A Master's Thesis

entitled

A STUDY OF RESEARCH CRITERIA AND TEACHER EVALUATION METHODS
FOR THE PURPOSE OF ESTABLISHING A MODEL FOR TEACHER
EVALUATION IN THE PROVINCE OF NEW BRUNSWICK

by

EARL TONER

in partial fulfillment of the requirements
for the degree of
Master in School Administration

UNIVERSITE LAVAL
FACULTE DES SCIENCES DE L'EDUCATION
DEPARTEMENT D'ADMINISTRATION ET POLITIQUES SCOLAIRES
Juin, 1986
An ABSTRACT of

A Study of Research Criteria and Teacher Evaluation Methods for the Purpose of Establishing a Model for Teacher Evaluation in the Province of New Brunswick

by

Earl Toner

Submitted in partial fulfillment of the requirements for a Master's degree in School Administration

Université Laval

June, 1986

The endeavours of this study were to examine recent research on evaluation criteria and review present practices in the methods for evaluating teaching performances in Canada and parts of the United States.

After completing a historical review of educational evaluation in the review of literature to acquire a background knowledge in teacher evaluation, various literature was examined concerning the purposes of teacher evaluation. It was concluded that evaluation of teaching personnel was done essentially for two reasons, primarily to improve the teaching-learning environment and secondly to facilitate administrative accountability responsibilities. The literature on evaluation criteria research yielded few positive positions on basic generic teaching behaviors, