of respondents falsifying any information given. However, the writer feels that with the lack of possible incentives for providing misleading information, and with an 80 percent return, the expectation of receiving a majority of valid information is a reasonable one.

Teachers returned their questionnaires through the coordinators, and this may be viewed as a limitation. Some pressure may have been felt knowing information was available for perusal by district office personnel. There was an attempt to alleviate this possible pressure by emphasizing the use to be made of the data collected.

There are limitations to the inservice plan developed. Three days may be viewed as an extravagant use of inservice days for just one area of concern. However, coordinators must impress upon those involved the connection which evaluation has to other areas of concern, such as teaching methodology. They must emphasize the possible effects inservice in one area, such as evaluation, may have on other areas of concern. Inservice in evaluation must be seen as a priority.

Another limitation of a plan is how it may be perceived as unmalleable. For this plan to be successful, as an initial step in effecting change in teacher practices, it must be adapted to individual school board situations.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

In order to attempt a well-organized and comprehensive discussion of the historical development of evaluation, this chapter is divided into three main sections.

The early beginnings section deals with the influence of the test and measurement movement on the initial development of evaluation.

The second section includes an examination of the appearance, over a short span of time, of numerous program evaluation models. This section also presents a discussion of the development of changing theories brought to light through educational evaluation research.

The final section attempts to examine student evaluation. Surfacing throughout the literature is the obvious contradiction between what researchers/educators are suggesting evaluation should be comprised of and what teachers are in practice doing in the name of evaluation. This writer attempts to provide a reason for this discrepancy.

Early Beginnings

The roots of evaluation can be traced far back in history to the times when testing was becoming popular. Chinese officials were conducting civil service examinations as early as 2000 B.C. (Dubois, 1970, p. 3). The examinations provided a singling-out vehicle to allow officials to decide on individuals appropriate for service. Intelligence testing became popular with the introduction of Binet.

> The first mention of Binet's work to the National Education Association was made at the 1909 convention. Just two years later its Committee on Resolutions gave whole-hearted approval to school admission and

> leaving standards based on intelligencetesting. A veritable deluge of tests was produced thereafter (Jonrich, 1962, p. 5).

Educators were becoming convinced of the usefulness of standardized tests to "evaluate" students. The early part of the twentieth century saw much attention given to psychological testing.

> In the early 1900's Robert Thorndike, called the father of the educational testing movement, was instrumental in convincing educators of the value of measuring human change. (Worthen and Sanders, 1973, p. 2).

Evaluation was this narrowly focused "measuring of human change" whose prime purpose was to differentiate between "good" and "poor" (referring to ability) students.

In the early 1900's psychological testing was being used in other areas besides education. However, educators as well as others were drawing inferences from what they witnessed and were applying these inferences to the school situation. Tyler and Wolf (1974) state:

> During World War I the psychological testing used to select those who could quickly learn to be officers and technical personnel from the two million men enlisted in the military service impressed not only the psychologists but also leaders in education, business and civic affairs. Schools, colleges, and large industrial organizations were seen as the proper civilian settings for the initiation and development of testing, and educational institutions and the personnel departments of industrial organizations were engaged to select and sort persons. Since tests had proved useful in selecting and sorting military personnel, it seemed that similar tests could be developed for civilian conditions, and for children and youth as well as young adults (p. 4).

Educational testing thus gained important status in the school environment. Testing and measurement evolved as a discipline unto itself. Madaus et al. (1983) refer to the time period 1900-1930 as the "Age of Efficiency and

Testing". The use of test data was expanded from being an assorting vehicle to its being proof of school, program, or teacher effectiveness. Madaus et al. (1983) state:

> In the United States the earliest formal attempt to evaluate the performance of schools took place in Boston 1845. This event is important in the history of evaluation because it began a long tradition of using pupil test scores as a principal source of data to evaluate the effectiveness of a school or instructional program (p. 5).

Information gathered from tests and measurement was now seen in a different, more threatening light because of its move of emphasis from individual students to programs, schools and administrators. Stufflebeam and Webster (1983) explain that:

> Since the 1930's, American education has been inundated with standardized-testing programs. Probably every school district in the United States has some type of standardized-testing program, and, formerly, many educators have tended to equate the results of a standardized-testing program with the information needed to evaluate the quality of a school district, a school, a program, and, in some cases, even a teacher (p. 29).

The public became involved in educational evaluation as its cry for accountability was heard.

> The cry for accountability in education continued to increase until several state departments of education began to design state assessment systems and state legislatures began to require reports from all schools on student achievement in subjects such as reading and mathematics (Worthen and Sanders, 1973, p. 7).

Large scale testing programs were put in place in response to pressure exerted by the public. Also, Worthen and Sanders (1973) suggest that this was the time for massive development of school grading systems - with their dependency on the results of teacher-made tests.

The established test and measurement movement flourished during a time when educational research was based on a "scientific paradigm". Patton (1978) provides the following definition for the term "paradigm":

> A paradigm is a world view, a general perspective, a way of breaking down the complexity of the real world. As such, paradigms are deeply embedded in the socialization of adherents and practioners: paradigms tell them what is important, legitimate, and reasonable. Paradigms are also normative, telling the practitioner what to do without the necessity of long existential or epistemological consideration. But it is this aspect of paradigms that constitutes both their strength and their weakness - their strength in that it makes action possible, their weakness in that very reason for action is hidden in the unquestioned assumption of the paradigm (p. 203).

The world view of educational research was dominated by scientific thought. The test and measurement movement could work quite well within such a context. Quantitative data were the meat of the movement and experimental design a trustworthy framework.

> Evaluation research is dominated by the largely unquestioned natural science paradigm of hypothetico-deductive methodology. This dominant paradigm assumes quantitative measurement, experimental design, and multi-variate, parametric statistical analysis to be the epitome of "good" science (Patton, 1978, p. 203).

Educational evaluation research was desirous of being considered "good" science and therefore reinforced the use of scientific methods espoused by those involved in the test and measurement movement. In the following quote, Guba (1969) describes the effect of this on evaluation:

> First evaluation was given an instrumental focus; the science of evaluation was viewed as the science of instrument development and interpretation. Second, the approach tended to obscure the fact that value judgements are necessarily involved. Third, evaluation tended to be limited to those variables for which the science of measurement had successfully evolved instruments; other variables became known

> as "intangibles", a characterization equivalent to saying they couldn't be measured; hence had no utility, and ultimately, no importance (p. 31).

It became apparent to many in the field of educational research (Guba, 1981; Stake, 1967; Patton, 1980; and others) that the quantitative data and the scientific methods used to collect this data were not always yielding pertinent information. Those "intangibles" referred to by Guba needed to be examined more closely. It was time that evaluation was to be seen apart from the test and measurement movement.

Tyler, who is often referred to as the "father of educational evaluation", provided new direction for evaluation. Guba and Lincoln (1981) suggest that:

> Until this time [Tyler's time] evaluation has existed largely for the purpose of making judgements about individual students in relation to test norms and of labelling students as overachievers, underachievers, or "normal" achievers. Tyler forged a new dynamic for evaluation, making it the mechanism for continuous circular and instructional improvement. The scope of influence for evaluation was thereby greatly enlarged (p. 5).

Tyler aided in enlarging the scope of influence of evaluation by bringing about the differentiation between the concepts of measurement and evaluation.

> Tyler made it clear that they were separate processes, with measurement being simply one of several possible tactics to be enlisted in support of evaluation (Guba and Lincoln 1981, p. 5).

Efforts in evaluation shifted from the test and measurement emphasis on individual students and evaluation to a focus upon curriculum and evaluation. Madaus et al. (1983) state that:

> Tyler began by conceptualizing a broadened innovative view of both curriculum and evaluation. This view saw curriculum as a set of broadly planned school-experiences designed and implemented to help students achieve behavioral outcomes (pp. 8-9).
> Evaluation was conceptualized by Tyler as a comparison of intended outcomes with actual outcomes (p. 9).

Tyler influenced thought on educational objectives and their place in programs. Madaus et al. (1983) refer to the time period 1930-1945 as the "Tylerian Age" and they give a great deal of credit to Tyler for initiating new thinking in educational evaluation research. Both definition and purpose needed to be reexamined, so that Stake's (1967) observation that "to the early evaluators, and to many others, the countenance of evaluation has been nothing more than the administration and normative interpretation of achievement tests" (pp. 525-526) would no longer ring true. The reexamination of the field of evaluation research occurred particularly in the area of program evaluation.

Program Evaluation

From the late 1950's on an emphasis was given to examining program evaluation. Tyler, with his discussions concerning objectives, provided a train of thought for researchers to examine and the American government provided the necessary funds for the task.

With the launching of the Russian Sputnik I, in 1957, came the pouring of federal money into educational programs. Federal aid was made available through the Title III program (Stufflebeam, 1967, p. 126). However, in 1958 came the National Defense Act which demanded that federally funded programs be evaluated (Madaus et al., 1983, p. 12).

With the call for evaluation came the realization that evaluators were ill-prepared to take on such a task. Stufflebeam (1967) expressed his annoyance and disillusionment with the practice of educational evaluation, especially in terms of what little had gone on, and he suggested that:

> An 'unwilling Title III evaluator' is faced with a specific set of guidelines which provides no specific

> guidance for planning and conducting the needed evaluations (p. 127).

Evaluators did not know how to conduct themselves in their work and did not have the means of finding out. The field of educational research seemed to have very little to offer in the way of constructive advice. Madaus et al. (1983) state:

> As a result of growing disquiet with evaluation efforts and with the consistent negative findings, the professional honorary fraternity, Phi Delta Kappa set up a National Study Committee on Evaluation (P.D.K. 1917). After surveying the scene, this committee concluded that educational evaluation was "seized with a great illness"; and called for the development of new theories and methods of evaluation as well as new training programs for evaluators (p. 14).

Madaus et al. (1983) refer to 1958-1972 as the "Age of Expansion". During this time researchers in educational evaluation attempted to provide some "new theories and methods of evaluation". Numerous evaluation models were developed during this period. As new models were appearing in the literature, it was becoming evident that the influence of the scientific paradigm on the theories of educational researchers was beginning to wane. Questions were being asked about why educational evaluation should be developed exclusively within that framework. Data from evaluation projects were not being utilized and the exclusive use of scientific methods was one element being blamed for this lack of use.

Some researchers examined other possible methods of evaluation. Many espoused the benefits of using a naturalistic-responsive type of evaluation in the area of education.

> It appears that the descriptive methods of the historian or anthropologist and the case study method of the psychiatrist are more appropriate to the task of educational program evaluation than the

> experimental methods of the psychologist or biologist (Provus, 1971, p. iii).

There appeared to be the beginnings of a major shift of thought operating more out of a naturalist paradigm rather than the dominant scientific one. This shift encouraged the use of qualitative data such as was described by Patton (1980):

> Qualitative data consist of detailed descriptions of situations, events, people, interactions and observed behaviors; direct quotations from people about their experiences, attitudes, beliefs, and thoughts, and excerpts or entire passages from documents, correspondence, records, and case histories. The detailed descriptions, direct quotations, and case documentation of qualitative measurement are raw data from the empirical world (p. 22).

This "raw data" needed to be collected by working from within the natural environment:

> Educational evaluation does not need the antiseptic world of the laboratory but the septic world of the classroom and the school in order to provide useful data (Guba, 1969, p. 34).

With the new models came the broadening of the definition of evaluation. Each researcher provided thoughts on what components were necessary to conduct an evaluation.

Robert Stake (1967) suggested that both description and judgement were essential in evaluation activities (p. 525). He stated:

> It is a great misfortune that the best trained evaluators have been looking at education with a microscope rather than a panoramic viewfinder (p. 536).

He attempted to broaden the scope of evaluation with the suggested emphasis on description and judgement but still kept closely aligned to scientific methods with his suggested countenance model. In the following quote he speaks of the essentials necessary for a "formal" evaluation:

> Formal evaluation of education is recognized by its dependence on check-lists, structured visitations by peers, controlled comparisons, and standardized testing of students. Some of these techniques have long histories of successful use. Unfortunately, when planning an evaluation few educators consider even these four (p. 523).

He suggested what should be included in an evaluation with emphasis on "standardized testing" and "controlled comparisons". The influence of the test and measurement movement is quite clear. Guba and Lincoln in Effective Evaluation (1981) state that:

> He [Stake] continued with an emphasis on formal evaluation, and this emphasis tied evaluation even more closely to the scientific paradigm and its attendant measurement processes (p. 14).

The countenance model was closely linked to scientific methods; however, with the passage of time Stake appeared to become less adamant about maintaining these close ties. He coined the term "responsive evaluation" which dictated the use of qualitative data collected in the natural environment. He stated that:

> The responsive evaluator lets the action of the program stimulate evaluative responses. These he collects and works into some form of illuminative narration or case study, which audiences can interpret for themselves. So the responsive evaluator is guided largely by the particular situation (p. 34).

Observation and feedback are important throughout the evaluation, these making possible alterations in design and methods in response to the audience's needs. There is a major shift of emphasis from Stake's scientifically oriented countenance model to his suggested "responsive" naturalistic paradigm and the methods this paradigm dictates.

Other models appearing in the research have stressed the importance of evaluation in providing decision-makers with relevant information. Stufflebeam

(1973) defined evaluation as "the process of delineating, obtaining and providing useful information for judging decision alternatives" (p. 128). He continued on to suggest keeping the following points in mind when one considers evaluation:

1. Evaluation is performed in the service of decision-making, hence, it should provide information which is useful for decision-makers.
2. Evaluation is a cyclic, continuing process, and, therefore, must be implemented through a systematic program.
3. The evaluation process includes the three main steps of delineating, obtaining, and providing. These steps provide the basis for a methodology of evaluation.
4. The delineating and providing steps in the evaluation process are interface activities requiring collaboration between evaluator and decision-maker, while the obtaining step is largely a technical activity which is executed mainly by the evaluator (p. 129).

Stufflebeam's Context-Input-Process-Product model broadens the scope of evaluation. He emphasizes the importance of evaluators working hand in hand with the decision-makers - "collaboration" is essential. This collaboration aids in the collection of data, both qualitative and quantitative, which is then used by the decision-maker. Evaluation is described as a cyclical process that undergoes change when necessary. This is quite a variance from Tyler's more linear model where measurement of the attainment of objectives is at the core of evaluation.

Similarities can be seen in the models offered by Stufflebeam and Provus. In the late 1960's Provus introduced his elaborate Discrepancy Model for Program Evaluation. The five stages delineated in the model were Design, Installation, Process, Product, Cost and "at each of these stages a

comparison is made between reality and some standard or standards. The comparison often shows differences between standard and reality; the difference is called discrepancy". (Provus, 1971, p. 46). Provus emphasized the need for flexibility in evaluation models. Continuous feedback and essential changes being made to the evaluation framework were to be both possible and welcomed.

> One of the strengths of this model is that it provides for midcourse correction, so that when we educational navigators find that we are a degree or two off course we can correct ourselves and be more likely to arrive at our destination (Carroll, 1971, p. 6).

Evaluation was thus considered a process, but a process necessary for course improvement as well as assessment of programs.

> Both administrators and researchers must see evaluation as a continuous information management process serving program improvement as well as program assessment purposes (Provus, 1971, p. 207).

Another dimension was being added to the evaluation process, that of program improvement.

The Discrepancy Model was applauded for having the essentials necessary for a good working model. Stufflebeam (1971) states:

> Generally, I think that evaluation designs should encompass the delineation of the information to be collected, the means for collecting that information, and the means for helping decision makers to use that information. I think that model has done an excellent job of covering these steps (p. 104).

There was a lot of positive reaction to the introduction of the Discrepancy Model which portrayed the evaluator as both a scientific and naturalistic inquirer. Provus was very outspoken about his reservations concerning the use of an experimental design at the outset of an evaluation and his belief in the resulting detriment to the program due to this use. Provus (1971) states:

> An evaluation that begins with an experimental design denies to program staff what it needs most: information that can be used to make judgements about the program while it is in its dynamic stages of growth. Furthermore, the imposition of an experimental design in the formulative stage of a program inhibits the staff's natural desire to improve a program on the basis of experience. Evaluation must provide administrators and program staff with the information they need and the freedom to act on the information (p. 11).

Provus stresses not using an experimental design during the initial stages of an evaluation, but he does see a place for its use in the latter stages (Stage 3 and 4). Stufflebeam (1971) suggests that this is a strength of Provus' model:

> Traditional research design has been placed in its proper perspective. Many of us have criticized the use of experimental design for a long time, mainly because we thought it inappropriate to employ experimental design in the early stages of the program. . . . I am glad to see the notion of experimental design coming in later after a program has been stabilized (p. 104).

The influence of the naturalistic paradigm on Provus' model is especially noticeable during the initial stages of the evaluation. During the Design stage, qualitative methods are used to collect pertinent data and all levels of the program staff members set the design and the design criteria (Provus, 1971, p. 46). However, fault was found with this aspect of the model because of its dependence on members involved being able to reach a consensus. Donald Carroll (1971) presented this as a concern:

> I am, however, concerned about the program design stage of the model and this idea of achieving a consensus in a design. People can hold very strong feelings about curriculum and I think this is one place where we are at times guided by gut level feelings. I wonder if we can achieve consensus called for in the early stage of a design (p. 126).

This is a point which needs to be examined; however, its examination should be done within the context of individual evaluation projects. Provus has provided the practitioner with a viable model for possible use.

Along with Provus, Michael Scriven was also involved in this expansionist period of program evaluation. He closely examined the place of goals in an evaluation and came to the understanding that, being goal-oriented was detrimental to the evaluation activity.

> It seemed to me [Scriven], in short, that consideration and evaluation of goals was an unnecessary but also contaminating step. I began to work on an alternative approach - simply, the evaluation of actual effects against (typically) a profile of demonstrated needs in this region of education. . . . I call this goal-free evaluation (Scriven, 1974, p. 35).

Scriven viewed the stressing of goals as just another obstacle in the way of providing a panoramic view. Working with goals narrowed the possible scope of the evaluation. Scriven was similar to Provus in his belief of the necessity of the possibility for change in a model. The evaluator must be responsive to the audiences/clients and make use of qualitative methods in the gathering of data. This goal-free model was seen as a major break from "traditional" models and methods.

Eisner (1983), in his presentation and discussion of educational connoisseurship and criticism, provides a unique view of the role of educational evaluation. He, like other writers of this time period, felt that evaluation practice was far too entrenched in the ideology of the earlier testing and measurement research and theory. He suggested that what was of more importance was the artistry of the teacher and the place of theory in its development. Eisner (1983) states:

> Therefore, what I believe we need to do with respect to educational evaluation is not to seek

> recipes to control and measure practice, but rather to enhance whatever artistry the teacher can achieve. Theory plays a role in the cultivation of artistry but its role is not prescriptive; it is diagnostic (p. 339).

The enhancement of teacher artistry is a new addition to the purposes of evaluation already recorded in the literature. Eisner (1983) stresses the importance of artistry in the educational process and the importance of the teacher becoming a connoisseur. He states:

> I do not believe that education as a process is likely to be controlled by a set of laws that can be transformed into a prescription or recipe for teaching. I do not believe that we will ever have a "Betty Crocker" theory of education. Teaching is an activity that requires artistry; schooling itself is a cultural artifact; and education is a process whose features may differ from individual to individual, context to context (p. 339).

The teacher as a connoisseur will develop an appreciation for the experiences of classroom life:

> The major contribution of evaluation is to be a heightened awareness of the qualities of that life so that teachers and students can become more intelligent within it (Eisner, 1983, p. 339).

Eisner's educational connoisseurship has a partner: "educational criticism". It appears that Eisner, like many others concerned with educational evaluation in the 1960's, sees the importance of description, interpretation and judgement in the world of education. Eisner speaks in terms of the educational critic doing these jobs while others such as Scriven (1974), Stufflebeam (1971), Provus (1971) have suggested the above three activities as the job of the program evaluator. Eisner (1983) suggests that:

> The task of the critic is not simply one of being a natural observer (an impossible position in any case), nor is it one of disinterested interpretation. The critic uses what he or she sees and interprets it in order to arrive at some conclusions about the

> character of educational practice and its improvement (p. 344).

The job of educational criticism is to describe, interpret and judge, and by so doing it provides necessary information to decision-makers (Eisner, 1983). This providing of information to decision-makers is an aspect of Eisner's theory that coincides with vocalized thoughts on educational evaluation for that time. However, generally, Eisner's thoughts on educational connoisseurship and criticism take the reader further away from the protection provided by tunnel vision - or as Stake (1967) suggests a "microscopic view" of evaluation - into the vast lands provided by a panoramic view. Eisner believes that educational connoisseurship and criticism have tremendous potential for aiding not only the development of educational evaluation but also the improvement of the process of education. His views are not emerging from a scientific tradition but from an artistic one (Eisner, 1983, p. 341). His concepts of connoisseurship and criticism have provided the reader of educational evaluation research with new ways of looking at evaluation.

From the late 1970's on there was an end to the influx of numerous descriptions of evaluation models appearing in the literature. There appeared to be a settling down of the ripples caused by the injection of government funds into the area of program evaluation and pressures exerted by the public's cry for accountability. Writings in the area of educational evaluation continued; however, the emphasis was more on examining the make-up and purposes of evaluation. The utility of educational research was being questioned as well as the lack of communication between the researcher and the evaluator in the field.

Guba and Lincoln (1981) in the 1980's emphasize the "responsive" evaluator, as did Stake originally in the 1960's. They espouse the importance of using

a naturalistic design and of collecting qualitative data. Their definition of evaluation included its use for providing pertinent information to decision-makers, but also included was its purpose in making possible the judging of whatever was being evaluated. "We define evaluation as a process for describing an evaluand and judging its merit and worth" (Guba and Lincoln, 1981, p.35). The organizer for the responsive evaluation as seen by Guba and Lincoln (1981) was "the concerns and issues of stakeholding audiences" (p.23). In order to serve these audiences, evaluation must be a continuous and interactive process put to work within the context of a naturalistic paradigm. A responsive evaluation is stifled when it is set in a scientific design:

> Preordinate designs are completed at the beginning of an evaluation; and, indeed, it is a major setback if they have to be changed in midstream. In contrast, responsive designs are continuously evolving and never complete; here it is a major setback if the evaluator does not actively work at continuous design change as a result of his ever-growing knowledge and insights (Guba and Lincoln, 1981, p. 30).

This ever-growing knowledge and insight is gained by the evaluator's becoming submerged in the natural setting. Within this natural setting the evaluator uses "methods that are subjective and qualitative rather than quantitative; for example, observations and interviews" (Guba and Lincoln, 1981, p. 31).

Communication between evaluator and audience(s) is an important aspect of Guba and Lincoln's responsive evaluation. The validity of an evaluation is dependent upon it.

> For the responsive evaluator, communication with his audiences is of the essence, for the most meaningful test of the validity of an evaluation is that it improves the audience's understanding of the evaluand (Guba and Lincoln, 1981, p. 32).

The need for scientific "hard" data to complete a successful evaluation is no longer the unquestioned truth.

Other writers in the 1970's and 1980's complemented Guba's and Lincoln's work. Patton (1981) coined the term "creative evaluation" which placed great emphasis on what he called "situational responsiveness". Each evaluation is a unique situation and it is important to consider it as such. Patton (1981) stated that:

> Creative evaluation involves situational responsiveness, methodological flexibility, conscious matching of evaluation approaches to the needs and interests of those with whom we are working and sensitivity to the unique constraints and possibilities of particular circumstances (p. 28).

Patton (1981) puts forth an argument for a paradigm of choices. Within these choices the evaluator is provided with the possibility of methodological flexibility. Patton presents an argument for not siding for any particular method but rather to make use of whichever methods are appropriate for the individual evaluation situations. The possibility of choice of methods is essential to Patton's "creative" evaluator. Both quantitative and qualitative data have a place if the situation calls for them.

> In creative evaluation a paradigm of choices, which recognizes a broad range of viable research methods and approaches, replaces unquestioning adherence to a prescribed set of disciplinary procedures that have the status and effect of being a community paradigm of acceptable methods and ideas (Patton, 1981, p.28).

Not only is choice of method important to creative evaluation, but also important is the personality of the evaluator. Patton has some concerns about the "humanness" of the evaluator. Below he describes what may happen to individuals who become involved in new situations:

> Basically, what appears to happen is that upon entering a new situation, we make sense out of that situation (that is categorize and label) by focusing on those aspects of the situation that are most familiar to us and those elements of the situation that are most similar to our previous experiences. We thus force the new problem or situation to be "representative" of things we already know, selectively ignoring information and evidence that is unfamiliar or does not fit our stereotypes developed through past experiences (Patton, 1981, p. 33).

The probability of selectively ignoring information and evidence that is unfamiliar or does not fit a stereotype can be detrimental to any evaluation.

Patton (1981) believes there is a way of dealing with this very "human" problem. The main defense evaluators have is to become aware of their "heuristics"; a term referring to "all noncomprehensive decision strategies, such as rule of thumb, standard operating procedures, tricks of the trade, and in some respects even scientific paradigms" (p. 31). Evaluators need to be more aware of why they see the world as they do.

> Man looks at his world through transparent patterns or templets which he creates and then attempts to fit over the realities of which the world is composed. . . . Let us give the name constructs to these patterns that are tentatively tried on for size (Kelly, 1955, pp. 8-9).

Evaluators need to be aware of the influence of their "personal constructs" on the way they behave. Being aware of what affects ways of thinking can open up other avenues of thought for evaluators. This awareness must be present from the early beginnings of the evaluation activity.

> Throughout the consultative process, creative evaluators work to maintain an awareness of how their own routine heuristics, academic training, and preconceived notions may be limiting their vision and narrowing possibilities prematurely. Thus, creative evaluators are keen observers of their own actions and influences as part of the ongoing process of observing and studying the entire evaluation endeavor (Patton, 1981, p. 67).

In their writings, Patton (1981) and Guba and Lincoln (1981) have moved away from presenting "a model" as the answer to problems in the world of evaluation. Instead, they have examined closely evaluation practices and the person executing them, the evaluator. They suggest that in order for an evaluation to be of use the evaluator must be both responsive and creative. An evaluation must not be a set plan where change is considered negative, nor should the "stakeholders" be excluded from the initial planning of the evaluation activity. It is essential that communication lines between all parties be kept open. Evaluation is an "active-reactive-adaptive" process, with each evaluation being treated as unique.

The future of program evaluation remains questionable. However, at this point in time it is clear that there has been a pronounced shift of emphasis from a dominant scientific paradigm of thought with its call for quantitative data to what Patton (1981) refers to as a paradigm of choice, in which both qualitative and quantitative data and naturalistic and scientific methods have a place. The role of the evaluator has changed from a dictator to a responsive enquirer. The groundwork has been laid for fruitfulness in future decades.

Student Evaluation

Gathering information on the students they teach has always been a part of teachers' routines. The gathering of this data was, and still is referred to as "evaluation". However, as with program evaluation, definitions and theories of student evaluation have changed over time.

The development of the concept of student evaluation was very much affected by the philosophy and practices of the test and measurement movement. Comparing students and providing reports on their progress through the use

of quantitative data such as teacher-made tests was a sign of a graded system which emphasized the value of end products over the actual processes experienced in creating those products. Hayman and Napier (1975) suggest that:

> A traditional evaluation mentality views education as a series of beginnings and endings in which major outcomes of school are somehow quantifiable and packaged in the form of grades, reports, or other product measures (p. 7).

The purpose of evaluation was seen as the gathering of information to produce these "grades, reports, or other product measures", and this fostered a very narrow view of evaluation.

> Many classroom teachers have a restricted view of evaluation because it has so often been associated with the giving and receiving of grades (Hayman and Napier, 1975, p. 81).

The public cry for accountability brought with it an excessive use of standardized and teacher-made tests to authenticate teachers' numerical judgements of students' work. Essentially, evaluation developed into a procedure for testing:

> While acknowledging the contributions of our predecessors, we must also recognize the limitations of past evaluative practices, many of which were based on the idea that learning consists primarily in the acquisition of facts. Evaluation was therefore limited to paper and pencil tests (Department of Education, 1985, p. 1).

The limited techniques used in evaluation were reflective of the prevalent narrow definition of evaluation as a reporting and grading system for students' "acquisitions of facts".

Writings on evaluation appearing in the literature over the past forty years describe many changes in definitions and theories in the field of evaluation. Similar to the separation made between testing and evaluation apparent in the literature beginning with the "Tylerian Age", the literature

now separates grading/reporting and evaluation. Evaluation is no longer portrayed as solely a testing and reporting procedure but rather as a very complicated process taking time and organization. The educational authority of Newfoundland (1985) states that evaluation is:

> the accumulation of comprehensive evidence concerning the abilities, strengths, and problems of children. Evaluation is more inclusive than measurement (p. 1).

The purposes of evaluation have been extended to encompass diagnosis and remediation for individual students. Stufflebeam, at the beginning of the book Educational Evaluation and Decision-Making (1971), provides readers with this thought: "The Purpose of Evaluation is Not to Prove But to Improve". Evaluation is concerned with the ongoing assessment of students with the aim of improving both student achievement and teachers' teaching methods. The Newfoundland Department of Education (1985) lists the following purposes for evaluation in the language-arts area:

> Evaluation, then, has three purposes: (1) to determine what needs to be taught to the class as a whole and to individuals within the class (on a one-to-one or group basis); (2) to provide both a personal focus for the individualized discussion of student work and the opportunity to encourage and promote progress through such conferences; as well as to provide the opportunity to make both critical analysis of student work and suggestions for improvement; (3) to record and report progress to parents and to the administration on a regular basis (p. 5).

Reporting/grading is now considered one of the purposes of evaluation, but not the only one or the most important one. To conduct evaluation as suggested by the educational authorities of the province, the evaluator would need to be very similar to the individual described by Patton (1981) as the "active-reactive-adaptive" and "creative" evaluator.

Much has been written in the literature concerning the purposes of evaluation and the appropriate methods to be used during evaluation activities. In the literature, evaluation has progressed and become disentangled from the confusion caused by its identification with the test and measurement movement. However, in practice, this does not appear to be true. This writer agrees with the conclusion reached by John Goodlad and Klein (1974) in a study of 67 American schools.

> One conclusion stands out clearly: many changes we have believed to be taking place in schooling have not been getting into classrooms; changes widely recommended for the schools over the past 15 years were blunted on school and classroom door (p. 97).

This writer would change their modest estimate of 15 years to a more reasonable figure of 50 years.

In the world of evaluation, why don't theory and practice gel? Goodlad and Klein's following comment could be considered somewhat of an answer to the above question.

> Perhaps the most telling observation about our educational system is that there is not, below the level of intense criticism and endless recommendations for improvement, any effective structure by means of which countervailing ideas and models may be pumped in and developed to becoming real alternatives (p. 100).

This lack of "effective structure" described by Goodlad and Klein is one answer for lack of change in teacher practices in student evaluation. However, this writer believes there is another, more powerful force, responsible for the fifty-year lag between theory and practice; this being what Crocker (1984) labels as teachers' "functional paradigms". A teacher's functional paradigm is described by Crocker as "a particular generalized way of thinking" out of which teachers operate.

In regard to student evaluation, teachers' "particular generalized ways of thinking" have developed without the assistance of a strong theoretical background. Teacher education does not provide for the necessary training. It is quite possible to complete a degree in education without taking any course whose main content deals with evaluation as described in the literature of recent years. Teachers are working from a base of past personal experience as students in the school system or from experience gained through their own teaching activities. They are operating under the influence of their "heuristics" described by Patton (1981) as "rules of thumb" or "tricks of the trade" or "scientific paradigms". Generally, teachers may well be unaware of the heuristics which influence their behavior. Testing and grading/reporting have been the traditional concept of evaluation and therefore are still often accepted as such. One of the main conclusions of Goodlad and Klein's (1974) study involving 67 United States schools was the following:

> The schools and classrooms of our sample, with very few exceptions, were committed in actual functioning if not in intent, to graded expectations, graded standards, graded norms and the characteristics of curriculum, materials and instruction that normally accompany the well-established traditional graded school (p. 86).

It appears that educational research has done little to affect actual practice in evaluation. Guba (1967) suggests that in education:

> The assumption was blithely made that educational research, once published, would by some mysterious process be turned into a practical teaching method or new curriculum (p. 61).

Goodlad and Klein's "effective structure" for putting theory into practice has not been developed, but the tremendous need for it is apparent if there are to be changes made in how teachers evaluate their students.

Conclusion

With the realization that there were problems in the approaches taken to program evaluation, it was apparent that research in the area was lacking and that this void needed to be filled. The response was an attempt by researchers to fill the void.

Much of what has been suggested under the name of program evaluation can quite easily be translated to fill the needs of teachers involved in student evaluation. However, before use can be made of theories and strategies suggested, such as Guba's "responsive evaluation" or Patton's "active-reactive-adaptive" process of evaluation, there needs to be a major shift in teachers' "functional paradigms". Teachers must question themselves on why they act as they do and examine the "heuristics" which affect the way in which they act. A shift in practice is essential. Teachers as evaluators need to move from assessing students with the exclusive use of quantitative methods, in order to produce marks, to evaluating students through an ongoing process which respects the use of qualitative methods for the purpose of assessing needs and improving education for individual students.

This writer agrees with Jackson (1968) in his statement:

> We [teachers] must be prepared and willing to give up many of our comfortable beliefs about what classroom life is all about" (p. 176).

The time for educators to partake in a "meta-evaluation" is overdue.

CHAPTER III

SURVEY OF GRADE FOUR LANGUAGE ARTS TEACHERS' EVALUATION PRACTICES

The literature is inundated with theories, models, and strategies in the area of evaluation. The Newfoundland Department of Education and school district personnel have developed policy and suggested avenues open to classroom teachers in their evaluations of students. However, in order to examine more closely what may actually be occurring at the classroom level, a survey was conducted with grade four teachers in the province.

The writer chose to use a survey method in order to include a large number of teachers in the sample and to gather as much qualitative data as possible. This chapter includes an explanation of how the survey was completed. Both the instrument and sampling procedure are described in some detail. A data analysis section is provided. Discussion of results and a summary are presented in the concluding section.

Instrument

The curriculum area of interest in the study has been that of language arts. An instrument was developed for use with a sample of grade four language arts teachers. The questionnaire (see Appendix A) was comprised of a list of activities which could be used in a classroom situation to measure student growth in the language arts. The list was compiled by gathering information from the Department of Education's booklet Evaluating Language Arts Performance in the Elementary School, from suggestions presented in Nelson's Teacher's Resource Book A which accompanies the province's grade four language arts program, and from informal conversations with teachers about strategies presently in use in their classrooms. The activities were randomly listed and teachers were requested to consider: (1) the frequency

of use of the activities listed for evaluation, (2) the importance placed on the activity for ongoing evaluation, and (3) the importance of the activity in the assigning of final grades.

Description of Sample and Sampling Procedure

The questionnaire was administered to a sample of the province's grade four teachers. In the selection of teachers the writer used the Department of Education's 1985-86 statement of enrollment of students by district, school, and grade. Teachers of multi-graded classes were not included in the study. All school boards in the province were contacted except those with fewer than 50 grade four students.

The total enrollment of grade four students in each district was examined and the number of grade four classes was calculated by using a formula of twenty-five students per class to reflect the potential number of grade four classes. Each of the school districts received questionnaires for at least half of the possible grade four teachers.

The schools to be involved in the study were selected. The only criteria used was that they have at least twenty grade four students. This was to ensure that multi-graded classes were not included. The writer used the alphabetical listing of schools per district compiled by the Department of Education to assign questionnaires to grade four teachers in the schools. She attempted, where possible, to choose one teacher per school. However, this proved difficult in small districts where the majority of elementary students attended the same school. In cases where there were more grade four teachers than questionnaires designated for the school, the teacher whose surname came first alphabetically was given the questionnaire.

To help ensure an adequate return of questionnaires, the researcher enlisted the aid of district personnel. An initial telephone call was made to each district's language arts coordinator and each was given a description of the study and asked for assistance in the administering of the questionnaires. Coordinators agreed to give the questionnaires to the appropriate teachers in the schools designated by the researcher. In a case where a multi-graded classroom was designated, coordinators were requested to choose the next suitable school listed.

After the writer was assured of cooperation, the questionnaires were sent to the coordinators previously contacted at each school board. They received the necessary number of questionnaires, a covering letter designating schools, a time line for the return of the questionnaires, and a postage-paid return envelope. A second telephone call was made to those coordinators whose questionnaires had not been returned by the time of the deadline assigned.

A total of one hundred ninety-six questionnaires were sent to the province's school districts (approximately half the number of grade four teachers in the province), and one hundred sixty-three (83 percent) were returned.

Data Analysis

This section provides a description of results of the survey. Tables are included to illustrate findings. The reader will note that percentages recorded on the tables may not always sum to a precise 100%. The investigator acknowledges this consequence which is a result of the process of rounding off decimal places for subsets.

Demographic information, such as gender, academic qualifications, and teaching experience, is reported in the first section.

The second section deals with the subjects' responses to the item concerning the frequency of use of the activities listed for student evaluation.

The level of importance placed upon these activities in regard to ongoing evaluation or to the assigning of final grades, by subject, is reported in this section. Some of the tables provided (see Tables VI and VII) have combined response categories in the hope of presenting a clearer picture of results.

Trends and responses are noted and discussed in the final sections.

Results - Demographic Information

The first page of the questionnaire was designed to elicit demographic information on the subjects. Tables I and II indicate that the majority of respondents (78 percent) were female. Seventy-six percent of the respondents had completed elementary methods courses.

TABLE I

GENDER

Percentages	Subjects
20%	Male
78%	Female
2%	No response

TABLE II

PROGRAM OF STUDIES

Percentages	Program
14%	Primary
76%	Elementary
7%	High School
2%	Primary and elementary
1%	No response

Many of the subjects had been highly trained academically. Ninety-three percent of teachers had obtained at least their fourth grade teaching certificate. Twenty-seven percent had their sixth grade or above. Ninety-one percent had had at least four years of university training (see Table III below).

TABLE III

ACADEMIC QUALIFICATIONS

Years of University Training	Percentage of Subjects	Teaching Certificate	Percentage of Subjects
7+	2%		
7	4%	7	6%
6	22%	6	21%
5	45%	5	47%
4	18%	4	19%
3	3%	3	3%
2	5%	2	4%

To complement their training, many of the subjects had had numerous years of teaching experience.

Table IV indicates that 87 percent of the subjects who returned the questionnaire had had more than ten years teaching experience and that 96 percent of the teachers had had at least eight years. No respondent had had less than four years of experience. All of the subjects had had previous exposure to grade four students. Forty-one percent had taught grade four students for more than eight years and 90 percent had had at least two years of teaching at that level.

TABLE IV

SUBJECTS' TEACHING EXPERIENCE

Total Number of Years Teaching Experience	Percentage of Subjects	Total Number of Years Teaching Grade Four	Percentage of Subjects
10+	87%		
10	4%		
9	2%	8+	41%
8	4%	8	6%
7		7	3%
6		6	7%
5	2%	5	10%
4	1%	4	10%
3		3	8%
2		2	5%
1		1	10%

In contrast to the teachers' wealth of experience was their lack of training in the area of evaluation. Table V below indicates responses to the item concerning how many courses the teachers had completed in the area of testing and evaluation.

TABLE V

EVALUATION AND TESTING COURSES COMPLETED

Evaluation and Testing Courses Completed	Percentage of Subjects
3+	7%
3	11%
2	25%
1	25%
0	18%
No response	14%

Very few of the subjects had completed three or more courses. One-half of the subjects who responded had had only one course or no course completed. Fourteen percent did not respond. One possible reason for the high no-response return could be confusion as to what constitutes a course in evaluation.

Data Gathered on Frequency of Use of Activities for Evaluation

On pages two and four of the questionnaire (see Appendix B), subjects provided information concerning the frequency of use of activities for evaluation. Table VI lists those activities used almost always or frequently by at least 50 percent of the respondents.

It would seem that some of the suggestions of the Department of Education and the Nelson program are being used by a majority of teachers. The high percentage of use of the writing folder and the student profile is indicative of this. However, the tendency to rely on the use of pen and paper tasks is quite evident. The majority almost always or frequently use spelling tests, end of term tests, and teacher-made quizzes. Even commercial

workbooks are used almost always or frequently by more than one-half of the teachers.

TABLE VI

ACTIVITIES FREQUENTLY OR ALMOST ALWAYS USED

Activity	Percentage of Subjects Frequently/Almost Always Use Activity
Spelling test	70%
Teacher-made cloze test	51%
Test (end of term)	65%
Writing folder	90%
Commercial workbooks	57%
Quiz (teacher-made test on a block of content)	61%
Student profile	84%
Teacher questioning (oral or written)	78%
Teacher observation during directed reading activities ...	75%

Table VII lists the activities which at least 50 percent of subjects report as having been used hardly ever or only occasionally.

The activities least used appear to be those that are more oriented towards the individual student rather than toward the class as a whole. Sixty-one percent hardly ever or occasionally use the oral presentation as an activity for evaluation. Further, the activities that depend upon a student's initiative are also in this category. The majority of subjects hardly ever use the reading contract with the student and 86 percent hardly ever or occasionally use a student self-assessment checklist for their evaluation.

TABLE VII

ACTIVITIES HARDLY EVER OR OCCASIONALLY USED

Activity	Percentage of Subjects Hardly Ever or Occasionally Use Activity
Oral presentation by student	61%
Teacher-made observation checklist	62%
Reading contract made with student	82%
Listening activity using audio-tapes	58%
Miscue inventories - oral reading tests	78%
Student self-assessment checklist	86%
Writing conference	53%
Teacher-made worksheets	66%
Group project	61%
Anecdotal records	60%
Audio-tapes used to record reading experiences	91%
Standardized reading tests	87%
Dramatizations of readings	71%

Activities that depend on teachers' writing or recording of events appear to be used infrequently. Teachers hardly ever or occasionally use anecdotal records and many infrequently, if at all, use teacher-made observation checklists.

As well as being asked to state the frequency of use of the activities listed on the questionnaire, subjects were asked to rate the importance of the activity in ongoing evaluation and in the assigning of final grades (see Appendix B). A scale of three categories - Not Important, Important, Very Important - was provided for the respondents.

The majority of activities were considered, by the subjects, to be important in both ongoing evaluation and in the assigning of grades. However, when one examines the frequency of use of these activities, it is apparent

that even though they are placed in categories of important to very important, they are rarely used. Table VIII lists the activities which fall into this category.

Only 37 percent of respondents use students' oral presentations to the class for evaluation, even though 77 percent believe them to be important in ongoing evaluation and in the assigning of final grades. Listening activities with audio-tapes are used as evaluation activities by only 39 percent of the respondents, even though 67 percent of subjects believe them to be important for use in ongoing evaluation and 61 percent of subjects believe them to be important in final grading. The same inconsistency is apparent in many of the activities listed: 20 percent of subjects use miscue inventories frequently or almost always, even though 59 percent think them important in ongoing evaluation and 50 percent believe them to be important in final grading. It appears that even though the respondents acknowledge the importance of the use of these activities in evaluation and grading, they are using them infrequently, if at all.

There were only four activities on the questionnaire which were considered by more than 50 percent of the respondents as not important for ongoing evaluation or grading (see Table IX).

TABLE VIII

COMPARISON OF FREQUENCY OF USE OF ACTIVITY WITH THE DEGREE OF IMPORTANCE PLACED ON ACTIVITY FOR EVALUATION PURPOSES

Activity	Percentage of Subjects Using Activity Frequently/Almost Always	Importance of Activity in Ongoing Evaluation Percentage of Subjects Responding Very Important/Important	mportance of Activity in the Assigning of Final Grades Percentage of Subjects Responding Important/ Very Important
Oral presentation	37%	77%	62%
Teacher-made observation checklist	33%	60%	60%
Group project	34%	77%	68%
Teacher-made cloze test	51%	78%	78%
Listening activities using audio-tapes	39%	67%	61%
Miscue inventories - oral reading tests	20%	59%	50%
Activities suggested in the Networks program	46%	91%	90%
Writing conference	45%	74%	56%
Teacher-made test or writing convention	46%	72%	71%
Teacher-made worksheets	33%	62%	52%
Anecdotal records	34%	56%	50%
Quiz	61%	80%	80%
Audio-tapes used to record reading experiences of individual students	7%	43%	34%
Standardized reading tests	9%	41%	34%
Dramatizations of readings	29%	59%	42%

TABLE IX

ACTIVITIES CONSIDERED NOT IMPORTANT FOR EVALUATION AND GRADING

Activity	Percentage of Subjects Not Important for Ongoing Evaluation	Percentage of Subjects Not Important for Assigning Final Grades
Reading contract	59%	66%
Student self-assessment checklist	55%	49%
Audio-tapes used to record reading experiences	48%	55%
Standardized reading tests	48%	44%

These activities are also listed on Table VII as ones hardly ever or occasionally used by at least 80 percent of subjects. Teachers are not using them, and they apparently feel the activities are not important for evaluation.

The responses given to the two categories, importance placed on an activity in ongoing evaluation and the importance placed on an activity for the purpose of providing final grades, were very similar. There was never more than a 20 percent difference between the number of subjects who felt that an activity is important for ongoing evaluation and the number of subjects who felt an activity is important for the assigning of final grades. In most cases there was less than a 10 percent difference.

In the majority of cases, if activities were considered important in ongoing evaluation they were considered important in the assigning of final grades. In some instances identical numbers of subjects responded to the two questions in the same way. Eighty percent of subjects believe a quiz is important in ongoing evaluation, and the same numbers of subjects believe a

quiz to be important in the assigning of grades. Ninety-two percent feel the student profile is important for both the assigning of grades and for ongoing evaluation. Anecdotal records are considered important for ongoing evaluation by 56 percent of the subjects, and 50 percent believe them to be important in the assigning of final grades. There appears to be little differentiation made between activities used for ongoing evaluation and those used for the assigning of grades.

Discussion

The purpose of the study was to provide a picture of what is happening at a classroom level in the evaluation of grade four students in the language arts area.

Clearly, the province's schools have a large number of educated and experienced grade four teachers. However, some questions need to be raised concerning the appropriateness of the subjects' training and experience in light of the evaluation tasks required of teachers in their daily routines. Teachers are constantly expected to evaluate a student's performance, but few have had formal training to prepare them to do so. Only 18 percent of the subjects had completed three courses in testing and evaluation, and 43 percent had completed only one course or none at all. Fourteen percent of the subjects provided no response to this particular item on the questionnaire. This percentage of "no response" could be viewed as indicative of the confusion surrounding what constitutes a course in evaluation. When one scans the Calendar of Studies for Memorial University of Newfoundland, which most of the subjects have attended as undergraduates, it is clear that there is a limited offering of what even the most broadminded educator might consider a practical methods course in the evaluation of students. A minimal

number of education courses have an evaluation component; for example, courses such as Reading in the Elementary Grades or Educational Drama and Curriculum. Otherwise, when they responded to this particular item on the questionnaire, subjects were generally referring to courses directed more towards a test and measurement stance, which might include courses such as Tests and Measurement, Psychology Test and Measurement, or Educational Assessment. There are no courses offered whose main concern is the evaluation of students either generally or in specific curriculum areas. Apparently our well-trained teaching force has spent little time studying how to evaluate their students, even though many of their tasks as teachers depend upon their having a sound base in the area.

The activities listed on the questionnaire for use in the evaluation of students in the language arts area were those recommended by the writers of the grade four program Networks and by the province's authority, the Department of Education. The activities are meant to encompass all four major components of the language arts - reading, writing, listening, and speaking.

However, as can be seen by scanning Tables VI and VII, the most frequently used evaluation activities listed are those dealing with traditional tasks of a pen and paper nature used especially for end of term reports. A good example of this would be the number of tests, described by different names such as quizzes, cloze-tests, end of term tests, being used almost always or frequently by 50 percent or more of the respondents.

These various time-consuming tests are administered to measure a student's ability in the reading and writing areas. The importance of testing students is not put to question by the researcher; however, the validity of the type and amount of testing occurring in the name of evaluation in the elementary classroom is.

Through their responses on the questionnaire, teachers have expressed their belief that the activities involving the other components of the language arts curriculum, listening and speaking, activities such as using audio-tapes for dramatizations of readings or oral presentations by students, are important; nevertheless, they do not often make use of them in the evaluation process.

When one compares the frequency of use of activities with the importance placed upon these activities in light of evaluation by the subjects, inconsistencies appear (see Table VIII). Both the Department of Education and the grade four program writers recommend the use of these activities and teachers acknowledge their own belief in the importance of their use, but this does not appear to affect practice at a classroom level.

This writer believes that the low level of expertise of the majority of teachers in the area of evaluation has had a detrimental effect on teachers' evaluation practices. What knowledge has been acquired concerning evaluation has often been limited to a teacher's own experiences as a student in the school system, together with bits of information collected through courses, inservice sessions, and teaching experiences, much of which has been grounded in test and measurement theory and practice. Evaluation for the purpose of diagnosing and remediating and enriching has been left to the realm of the Special Education teacher or left to remedial teachers who have always seen this as part of their jobs. Testing and the assigning of grades has been the main thrust of evaluation for the average classroom teacher.

Compounding the problems developing because of the lack of expertise in the area of evaluation is the time constraint ever-present in the teaching field. The five-hour day is sliced into strips of time supposedly sufficient to cover what needs to be completed in the numerous curriculum areas. With time

being parcelled out, sparingly, the least time-consuming activity may well be the most practical. Many of the activities included on the questionnaire are time consuming. The mere mention of using teacher-made checklists for individual students appears over-whelming to many teachers, as does the time allotment necessary to allow students the opportunity to partake in different forms of oral language activities.

Matched with time constraints is the administrative pressure for producing grades. With the aim of producing grades, in many schools students are expected to complete end of term and end of year tests even though in the language arts area this is not seen as a particularly efficient means to measure student growth. In the area of writing, the Department of Education's guide for teachers, Language Growth, recommends that students be provided with the opportunity of experiencing a complete process of pre-writing, writing, revising, and editing in the developing of an end product before it should be graded. The politics of grading appears to have superseded comprehensive evaluation practices.

Summary

The results of the survey have reinforced the writer's belief that evaluation at a classroom level is still entrenched in outdated and often poorly understood test and measurement theory. The writer has presented possible reasons for this state of affairs.

In the past, attempts to affect change in this area have been minimal. Methods of evaluation have been suggested by authorities in the province and by program writers, but this has accomplished very little. The surface-down approach is insufficient when what is needed for educators out in the field is an examination of the roots of their thinking concerning evaluation.

Individuals need to consider why they evaluate as they do. An examination of their "paradigms of thought" or "personal constructs" is essential if change is to occur.

CHAPTER IV

A RATIONALE FOR INSERVICE

Introduction

The survey described in Chapter III indicates that there is a problem with teachers' methods of evaluating grade four students in the language arts area. The information gathered by using the questionnaire suggests that the teaching workforce consists of a highly trained and experienced group of professionals. However, upon close examination of the results of this survey the reader becomes attuned to the fact that these people are lacking the necessary equipment to complete the task of evaluating their students. Previous teaching experience and academic training of teachers presently in the classrooms is inadequate in providing the means for teachers to accomplish this essential task. Sawin (1969) states that:

> Evaluation is a many-sided process. It is not merely a matter of giving and scoring tests. Evaluation encompasses every usable technique or device available for obtaining the information about students needed for planning and conducting effective instruction (p. 20).

Teachers in their classrooms are having difficulty dealing with this "many-sided process". Administering and scoring pen and paper tasks appears to consume the majority of time alloted for evaluation. This is understandable, considering the background of those teachers involved. The content of their academic training apparently consists of little relevant to the real world of evaluation. Teaching degrees are awarded without education programs having placed emphasis on training individuals to evaluate their students. As well, teachers have had minimal, if any exposure to others in the teaching world who would be considered appropriate models of student evaluators. The dealings they have had with teachers before entering the teaching profession

and during their student lives in many cases is of little help to them as evaluators. In light of their training and experience, the previous educators of present day teachers would likely have been even more entrenched in a test and measurement mind-set.

After entering the profession, teachers work in a very "isolated" environment. Rarely are they observing other teachers working in classroom situations and therefore seldom, if ever, are they exposed to the evaluation strategies of others. Individuals have their students, their classrooms, and their closed doors.

It is evident that teachers already in the workforce need to develop a better understanding of the concept of evaluation and of how to accomplish the task of adequately evaluating their students.

Applications of Program Evalution

Research in the area of student evaluation has been somewhat limited. However, the writer believes that much of what has been reported in the literature concerning program evaluation can be extracted and applied to student evaluation to help teachers acquire a better understanding of the concept and the process. There are similarities between the philosophy of evaluation presented by those concerned with classroom practices and by those involved in the burst of educational research in program evaluation conducted over the past forty years. In the minds of those described above, evaluation has broadened out to encompass more than mere testing and reporting purposes. It is considered a continuous process that is basic to the improvement of life for what Guba and Lincoln (1981) refer to as the "evaluand": the subject of the evaluation. When referring to program evaluation, Provus (1971) states that:

> Both administrators and researchers must see evaluation as a continuous information management process serving program improvement as well as program assessment purposes.

Perrone (1977), in the following quote, also emphasizes the importance of continuous evaluation and its contribution to the growth of programs, teachers and children.

> Most definitely, I do not oppose evaluation; I consider it basic to the growth of programs, teachers, and children. But evaluation needs to be embedded in the classrooms. It needs to be consonant with purpose. Assessing children's growth, for example, is an intense activity, and it should occur daily, continuously. It is integral to everything that goes on in a classroom (p. 10).

Perrone is writing from a classroom perspective of teacher as evaluator, whereas Provus's perspective encompasses the larger domain of major program evaluations oftentimes conducted by hired, trained personnel. However, both stress the importance of evaluation as being a continuous process with a much broader scope than the testing and reporting of results.

The research in program evaluation has aided in broadening the scope of evaluation. It has accomplished this by examining the basic reasoning behind the theories and practices in evaluation. By so doing, research has produced much food for thought for those educators now in classrooms who are attempting to evaluate their students.

In the research, questions have been raised concerning the adherence of many in the field of evaluation to a scientific paradigm of thought. Guba and Lincoln (1981), Patton (1978) and Eisner (1983) question the validity of adhering solely to a mind-set that sees quantitative research as the only reliable methodology.

> At best some social scientists are willing to recognize that qualitative methodology may be useful at an exploratory stage of research prefatory to quantitative

> research. What they deny is that qualitative methodology can be a legitimate source of either data collection, systematic evaluation, or theory construction (Patton, 1978, p. 215).

A naturalistic paradigm of thought is offered as an alternative to the scientific. This framework accepts the validity and essentiality of qualitative data in any evaluation situation. Guba and Lincoln (1981) state that:

> It is our judgement that in the field of behavioral science, of which evaluation is surely a part the naturalistic paradigm should be the paradigm of choice (p. 77).

Patton (1981) extends this concept of choice from decisions made concerning two options to decisions made in light of multiple possibilities. Below he describes the term he coined "creative evaluation":

> Creative evaluation means working within a framework, or paradigm, of multiple possibilities where new situations are approached without preconceptions about which particular methods or approaches ought to be applied. Creative evaluation means being situationally responsive, methodologically flexible, consciously committed to matching evaluation approaches to the needs and interests of those with whom one is working, and genuinely sensitive to the unique constraints and possibilities of particular people and circumstances (p. 67).

Each evaluation situation is unique; therefore, decisions concerning methodology are made in light of this uniqueness.

The writer feels confident in drawing a parallel between what has occurred in the development of program evaluation practices, as reported in the literature, and what has occurred in the development of evaluation practices at the classroom level.

A scientific paradigm of thought has reigned in the classroom. Teachers, administrators and the public, have been concerned with numbers and grades to the detriment of proper evaluation, with numbers being perceived as the

content of science and the signal of validity. Educators who view evaluation as only a test and number activity are operating, for the most part, out of a scientific paradigm of thought.

The alternative paradigm of thought, the naturalistic, appears to be more compatible with the world of education, where classroom teachers work with students of various capabilities and personalities. The data indicative of this paradigm of thought are qualitative, which include information collected through interviews, projects, speeches, observation checklists and other activities. Teachers are able to collect data about their "evaluands" due to their prime position in the natural setting, the classroom.

Even though this alternative paradigm appears to gel with classroom life, this writer is in agreement with Patton (1981) and his suggestion of a third paradigm: one of choice. Both the scientific and naturalistic paradigms of thought have something to offer for evaluators. Use must be made of whatever is necessary for the success of the evaluation. Appropriate choices are essential.

Educational research in program evaluation describes the people, the evaluators, who make these choices. The characteristics of good evaluators described in the literature can be applied to both the program evaluator or the teacher in the classroom. Patton (1981) refers to creative evaluation and the evaluator as a person who is "active-reactive-adaptive" (p. 67).

Evaluators must observe happenings and react to them in an appropriate manner. Where necessary, evaluators need to adapt in response to observations and results of reactions. The same is true of teachers in the classrooms.

> No elementary teacher can be as effective as she might otherwise be unless she possesses considerable skill in assessing how well her children are doing,

> the difficulties they are encountering - and why (Hedges, 1969, p. iv).

After the "why" is answered, the question of how to adapt needs to be proposed.

Researchers have provided labels for evaluators. Guba and Lincoln (1981) refer to the "responsive evaluator" as did Stake as early as 1967. Eisner (1983) addresses the importance of the evaluator being both a critic and connoisseur. Although labels vary, the literature offers a picture of evaluators as individuals who are able to recognize and utilize all options at their disposal. This makes possible the collection of comprehensive data necessary for the evaluation process which includes action, reaction, and adaptation.

The Process of Educational Change

The review of the literature provided information on "what should be", while Chapter III informed us of "what is". However, to get from what is to what should be involves a process of change. Teachers in their classrooms must change their evaluation practices. On the surface this may appear to be a simple solution, but in reality it is a tremendous task. Fullan (1982) provides this insight into educational change:

> The difficulty is that educational change is not a single entity. It is to a certain extent multidimensional. There are at least three components or dimensions at stake in implementing any new program or policy: (1) the possible use of new revised materials (direct instructional resources such as curriculum materials or technologies), (2) the possible use of new teaching approaches (i.e., new teaching strategies or activities), and (3) the possible alterations of beliefs (e.g., pedagogical assumptions and theories underlying particular new policies or programs) (p. 30).

Those individuals attempting to affect change need to be aware of this "multi-dimensionality". Most times it is ignored, and possibly that is why. Brown (1985) suggests:

> Our schools do not reflect what we know about what education should be. In most schools time has practically stood still. Education does not encompass the innovations which have been so widely publicized (p. 231).

A new language arts program has been placed at the doors of our schools. Evaluation philosophy and strategies have been presented in conjunction with this, but whether implementation is successful or not must rely most heavily on the classroom teacher's ability to "act-react-and adapt". There is much more involved than an acceptance of new texts. The way individuals teach and think comes into question and major changes must ensue. Fullan (1982) refers to this in the following quote:

> Change in teaching approach or style in using new materials presents greater difficulty if new skills must be acquired and new ways of conducting instructional activities acquired. Change in beliefs are yet more difficult: they challenge the core values held by individuals regarding the purposes of education; moreover beliefs are often not explicit, discussed, or understood, but rather are buried at the level of unstated assumptions (pp. 34-35).

Grade four teachers need to consider "new ways of conducting instructional activities" and possibly a "change in their beliefs"; the latter being the most difficult. As Brown (1985) suggests:

> The most difficult to implement is a change that requires a change in beliefs. The effects on the individual when core values are challenged can be traumatic. Often teachers' beliefs are not explicit but rather they exist as unstated assumptions. As the studies have shown, individuals have to change their beliefs in order for real change and personal growth to occur (p. 158).

It is apparent that change in practice at a classroom level cannot occur overnight. In this specific case there would appear to be a number of grade four classroom teachers who, on the surface, need to change their methods of evaluating students, while underneath they need to consider their beliefs

about evaluation. For those considering entering the profession, the writer recommends a change at the university level; teacher training programs must include extensive work in the area of student evaluation. However, this is of little use to those teachers already in the workforce who have no intention of returning to the university milieu. Inservice at the school board level is the only realistic option.

Chapter V presents the reader with a plan for coordinators to use while inservicing grade four language arts teachers and the appropriate administrators. It is a plan which might be adapted and used by various school boards in the province.

The writer is aware of limitations of so-called one-shot inservice and recommends that this plan be viewed as a stepping stone for initiating change at the classroom level. The writer agrees with Fullan (1982) in his statement that:

> Most forms of inservice training are not designed to provide the ongoing, interactive, cumulative learning necessary to develop new conceptions, skills, and behavior. Failure to realize that there is a need for in-service work during implementation is a common problem. No matter how much advance staff development occurs, it is when people actually try to implement new approaches that they have the most specific concerns and doubts. It is thus extremely important that people obtain some support at the early stages of attempted implementation (p. 66).

After the three-day inservice, coordinators need to spend time in classrooms to work with teachers during their attempts to examine personal evaluation strategies. Contact is essential for the promotion of change. House (1974) suggests that:

> Schools and teachers are fragmented and decentralized and exist within the same societal structures. The flow of direct personal contact within and between

> educational structures has much to do with the diffusion of educational innovation (p. 15).

As well as time spent by the coordinator in classrooms, communication between teachers is essential. Teachers within schools or districts need to meet periodically to discuss their experiences in classrooms.

Besides being aware that change is unlikely to occur if this is considered just a three-day plan which includes no personal contact during teachers' attempts at change in strategies, those involved in the quest to change educators' ideas and strategies in the area of evaluation must have some perspective of a time-line.

> Any conception of the change process in regards to schools, explicitly or implicitly, involves a time perspective (Sarason, 1971, p. 219).

Sometimes in education this time perspective is ill-considered. Often, goals may be unrealistic when there is minimal understanding of the change process. The greatest disparity may arise between the expectations of coordinators who play the role of change agents and the classroom teachers. Sarason presents these ideas on the topic:

> Let us assume that the agents of change (coordinators) have worked out in a systematic fashion the relationship between their conceptions of the setting and a time perspective by which the intended change should be judged. A second aspect of the time perspective problem then arises: comparing the time perspective of the agents of change with that of those who are the targets, and that of those who will, in one way or another, participate in the process. This comparison is crucial because if, as is usually the case, the differences in time perspective are great, the seeds of conflict and disillusionment are already in the soil (p. 219).

Avoidance of this sowing of "seeds of conflict and disillusionment" is possible if realistic and common goals are sought by all involved in the process.

After emphasizing the use of the plan as an initiator of change in classroom practice and discussing the tremendous complexity of the change process itself, the writer leaves the plan to stand on its own.

Pains were taken to propose a plan which would incorporate the necessary ingredients for success in inservice delivery.

The coordinator's role in the inservice goes beyond disseminating information to providing a workable model for teachers. Strategies used within the plan were chosen for their simplicity and adaptability to classroom life. The evaluation component is emphasized throughout the three-day session to promote its serious consideration. In light of the research available on the difficulty of changing beliefs, the writer incorporates sections in the plan through which beliefs are brought to the forefront and questioned: reflection is a major component.

Summary

The writer is suggesting that it is possible to move from "what is" to "what should be". However, this is not a task to be taken lightly.

In order for teachers to become more proficient evaluators of their students, school boards in the province will have to declare inservice in evaluation a priority.

The three-day inservice plan described in Chapter V is available for use as a stepping stone for affecting change in teachers' evaluation strategies. However, this must be supplemented with personal contact between those involved in the process. Coordinators must work with teachers in their classrooms and time must be provided for teachers, periodically, to meet together.

The complexity of the change process must be understood by those involved. A realistic time-line for the achievement of goals must be agreed upon.

CHAPTER V

AN INSERVICE PLAN

Introduction

In this chapter the writer has attempted to provide a framework for inservicing grade four language arts teachers in the evaluation of students. It is hoped that this will be the initial step in the beginning of a much needed examination of why and how students are evaluated in the school system. The plan is developed for use by language arts coordinators employed in the various districts. However, considering the uniqueness of each district it is essential that the framework be adapted to individual situations. A very conservative time-line is suggested for delivery of this inservice in the hope that implementation will be viewed as both possible and practical.

The writer intends that the inservice plan be viewed as a flexible tool which is to be continuously exposed to the participants' scrutiny. This scrutiny is the backbone of the evaluation process which is continuous throughout all stages of the inservice. Modifications are to be made to the plan when deemed necessary by those involved. The writer believes the flexibility inherent in the inservice is one of the plan's positive characteristics.

Another strength of the plan may be the consideration given to the role of coordinator as a model for teachers. While developing the inservice program, the writer attempted to utilize strategies and materials easily accessible for either inservice situations or classroom instruction. The method of delivery was arranged to provide opportunity for participants to glean an array of activities for possible use in their own classrooms.

In an attempt to present the inservice framework in a clear format, the writer includes information maps and sample time-tables to accompany written descriptions and directions for implementation.

The inservice plan is divided into five sections. The first section provides an inservice design map and discussion of the overall plan. The following four sections provide session maps and discussion of plans for possible delivery of each of the sessions.

Inservice Plan for Grade Four Language Arts Teachers

INSERVICE DESIGN MAP

PARTICIPANTS:	grade four language arts teachers, principals, professional school board personnel
TOPIC:	evaluation of grade four students in the language arts area
Session One:	reflection and discussion of personal philosophies of evaluation
Session Two:	examination of evaluation strategies now in use by participants
Session Three:	dissemination of information on the history of evaluation
Session Four:	discussion and presentation of appropriate strategies for use in student evaluation
GOALS OF THE INSERVICE:	to provide participants with the opportunity to interact with other professionals in the area
	to provide an opportunity for participants to reflect upon and examine personal philosophies of evaluation
	to provide a forum which enables participants to present and to discuss their personal philosophies
	to provide participants with information dealing with the historical background of evaluation

GOALS OF THE INSERVICE: (Cont'd)	to promote a better understanding of evaluation and its connection with testing and grading
	to encourage participants to examine evaluation strategies they now use
	to provide participants with information dealing with appropriate evaluation strategies as suggested by the Department of Education and the grade four program writers
PROCEDURE:	three days of inservice are granted to the participants
	a suitable meeting place is arranged for the sessions
Day I	Session One
	9:00 - 12:00 (15 minute break) 1:00 - 3:00
Day II	Session Two
	9:00 - 10:45
	Session Three
	11:00 - 12:00 1:00 - 3:00
Day III	Session Four
	9:00 - 12:00 1:00 - 3:00
TEACHING STRATEGIES:	Lecturing
	Brainstorming
	Grouping
	Discussing

EVALUATION:	(continuous throughout the inservice)
	records and checklists kept by coordinator
	recommendations or suggestions of participants
	evaluation discussion at the close of each session
	final written evaluation report completed by participants
MATERIALS:	information sheets
	agendas
	folders for participants
	list of goals
	starters for brainstorming
	outlines for essential overheads
	student evaluation strategies forms
	participant profile sheets
	list of statements describing beliefs held concerning evaluation
	true-false forms
	fictional case study of a teacher
	case study discussion starters
	evaluation forms
	anecdotal record sheets
	observation checklist forms

Inservice - Description

The inservice design map provides a framework for the delivery of the inservice described. The map is to be used by coordinators in the planning

stages of the inservice. General information is provided on the inservice map, whereas more specific details are outlined on the individual session maps. Sample agendas are included for all sessions.

The suggested participants for the inservice are grade four language arts teachers, principals and school board personnel. The latter two are included to ensure the involvement of all those with input into the evaluation process. The opportunity for development of a support system for teachers is created when administrators are included in inservice dealing with curriculum matters.

Goals of the inservice are listed on the map. However, these may be modified after discussions with participants. These goals are general statements of expected outcomes from delivery of the inservice.

The way the map is to be used is a decision for individual coordinators. The three-day procedure suggested in this plan was decided upon with a particular school board in mind. Alterations to the plan may be necessary in order to accommodate school boards in varying districts.

The inservice plan suggests the order and content of four sessions. The writer organized the sessions to follow a "bottom up" approach. In the past, many suggestions on student evaluation have been passed down from the Department of Education or school board personnel to the classroom practitioner. However, rarely has this "top down" approach been more than a fruitless endeavor.

The plan suggests the first session evolve around consideration of teachers' personal philosophies of evaluation - the roots of their evaluation practices. Participants are required to examine their ways of thinking about evaluation, their "paradigms of thought", and to question why they think as

they do. The second session is more specific with its examination of actual evaluation activities being used by participants in their classrooms.

Session three provides information concerning the history of evaluation. A distinction is made between testing and evaluation. Recent research in the area of evaluation is presented.

Session four requires that teachers examine their methods of evaluation in light of the information provided in the previous session and in light of the results of the period of reflection offered in the first session. The Department of Education's recommendations for student evaluation in the language arts area are presented.

The writer attempted to develop an inservice plan which coordinators could use with the least amount of difficulty. Therefore, where possible, materials suggested for use in the activities described are provided in the inservice appendices.

SESSION I - DESIGN MAP

PARTICIPANTS:	grade four language arts teachers, principals, professional school board personnel
TOPIC:	personal philosophies of evaluation - reflection and discussion
GOALS OF THE SESSION:	to provide a time for participants to reflect on beliefs held in the area of evaluation to provide an environment to promote reflection on philosophies held to encourage participants to examine how and why they think about evaluation as they do to provide a forum for discussion of beliefs held to create an awareness, on the part of the participants, of the possible limitations and strenghts of the personal philosophies described and discussed
PROCEDURE:	give out materials needed for the session assign participants to their small groups read through the agenda with the large group (* discussing plans for the day) as the day progresses keep an anecdotal record of unusual - interesting occurrences use a checklist to periodically record areas where change is needed in order to improve the inservice

PROCEDURE: (Cont'd)	use small and large group work sessions as suggested by the agenda
TEACHING STRATEGIES:	large group discussions
	small group discussions
	brainstorming
	writing exercises
EVALUATION:	participant input at all stages of the inservice
	anecdotal records and checklist compiled by coordinator
MATERIALS:	list of goals
	agenda
	list of evaluation statements
	case study
	case study discussion starters
	teacher profile forms
	anecdotal record sheet
	evaluation checklist
	participants' folders

AGENDA

9:00 - 9:30	Introduction
9:30 - 10:00	Discussion of Goals of the Inservice
10:00 - 10:45	Small Group Work Using List of Evaluation Statements
10:45 - 11:00	Coffee
11:00 - 12:00	Development of an Acceptable List of Evaluation Statements (Small and Large Group Work)
12:00 - 1:00	Lunch
1:00 - 1:45	Reading and Discussion of a Case Study Concerned With the Description of a Grade Four Language Arts Teacher
1:45 - 2:30	Writing Time - Participants' Development of a Personal Teacher Profile
2:30 - 3:00	Concluding Section

(Total - 4 hours, 45 minutes)

Session I - Description

Housekeeping matters are dealt with at the beginning of the first session. Participants receive folders for the collection of materials used throughout the inservice, and they are assigned to small groups for the duration of the inservice. They will work in these groups of four to six individuals occasionally during the three-day period.

The first session includes an introduction to the whole inservice. The overall goals of the plan are presented and discussed. The coordinator may modify or further delineate these goals in response to participant input. Also, an explanation of the evaluation process is given by the coordinator, who stresses the importance of continuous diagnostic evaluation occurring throughout the three-day period. Emphasis is given to the role of the participant in the evaluation process. Also, participants are made aware of the attempt of the coordinator to vary teaching strategies during the inservice. The importance of participants' being active learners and the use of grouping to accomplish this is stressed.

After the introduction, the first session begins with an examination of goals stated specifically for this session. A connection is made, by the coordinator, between evaluation and achievement of goals.

Group work is initiated during the first session. After examining the goals of the session, participants break into their previously arranged small groups and begin the task of considering a list of statements describing evaluation (see Appendix for Session I).

They are requested to consider the list, to indicate agreement or disagreement with the statements, and to compile a group list of what they

consider to be acceptable descriptions of evaluation. This list is created from provided statements and those proposed by group members.

After the completion of this activity, participants return to a large group setting. The coordinator, with the aid of an overhead projector and the use of a discussion format, records on chart paper a whole group list. Later, copies of the list are filed in the participants' folders for future reference. This activity will provide time for reflection and discussion of differing views held concerning evaluation.

The afternoon session begins with silent reading of a case study describing a grade four language arts teacher. Details in the study include: (1) the teacher's philosophy of evaluation; (2) the evaluation strategies used by the teacher; and (3) a description of the teacher's training and experience in evaluation.

After the initial reading, the small groups gather and discuss the study presented, keeping in mind the list of evaluation statements developed earlier in the session. A list of questions (see Appendix for Session I) is provided by the coordinator to help guide the discussion through an examination of beliefs and practices of the teacher described.

The session moves from an examination of a fictitious character to an examination of self. Participants are asked to reflect upon their own beliefs and methods of evaluation by writing brief profiles of themselves. These personal pieces of writing are filed in the participants' folders.

The last half hour of the afternoon is used for concluding remarks and discussion.

Participants review the goals listed for the session and discuss whether

these goals have been achieved. Discussion centers around possible changes for improvement of this session or the sessions to follow.

A brief summary of the following day's activities is provided.

SESSION II - DESIGN MAP

PARTICIPANTS:	grade four language arts teachers, principals, professional school board personnel
TOPIC:	evaluation practices of grade four language arts teachers
GOALS OF THE SESSION:	to develop a list of evaluation activities actually used by the participants in a classroom situation to promote an awareness of the reasoning behind individual's evaluation practices to expose participants to the ideas and practices of other members of the profession
PROCEDURE:	pass out materials needed for the session instruct participants on how to use the strategies list divide into small groups (same groups as previous session) for brainstorming session use the overhead projector to develop a final list of evaluation strategies keep an anecdotal record of unusual or interesting occurrences plus a checklist to record areas where changes in inservice strategies may need to be considered
TEACHING STRATEGIES:	large group discussion small group discussion brainstorming

EVALUATION:	anecdotal records and checklist compiled by coordinator
	participant input
MATERIALS:	agenda
	strategy sheet for individual participant
	strategy sheet for small groups
	overhead projector and overheads
	list of goals
	anecdotal record form
	evaluation checklist
	list of thought provokers for the concluding section

AGENDA

9:00 - 9:15	Discussion of Goals for the Session
9:15 - 9:30	Individual Participant Consideration of Evaluation Strategy List
9:30 - 10:15	Small Group Brainstorming Session
10:15 - 10:45	Large Group Activity
10:45 - 11:00	Coffee Break

(Total - 2 hours)

Session II - Description

The coordinator begins the session by initiating a discussion of the goals listed on the session map. It is essential that this list be available for inspection by participants, with the possibility of some change occurring in reaction to participant input. Where necessary, the coordinator notes recommendations made.

The bulk of this session is a pointed examination of evaluation strategies presently in use in the classrooms. Participants review their personal evaluation practices and list them on the strategy form provided (see Appendix for Session II). The form is developed so that it provokes more than a rote listing of strategies. Participants must consider the reasoning behind their choices and list both positive and negative aspects of each strategy.

The participants move from a personal examination of strategies used to a small group brainstorming session on the same topic. A strategy form identical to that given to individuals is provided for the small groups. One member of each group records the input of participants. By using a small group format, individuals are exposed to the ideas of others in a situation where discussion is encouraged. One of the goals of the inservice is to provide participants with exposure to other professionals in the workforce. This activity provides for not only exposure but also discussion and questioning.

The remainder of the session is to be used for a large group brainstorming activity which assists in pooling the thoughts of all participants. The coordinator acts as a recorder, using overheads to present to participants lists developed by the small groups. The overheads (see Appendix for Session II) are purposely structured to indicate patterns emerging in evaluation strategies used by participants.

The coordinator concludes the session with connecting comments. These are comments or thought provokers (see Appendix for Session II), aimed at tying the first and second session together. Participants are encouraged to consider their beliefs and attitudes towards evaluation and how these are reflected in their evaluation of students.

SESSION III - DESIGN MAP

PARTICIPANTS:	grade four language arts teachers; principals, professional school board personnel
TOPIC:	the history of evaluation
GOALS OF THE SESSION:	to provide participants with information on the history of evaluation to encourage teachers to reflect on knowledge they have acquired concerning the history of evaluation, previous to this inservice
PROCEDURE:	pass out materials needed for the session set up chart paper to be used in large group brainstorming session rearrange seating for afternoon lecture format set up overhead projector for use in delivery of information on the history of evaluation
TEACHING STRATEGIES:	writing activity brainstorming charting lecturing
EVALUATION:	anecdotal records and checklist compiled by coordinator participant input

MATERIALS:	list of goals
	true-false form
	overhead projector and overheads
	chart paper
	evaluation checklist
	anecdotal record form

AGENDA

11:00 - 11:15	Discussion of Goals
11:15 - 12:00	Activity Using True-False Forms
12:00 - 1:00	Lunch
1:00 - 1:45	Early Beginnings of Evaluation
1:45 - 2:30	Program Evaluation
2:30 - 3:00	Student Evaluation
3:00 - 3:15	Comments of Participants

(Total - 3 hours, 15 minutes)

Session III - Description

At the beginning of the session a list of goals is given to each of the participants. Discussion is directed by the coordinator to include an examination of the role of goals in the evaluation process. The importance of setting goals and of measuring their attainment is considered in the discussion.

The major focus of the session is the history of evaluation. Before giving out information gathered on the topic, the coordinator attempts to assess, to some degree, the participants' present level of knowledge in the area. Each individual completes the form (see Appendix for Session III) provided by the coordinator. The form includes general statements to be marked true or false, plus a few short answer questions. Space is left on the form for additional statements from the individual. The coordinator uses these forms as a tool to encourage participants to consider the amount of information they have or they lack in this particular area. Individuals file these forms in their folders with other material collected.

After completion of this activity based on individual reflections, a group activity commences. The coordinator uses the input from a large group brainstorming session to develop a large chart of information concerning the history of evaluation. This is tacked to the wall for later reference.

The first part of the session gives some indication of what participants know about the history of evaluation. The coordinator takes this into consideration when delivering the rest of the inservice.

The coordinator devotes the final two hours to presenting participants with information pertaining to the history of evaluation. An overhead projector and a lecture format are used to deliver the information.

To prepare for this session, coordinators need to become familiar with the historical development of evaluation. A skeleton of information is provided in chapter two of the thesis. Pertinent sources are listed in the bibliography.

During this session the coordinator uses overheads (see Appendix for Session III) to aid in the structuring of information into appropriate sections. The sections follow a similar format to those presented in chapter two of this thesis. The first section attempts to provide information on the test and measurement movement's influence on the development of evaluation. The second section reports on research conducted in program evaluation. Student evaluation is the topic of the third section.

After the presentation of information, the session is concluded with input from participants. The coordinator attempts to assess whether participants feel goals of the inservice have been achieved.

SESSION IV - DESIGN MAP

PARTICIPANTS:	grade four language arts teachers, principals, professional school board personnel
TOPIC:	evaluation in the language arts area
GOALS OF THE SESSION:	to highlight for participants the necessity of utilizing the findings of evaluation research to develop a list of appropriate evaluation strategies in light of previous sessions to present the Department of Education's suggested list of evaluation strategies to have participants evaluate the inservice as a whole
PROCEDURE:	give out materials necessary for inservice conduct large group discussion on converting research into practice arrange participants into original small groups to work on strategies list prepare overhead projector for use in the section on language arts and the Department of Education's list of evaluation activities
TEACHING STRATEGIES:	large group discussion small group discussion brainstorming lecturing

EVALUATION:	anecdotal records and checklist compiled by coordinator
	participant input through discussion
	written completion of an evaluation form
MATERIALS:	list of goals
	overhead projector and overheads
	research into practice forms
	evaluation strategy form
	list of Department of Education's recommended evaluation activities (overheads)
	list of points to address during final open discussion
	evaluation form for participants

AGENDA

9:00 - 9:30	Discussion of Goals
9:30 - 10:30	Activities Related to the Idea of Putting Research Findings into Practice
10:30 - 10:45	Coffee
10:45 - 12:00	Development of a List of Evaluation Strategies
12:00 - 1:00	Lunch
1:00 - 2:00	Language Arts Session
2:00 - 3:00	Discussion of Inservice Plan Completion of Evaluation Forms

(Total - 5 hours)

Session IV - Description

The session begins with reading and discussing the inservice goals. Where necessary, modifications are made.

In the early stages of the session the coordinator connects this session to the previous one by centering discussion around converting research into practice. Participants, while in small groups, consider whether this conversion actually is occurring at the school level. A form (see Appendix for Session IV) is provided by the coordinator. This is done to direct participants' thinking towards examining practice at the school level. Whenever participants suggest practice is not reflecting research findings, they record the possible causal factors.

Following this session participants develop a new list of "appropriate" evaluation strategies in light of the past three day's experiences. A form similar to the one used in session two is provided by the coordinator. Participants complete their forms and compare them to the ones filed in their folders after session two. While comparing the lists, participants note obvious similarities and differences between the two sets of forms. The coordinator observes this activity closely in an attempt to ascertain whether or not revisions are being made; any indication of changes in participants' attitudes and ideas concerning evaluation is noted.

The second half of the session is devoted to a concern for issues in the language arts curriculum area. A lecture format is utilized to present information on the four main components - reading, writing, listening, and speaking. Suggestions of activities for evaluating these components are put forth.

During this session, the Department of Education's guideline for evaluating language arts is discussed. A list of activities recommended by the Department's document Evaluating Language Arts Performance in the Elementary School is presented.

During the last hour of the session, large group discussion of the implications of the three-day inservice takes place. The coordinator draws together threads of the inservice sessions for the group's consideration. Emphasis is placed on the content and delivery of the inservice, with special mention of teaching strategies and evaluation procedures.

Time is taken to put forth some general statements about the anecdotal records and checklists kept by the coordinator over the three-day period.

Participants complete the evaluation forms (see Appendix for Session IV) provided by the coordinator. By using these the coordinator attempts to assess the success of the inservice and to find means for improving the plan itself.

APPENDICES FOR INSERVICE

APPENDIX: SESSION I

AGENDA

9:00 - 9:30	Introduction
9:30 - 10:00	Discussion of Objectives of the Inservice
10:00 - 10:45	Small Group Work Using List of Evaluation Statements
10:45 - 11:00	Coffee
11:00 - 12:00	Development of an Acceptable List of Evaluation Statements
12:00 - 1:00	Lunch
1:00 - 1:45	Reading and Discussion of a Case Study Concerned With the Description of a Grade Four Language Arts Teacher
1:45 - 2:30	Writing Activity - Participants' Development of a Personal Teacher Profile
2:30 - 3:00	Concluding Section

(Total - 4 hours, 45 minutes)

GOALS OF THE SESSION

- to provide a time for participants to reflect on beliefs held in the area of evaluation
- to provide an environment to promote of reflection on philosophies held
- to encourage participants to examine how and why they think about evaluation as they do
- to provide a forum for discussion of beliefs
- to create an awareness on the part of the participants of the possible limitations and strengths of the personal philosophies described and discussed

EVALUATION CONCEPTS

Below are a series of statements which reflect various views on evaluating learners. Read all of them carefully, then as a group do the following.

1. Place a ++ before the statement best reflecting majority opinion.
2. Place a + before all other statements which reflect majority agreement.
3. Place a -- before the statement least reflecting majority opinion.
4. Place a - before all other statements which reflect majority disagreement.
5. Feel free to leave statements blank.
6. Note the number of dissenters in your group overall.

___ In evaluating learners we mostly measure what we know how to measure rather than what we should measure.

___ Evaluation of learners is a continuous process of effective communication as needed, when needed, to whomever needs it.

___ The primary purpose of evaluating learners is to report to learners their achievement in relation to course objectives.

___ Paper and pencil tests are the best way to ensure that learners are achieving at least minimum competencies.

___ Learner evaluations should facilitate teaching and learning.

___ Learner evaluations should provide information on learners' talents, capabilities, and achievements, and should reflect learners' attitudes and interests.

___ Since all evaluation is by its very nature threatening, there is nothing that can be done to lessen the threat factor of tests and exams.

___ The primary purpose of evaluating learners is to report to learners their achievement in relation to other learners.

___ It is not right to assign grades to learners for things like homework, classroom learning activities, and cooperative attitudes.

___ Surprise tests are sometimes necessary to keep learners on their toes.

___ Learner evaluations have nothing to do with helping learners develop skills in self-appraisal.

___ Testing and grading systems must continue because there are no viable alternatives.

___ Learner evaluations should be undertaken mainly for improvement purposes.

___ Criterion-referenced testing focuses on success rather than failure.

___ In evaluating learners classroom observations can be just as valid as paper and pencil tests.

___ Grading of learners is necessary because life is competitive and grading prepares them for the real world.

___ Tests and measures used in evaluating learners should contain only items that reflect previously stated behavioural objectives.

___ It should be nice to use a great variety of measurement techniques in evaluating learners, but written tests are the only practical solution because of time restraints.

(Taken from Dr. M. Kennedy, Inservice, October 1985).

CASE STUDY

Mary Tucker has been teaching for the past eighteen years, with the last seven being spent in grade four classrooms. She has always enjoyed teaching and certainly wouldn't consider herself to be suffering from burnout because she has been on the job for eighteen years. As a matter of fact, Mary enjoys teaching now as much as she did when she began in the late 1960's.

Mary has survived lots of changes. Some she agreed with, some she didn't. She believes she could be more effective in her job if days had more hours and if she had more expertise in certain areas. This is especially true when Mary considers the task of evaluating her students. She is aware that programs recommend the use of observation checklists and student profiles, but she doesn't have time to fool with such things. Mary finds herself using an extensive amount of time just collecting enough marks to get ready for report card time. It seems she doesn't have time to recuperate from the first reporting session before second term reports are due. Three formal reporting sessions a year, for grade four students, seems to Mary an overindulgence in the activity. She thinks teachers should meet parents and discuss their children's work, maybe even use the student profile to which programs refer. She believes a list of marks is not always necessary when considering a child's progress. Mary can guess the marks she will assign to students without ever administering the tests and quizzes normally completed for this purpose. She believes a tremendous amount of instructional time dissipates during yearly testing activities.

She thinks it's ironic, when considering curriculum areas such as language arts, the evaluation strategies she ends up using. Supposedly, this curriculum area consists of four major components: reading, writing, listening and

speaking. Where one gets time to emphasize these is a question Mary would like answered. While evaluating, she gives tests whose objectives are to assess the reading and writing components of the program while ignoring the other two areas. Even if Mary attempted to evaluate those other two components, how would she go about accomplishing the task? She is very unsure of what strategies should be used when assigning marks to students for listening and speaking activities. Even if she included lots of oral language experiences in her daily class activities, how would she justify her marks to parents dissatisfied with them? She knows parents want to be able to see the tests their children failed.

Mary believes she needs some guidance in the evaluation area. Programs are expounding on the "wholistic" approach and the importance of formative evaluation. Mary doesn't remember any education courses dealing with evaluation. She has always evaluated her students the best way she knew how. Mary thinks her marks are fair, but she is not sure if they serve any real purpose. She spends so much time testing and marking, little time is left to work with the students to help alleviate problems they are experiencing. Mary believes the situation could be improved if administrators and parents exerted less pressure for marks. There isn't one student in Mary's classroom who she couldn't give a full account of if requested. However, for Mary, a nagging pang of guilt persists due to her practice of limiting the majority of student evaluation to recording numbers to represent student weaknesses. It seems to her that this should be the least important function she should serve as an evaluator.

She believes it is far more important for teachers to work with children in the hopes that students may develop in the areas that teachers have recorded as being weak.

Mary thinks it's about time educators set their priorities straight and alloted time appropriately. Maybe teachers would have more time for what they felt, and are told, is important if they could cut out half of what they believe to be pointless trivia.

CASE STUDY DISCUSSION STARTERS

- Describe the teacher's feelings about evaluation. How are they similar or dissimilar to yours? What would you say is her philosophy of evaluation?
- What strategies appear to prevail in her classroom and why?
- "I can guess the marks students will receive without ever administering the test and quizzes normally completed for this purpose." Please comment.
- What "changes" (see second paragraph) do you think this teacher has experienced? Changes in theory? Changes in practice?
- Do you think instructional time is being put to proper use?
- Is there too much pressure exerted on teachers to produce marks?
- Are the four components of language arts emphasized in your classroom? Do you evaluate your students in these areas? If not, give reasons.
- Ideas the group would like to propose for discussion.

PARTICIPANT PROFILE

Experience and training in evaluation:

Beliefs about evaluation:

Practices in evaluation:

Reasons behind practices used:

Suggestion for possible changes in practices:

ANECDOTAL RECORD SHEET

Goals of Session

Comments:

Goals added -

Goals deleted -

Small and Large Group Activities

Comments:

Did groups work well together?

Who were the leaders?

Who were not participating?

Other strategies for possible use?

Is regrouping necessary?

Case Study Work

Comments made by participants:

Writing Activity

Was there full participation?

Problems which arose?

Conclusion

Comments by participants:

Negative -

Positive -

Changes suggested:

OBSERVATION CHECKLIST

Names	Takes part in small group work	Vocal in large group session	Displays leadership qualities	Communicates well with others	Positive reaction to session	Displays changes in opinions or beliefs	Offers suggestions for improvement of inservice

APPENDIX: SESSION II

AGENDA

9:00 - 9:15 Discussion of Goals for the Session

9:15 - 9:30 Individual Participant Consideration of Evaluation Strategy List

9:30 - 10:15 Small Group Brainstorming Session

10:15 - 10:45 Large Group Activity

10:45 - 11:00 Coffee

(Total - 2 hours)

GOALS OF THE SESSION

- to develop a list of evaluation activities actually used by the participants in a classroom situation
- to promote an awareness of the reasoning behind individual's evaluation practices
- to expose participants to the ideas and practices of other members of the profession

EVALUATION STRATEGY FORM

Strategy used:

Why is it used?

Positive aspects of its use -

Negative aspects of its use -

Strategy used:

Why is it used?

Positive aspects of its use -

Negative aspects of its use -

Strategy used:

Why is it used?

Positive aspects of its use -

Negative aspects of its use -

Strategy used:

Why is it used?

Positive aspects of its use -

Negative aspects of its use -

Strategy used:

Why is it used?

Positive aspects of its use -

Negative aspects of its use -

Strategy used:

Why is it used?

Positive aspects of its use -

Negative aspects of its use -

OVERHEAD

List of strategies used by participants:

Strategies used to evaluate reading:

Pen and Paper	Other

Strategies used to evaluate writing:

Pen and Paper	Other

Strategies used to evaluate listening:

Pen and Paper	Other

Strategies used to evaluate speaking:

Pen and Paper	Other

THOUGHT PROVOKERS

What is your definition of evaluation?

Why do we evaluate as we do?

How much information do we have about appropriate evaluation practices?

Is our practice reflective of our beliefs?

What changes could be made to enhance practice?

How much politics is involved in evaluation?

ANECDOTAL RECORD SHEET

Goals of Session

Comments:

Goals added -

Goals deleted -

Small and Large Group Activities

Comments:

Did groups work well together?

Who were the leaders?

Who were not participating?

Other strategies for possible use?

Is regrouping necessary?

Were there a lot of strategies being discussed?

Were the strategy forms of use?

Any changes to be made?

What patterns emerged on the strategy forms?

Conclusion

Comments by participants:

Negative -

Positive -

Changes suggested:

OBSERVATION CHECKLIST

Names	Takes part in small group work	Vocal in large group session	Displays leadership qualities	Communicates well with others	Positive reaction to session	Displays changes in opinions or beliefs	Offers suggestions for improvements of inservice

APPENDIX: SESSION III

AGENDA

11:00 - 11:15 Discussion of Goals

11:15 - 12:00 Activity Using True-False Form

12:00 - 1:00 Lunch

1:00 - 1:45 Early Beginnings of Evaluation

1:45 - 2:30 Program Evaluation

2:30 - 3:00 Student Evaluation

3:00 - 3:15 Comments of Participants

(Total - 3 hours, 15 minutes)

GOALS OF THE SESSION

- to provide participants with information on the history of evaluation
- to encourage teachers to reflect on knowledge they have already acquired concerning the history of evaluation

TRUE-FALSE FORM

I. Please tick the appropriate response.

	True	False
1. Evaluation began when tests were devised for classroom use.		
2. Michael Patton is known as "the father" of educational testing.		
3. Quantitative data is the most important data to be utilized in an evaluation activity.		
4. Evaluation has always been a concern of educators.		
5. Evaluation has no connection to the test and measurement movement.		
6. Intelligence testing became popular with the introduction of Alfred Binet's work.		
7. Teachers have been evaluating since the early 1900's.		
8. At times, the information from standardized testing programs has been used to judge the quality of schools or teachers.		
9. A naturalistic paradigm of thought accepts the use of qualitative data in an evaluation activity.		
10. There is a connection between World War I psychological testing and the way children in present day classrooms are being evaluated.		

II. List any people or events you know which have affected the field of evaluation.

III. Give a short account of any changes you feel have occurred over the past fifty years in methods of evaluating students in the educational system.

IV. Additional comments.

OVERHEAD #1

Early Beginnings of Evaluation:

Test and Measurement Movement

Discussion Points

- Psychological testing -Introduction to school system
  - Alfred Binet
  - Robert Thorndike

- United States - large scale testing programs are implemented in response to public cry for accountability.

- * Massive development of school grading systems with their dependency on teacher-made tests.

- Test and measurement movement = scientific paradigm of thought with dependency on quantitative data.

- * Ralph Tyler - differentiated between the concepts of measurement and evaluation

- stressed the importance of educational objectives

OVERHEAD #2

Program Evaluation

Discussion Points

- Russian Sputnik (1957) launched
  - United States pours money into education
- National Defense Act (1958)
  - Federally funded programs have to be evaluated - no one capable of performing the task properly
- Numerous models of evaluation proposed in the research
  - Stake - Countenance Model
  - Stufflebeam - Context-Input-Process-Product Model
  - Provus - Discrepancy Model
  - Scriven - Goal-Free Evaluation
  - Eisner - Educational Connoisseurship and Criticism
- Exclusive use of a scientific paradigm of thought with its dependency on quantitative data is questioned
- Shift to a naturalistic paradigm of thought which encouraged the use of qualitative data
- Definition of evaluation is broadened

- Production of models cease and research considers the character of the evaluator and the utility of the data collected

- Guba and Lincoln - The Responsive Evaluator

- Patton - Creative Evaluator
  the active-reactive-adaptive evaluator

- Connection between research in program evaluation and the classroom teacher

OVERHEAD #3

Student Evaluation

Discussion Points

- Development of student evaluation and its connection to the test and measurement movement

- Restricted view of evaluation held by educators (giving and receiving of grades)

- Cry for accountability = excessive use of standardized and teacher-made tests

- Over time definitions and theories of evaluation have changed but practice has not

- Theory and practice - do they gel?

- Teachers' paradigms of thought/heuristics

ANECDOTAL RECORD SHEET

Goals of Session

Comments:

Goals added -

Goals deleted -

How much do participants appear to know about the history of evaluation?

Were they responsive to lecture style for disseminating information?

Conclusion

Comments made by participants:

Negative -

Positive -

Changes suggested:

OBSERVATION CHECKLIST

Names	Takes part in small group work	Vocal in large group session	Displays leadership qualities	Communicates well with others	Positive reaction to session	Displays changes in opinions or beliefs	Offers suggestions for improvement of inservice

APPENDIX: SESSION IV

AGENDA

9:00 - 9:30	Discussion of Goals
9:30 - 10:30	Activities Related to the Idea of Putting Research Findings into Practice
10:30 - 10:45	Coffee
10:45 - 12:00	Development of List of Evaluation Strategies
12:00 - 1:00	Lunch
1:00 - 2:00	Language Arts Session
2:00 - 3:00	Discussion of Inservice Plan Completion of Evaluation Forms

(Total - 5 hours)

GOALS OF THE SESSION

- to highlight, for participants, the necessity of utilizing the findings of evaluation research
- to develop a list of appropriate evaluation strategies in light of previous sessions
- to present the Department of Education's suggested list of evaluation strategies
- to have participants evaluate the inservice as a whole

RESEARCH INTO PRACTICE FORM

Ideas/Strategies Presented in the Literature	Put into Practice Yes	No	Participants' Suggestion for Why the Strategy is or isn't put into Practice
Strategy			Reason

EVALUATION STRATEGY FORM

List the evaluation strategies you believe should be used in classrooms.

Strategy:

Positive aspects -

Negative Aspects -

Strategy:

Positive Aspects -

Negative Aspects -

Strategy:

Positive Aspects -

Negative Aspects -

Strategy:

Positive Aspects -

Negative Aspects -

Strategy:

Positive Aspects -

Negative Aspects -

Strategy:

Positive Aspects -

Negative Aspects -

OVERHEAD OUTLINE

List of procedures used for diagnostic-formative-summative evaluation presented in Evaluating Language Arts Performance in the Elementary School, the Department of Education's booklet on evaluation of students in the language arts.

- Anecdotal Records
- Oral Quizzes
- Interviews
- Questionnaires
- Conferences
- Teacher-Made Tests (Group and Individual)
- Stored Samples of Written Work Checklists
- Report Cards
- Cumulative Records
- Intelligence Tests
- Standardized Achievement Tests
- Teacher/Peer/Self-Evaluation Using Checking and established Criteria
- Parent Involvement

ANECDOTAL RECORD SHEET

Goals of Session

Comments:

Goals added -

Goals deleted -

Small and Large Group Activities

Comments:

Did groups work well together?

Who were the leaders?

Who were not participating?

Other strategies for possible use?

Did it appear as if participants were questioning their evaluation practices?

What possible causal factors did they record on the forms?

What comments were made during the discussion concerning the list of evaluation strategies recommended by the Department of Education?

Was there any indication of changes in participants' attitudes and ideas concerning evaluation?

Comments made during the discussion about the content and delivery of the inservice.

Were the final evaluation forms completed by the participants without any difficulty?

Comments during the activity.

OBSERVATION CHECKLIST

Names	Takes part in small group work	Vocal in large group session	Displays leadership qualities	Communivates well with others	Positive reaction to session	Displays changes in opinions or beliefs	Offers suggestions for improvement of inservice		

WORKSHOP EVALUATION

In order to make improvements and revisions for any future inservices, I would appreciate your comments on this one. Please take a few minutes to complete this questionnaire.

1. Did the inservice meet your needs?

2. Were you, as learners, satisfied with the content?

3. List three things which were most beneficial to you.

4. List three things which were least beneficial to you.

5. Other comments.

BIBLIOGRAPHY

BIBLIOGRAPHY

Brown, I. J. (1985). A study of the evolution of school library media programs and the process of educational change with implications for program implementation. St. John's: Memorial University of Newfoundland.

Carroll, D. (1971). Critique of the model. In M. Provus, Discrepancy evaluation. California: McCutcheon Publishing Company.

Crocker, R. (1984). Elements of teacher functional paradigms. Paper presented at the conference on Research on Teaching and Teacher Education, Vancouver, British Columbia.

Department of Education of Newfoundland. (1985). Evaluating language arts performance in the elementary school. St. John's, Newfoundland: Department of Education.

Dubois, P. (1970). A history of psychological testing. Boston: Allyn and Bacon, Inc.

Eisner, W. E. (1983). Educational connoisseurship and criticism: Their form and functions in educational evaluation. In Madaus et al., Evaluation models. Hingham, M.A.: Kluver Academic Publishers.

Fullan, M. (1982). The meaning of educational change. Toronto: OISE Press.

Goodlad, J. and Klein, M. (1974). Looking behind the classroom door. Ohio: Charles A. Jones Publishing Company.

Guba, E. G. (1967, April). The expanding concept of research. Theory into Practice, 6(2), pp. 57-65.

Guba, E. G. (1969). The failure of educational evaluation. Educational Technology, 9(5), pp. 29-33.

Guba, E. G. and Lincoln, Y. (1981). Effective evaluation. California: Jossey-Bass Inc.

Hayman, J. L. and Napier, R. N. (1975). Evaluation in the schools: A human process for renewal. California: Brooks/Cole Company.

Hedges, W. (1969). Evaluation in the elementary school. New York: Holt, Rinehart and Winston, Inc.

House, E. R. (1974). The politics of educational innovations. Berkeley, California: McCutcheon Publishing Company.

Jackson, P. W. (1968). Life in classrooms. New York: Holt, Rinehart and Winston.

Joncich, G. (1962). Whither thou educational scientist. Teachers College Record, 64(1), pp. 1-11.

Kelly, G. A. (1955). The psychology of personal constructs. New York: W. W. Norton & Company.

Madaus, G. and Scriven, M. and Stufflebeam, J. (Eds.). (1983). Evaluation models viewpoints on educational and human services evaluation. Hingham, M.A.: Kluver Academic Publishers.

Patton, M. Q. (1978). Utilization-focused evaluation. California: Sage Publishers.

Patton, M. Q. (1980). Qualitative evaluation methods. California: Sage Publishers.

Patton, M. Q. (1981). Creative evaluation. California: Sage Publishers.

Perrone, V. (1977). The abuses of standardized testing. Indiana: Phi Delta Kappa Educational Foundation.

Phi Delta Kappa National Study Committee. (1971). Educational evaluation and decision-making. Illinois: F. E. Peacock Publishing.

Provus, M. (1971). Discrepancy evaluation. California: McCutcheon Publishers Company.

Sarason, S. (1971). The culture of the school and the problem of change. Boston: Allyn and Bacon, Inc.

Sawin, E. I. (1969). Evaluation and the work of the teacher. California: Wadsworth Publishing Company.

Scriven, M. (1974). Evaluation perspectives and procedures. In Popham (Ed.), Evaluation in education. California: McCutcheon Publishing Company.

Stake, R. (1967, April). The countenance of educational evaluation. Teachers' College Record, 68(7), pp. 523-540.

Stufflebeam, D. (1967, June). The use and abuse of evaluation in title III. Theory into Practice, 6(3), pp. 126-133.

Stufflebeam, D. (1971). Critique of the model. In M. Provus, Discrepancy evaluation. California: McCutcheon Publishing Company.

Stufflebeam, D. (1973). An introduction to the PDK book Educational evaluation and decision-making. In Worthen and Sanders, Educational evaluation: Theory into practice. California: Wadsworth Publishing Company.

Stufflebeam, D. and Webster, W. J. (1983). An analysis of alternative approaches to evaluation. In Madaus et al., (Eds.), Evaluation models viewpoints on educational and human services evaluation. Hingham, M.A.: Academic Publishers.

Tyler, R. and Wolf, R. (Eds.). (1974). Crucial issues in testing. California: McCutcheon Publishing Company.

Worthen, R. R. and Sanders, J. R. (1973). Educational evaluation: Theory into practice. California: Wadsworth Publishing Company.

APPENDIX A

QUESTIONNAIRE

QUESTIONNAIRE

Female ___

Male ___

Grade Four Language Arts Teachers

Please circle the appropriate response.

Years of university training	1	2	3	4	5	6	7	or more			
Program of studies enrolled in	Primary			Elementary				High School			
Courses taken in testing and evaluation	0	1	2	3	or more						
Present teaching certificate	1	2	3	4	5	6	7				
Total number of years teaching experience	1	2	3	4	5	6	7	8	9	10	or more
Number of years teaching grade four	1	2	3	4	5	6	7	8	or more		

On the following pages of the questionnaire you will find listed numerous activities for ongoing student evaluation. Some you may use extensively; others you may not. Please read each item carefully and place an X in the box representing the most appropriate response. Please note that there are 3 separate columns to which a response is needed. A sincere thank you is extended to all respondents. Without your invaluable cooperation and contribution this study could not be conducted.

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				Importance of activity in ongoing student evaluation			Importance of activity in the assigning of final grades		
	Hardly Ever	Occasionally	Frequently	Almost Always	Not Very Important	Important	Very Important	Not Very Important	Important	Very Important
Spelling Test										
Oral presentation by student to the whole class										
Teacher-made observation checklist (Teacher writes out a list of student behaviors and records his/her observations of these behaviors)										
Group Project										
Teacher-made cloze test (Eg. a selection in which every 5th or 7th word is replaced by a blank and students are requested to fill in the blanks with words that fit semantically and syntactically)										
Test (end of term)										

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				Importance of activity in ongoing student evaluation			Importance of activity in the assigning of final grades		
	Hardly Ever	Frequently	Occasionally	Almost Always	Not Very Important	Important	Very Important	Not Very Important	Important	Very Important
Writing folder (Collection of student's "published" and exploratory writing)										
Reading contract made with student (Student signs an agreement with teacher to read a book)										
Commercial Workbooks (Flip Flops and Zoom Shots)										
Listening Activities using audiotapes										
Miscue Inventories - Oral reading tests (Student reads aloud to the teacher who records the student's miscues during presentation)										

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				Importance of activity in ongoing student evaluation			Importance of activity in the assigning of final grades		
	Hardly Ever	Frequently	Occasionally	Almost Always	Not Very Important	Important	Very Important	Not Very Important	Important	Very Important
Student self-assessment checklist (Teacher and student compile a checklist to be used by the student to record development in particular areas)										
Activities suggested in the Networks program										
Teacher-made test on writing conventions										
Writing conference (Teachers arrange time for interviews with individual students)										
Teacher-made worksheets (Fill in the blank, multiple choice, short answer type worksheets developed by teachers)										

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				Importance of activity in ongoing student evaluation			Importance of activity in the assigning of final grades		
	Hardly Ever	Frequently	Occasionally	Almost Always	Not Very Important	Important	Very Important	Not Very Important	Important	Very Important
Anecdotal records (Teacher records of factual descriptions of critical incidents that occur during a school term)										
Quiz (Teacher-made test on a block of content)										
Audio-tapes used to record reading experiences of individual students										
Student Profile (Comprehensive collection of data on individual students - tests, projects, reading inventories, etc.)										
Teacher questioning (Oral and written questions given to students specifically to ascertain and emphasize level of listening)										

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				Importance of activity in ongoing student evaluation			Importance of activity in the assigning of final grades		
	Hardly Ever	Frequently	Occasionally	Almost Always	Not Very Important	Important	Very Important	Not Very Important	Important	Very Important
Teacher observation during directed reading activities and free reading time										
Standardized reading tests										
Dramatizations of readings										

APPENDIX B

RESULTS

Activities - First Scale

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation: Hardly Ever	Frequently	Occasionally	Almost Always	No Response
Spelling Test	8%	29%	18%	41%	3%
Oral presentation by student to the whole class	6%	30%	55%	7%	2%
Teacher-made observation checklist (Teacher writes out a list of student behaviors and records his/her observations of these behaviors)	31%	25%	31%	8%	4%
Group Project	9%	29%	52%	5%	5%
Teacher-made cloze test (Eg. a selection in which every 5th or 7th word is replaced by a blank and students are requested to fill in the blanks with words that fit semantically and syntactically)	11%	35%	37%	16%	1%
Test (end of term)	24%	11%	6%	54%	5%
Writing folder (Collection of student's "published" and exploratory writing)	1%	21%	7%	69%	1%

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				
	Hardly Ever	Frequently	Occasionally	Almost Always	No Response
Reading contract made with student (Student signs an agreement with teacher to read a book)	70%	12%	12%	2%	4%
Commercial Workbooks (Flip Flops and Zoom Shots)	13%	28%	26%	29%	4%
Listening Activities using audiotapes	17%	22%	41%	17%	4%
Miscue Inventories - Oral reading tests (Student reads aloud to the teacher who records the student's miscues during presentation)	31%	17%	47%	3%	2%
Student self-assessment checklist (Teacher and student compile a checklist to be used by the student to record development in particular areas)	67%	10%	19%	1%	3%
Activities suggested in the Networks program	1%	40%	14%	44%	1%
Teacher-made test on writing conventions	14%	29%	35%	17%	4%

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				
	Hardly Ever	Frequently	Occasionally	Almost Always	No Response
Writing conference (Teachers arrange time for interviews with individual students)	23%	33%	30%	12%	2%
Teacher-made worksheets (Fill in the blank, multiple choice, short answer type worksheets developed by teachers)	24%	23%	44%	9%	0%
Anecdotal records (Teacher records of factual descriptions of critical incidents that occur during a school term)	30%	25%	30%	9%	6%
Quiz (Teacher-made test on a block of content)	12%	33%	25%	28%	2%
Audio-tapes used to record reading experiences of individual students	55%	6%	36%	1%	2%
Student Profile (Comprehensive collection of data on individual students - tests, projects, reading inventories, etc.)	8%	34%	7%	50%	1%
Teacher questioning (Oral and written questions given to students specifically to ascertain and emphasize level of listening)	2%	42%	21%	36%	-

Activity used for ongoing student evaluation in the language arts area	Frequency of activity being used for student evaluation				
	Hardly Ever	Frequently	Occasionally	Almost Always	No Response
Teacher observation during directed reading activities and free reading time	2%	37%	21%	38%	1%
Standardized reading tests	53%	4%	34%	5%	4%
Dramatizations of readings	19%	23%	52%	6%	1%

Activities - Second Scale

Activity used for ongoing student evaluation in the language arts area	Importance of activity in ongoing student evaluation			
	Not Very Important	Important	Very Important	No Response
Spelling Test	29%	56%	11%	4%
Oral presentation by student to the whole class	18%	67%	10%	4%
Teacher-made observation checklist (Teacher writes out a list of student behaviors and records his/her observations of these behaviors)	31%	48%	12%	10%
Group Project	15%	67%	10%	7%
Teacher-made cloze test (Eg. a selection in which every 5th or 7th word is replaced by a blank and students are requested to fill in the blanks with words that fit semantically and syntactically)	16%	53%	25%	6%
Test (end of term)	24%	34%	35%	7%
Writing folder (Collection of student's "published" and exploratory writing)	3%	29%	66%	2%

Activity used for ongoing student evaluation in the language arts area	Importance of activity in ongoing student evaluation			
	Not Very Important	Important	Very Important	No Response
Reading contract made with student (Student signs an agreement with teacher to read a book)	59%	25%	4%	12%
Commercial Workbooks (Flip Flops and Zoom Shots)	25%	53%	18%	4%
Listening Activities using audio-tapes	21%	58%	11%	9%
Miscue Inventories - Oral reading tests (Student reads aloud to the teacher who records the student's miscues during presentation)	36%	52%	7%	4%
Student self-assessment checklist (Teacher and student compile a checklist to be used by the student to record development in particular areas)	55%	31%	3%	11%
Activities suggested in the Networks program	3%	64%	31%	1%
Teacher-made test on writing conventions	22%	54%	18%	6%

Activity used for ongoing student evaluation in the language arts area	Importance of activity in ongoing student evaluation			
	Not Very Important	Important	Very Important	No Response
Writing conference (Teachers arrange time for interviews with individual students)	21%	51%	23%	4%
Teacher-made worksheets (Fill in the blank, multiple choice, short answer type worksheets developed by teachers)	36%	53%	9%	2%
Anecdotal records (Teacher records of factual descriptions of critical incidents that occur during a school term)	34%	42%	14%	9%
Quiz (Teacher-made test on a block of content)	16%	51%	29%	4%
Audio-tapes used to record reading experiences of individual students	48%	39%	4%	10%
Student Profile (Comprehensive collection of data on individual students - tests, projects, reading inventories, etc.)	7%	41%	51%	1%
Teacher questioning (Oral and written questions given to students specifically to ascertain and emphasize level of listening)	2%	56%	40%	2%

Activity used for ongoing student evaluation in the language arts area	Importance of activity in ongoing student evaluation			
	Not Very Important	Important	Very Important	No Response
Teacher observation during directed reading activities and free reading time	10%	58%	31%	1%
Standardized reading tests	48%	37%	4%	11%
Dramatizations of readings	35%	52%	7%	6%

Activities - Scale Three

Activity used for ongoing student evaluation in the language arts area	Importance of activity in the assigning of final grades: Not Very Important	Important	Very Important	No Response
Spelling Test	42%	42%	10%	6%
Oral presentation by student to the whole class	31%	58%	4%	6%
Teacher-made observation checklist (Teacher writes out a list of student behaviors and records his/her observations of these behaviors)	34%	46%	10%	10%
Group Project	25%	59%	9%	7%
Teacher-made cloze test (Eg. a selection in which every 5th or 7th word is replaced by a blank and students are requested to fill in the blanks with words that fit semantically and syntactically)	20%	57%	21%	2%
Test (end of term)	24%	33%	36%	7%
Writing folder (Collection of student's "published" and exploratory writing)	3%	28%	66%	2%

Activity used for ongoing student evaluation in the language arts area	Importance of activity in the assigning of final grades			
	Not Very Important	Important	Very Important	No Response
Reading contract made with student (Student signs an agreement with teacher to read a book)	66%	18%	3%	12%
Commercial Workbooks (Flip Flops and Zoom Shots)	39%	44%	13%	4%
Listening Activities using audio-tapes	31%	47%	14%	8%
Miscue Inventories - Oral reading tests (Student reads aloud to the teacher who records the student's miscues during presentation)	43%	45%	5%	6%
Student self-assessment checklist (Teacher and student compile a checklist to be used by the student to record development in particular areas)	49%	28%	13%	9%
Activities suggested in the Networks program	7%	64%	26%	2%
Teacher-made test on writing conventions	23%	52%	19%	6%

Activity used for ongoing student evaluation in the language arts area	Importance of activity in the assigning of final grades			
	Not Very Important	Important	Very Important	No Response
Writing conference (Teachers arrange time for interviews with individual students)	37%	41%	15%	7%
Teacher-made worksheets (Fill in the blank, multiple choice, short answer type worksheets developed by teachers)	45%	45%	7%	2%
Anecdotal records (Teacher records of factual descriptions of critical incidents that occur during a school term)	41%	37%	13%	9%
Quiz (Teacher-made test on a block of content)	17%	49%	31%	4%
Audio-tapes used to record reading experiences of individual students	55%	31%	3%	11%
Student Profile (Comprehensive collection of data on individual students - tests, projects, reading inventories, etc.)	7%	42%	50%	1%
Teacher questioning (Oral and written questions given to students specifically to ascertain and emphasize level of listening)	5%	58%	36%	2%

Activity used for ongoing student evaluation in the language arts area	Importance of activity in the assigning of final grades			
	Not Vey Important	Important	Vey Important	No Response
Teacher observation during directed reading activities and free reading time	15%	55%	29%	2%
Standardized reading tests	56%	26%	6%	12%
Dramatizations of readings	52%	37%	5%	6%

APPENDIX C

LETTER TO LANGUAGE ARTS COORDINATORS

Apartment 8
87 Waterford Bridge Road
St. John's, Newfoundland
A1C 1C7

Enclosed you will find the questionnaires we recently discussed over the telephone.

Respondents of the questionnaires must be grade four language arts teachers. Based on statistics dealing with the numbers of grade four students per school board, I have attempted to omit any multigraded classrooms. If, however, I have mistakenly assigned a school in which a class is composed of more than grade four students please, if possible, reassign the questionnaire to another teacher.

A list of schools in your district is enclosed. Only one questionnaire is assigned to the majority of schools listed. In the case of more grade four teachers than questionnaires to be administered, please alphabetize the surnames of the teachers and give the questionnaire to the teacher whose name comes first.

I would greatly appreciate receiving completed questionnaires in the return envelope provided by February 21, or as soon after as possible.

If any problems arise, please contact me collect at 722-7504.

I would like to take this opportunity to thank you for your cooperation. The positive response received during the phone calls to coordinators across the province has been overwhelming. Any information gathered through the study will be shared with you.

Sincerely,

Susan Tilley

ST/mk

Enclosure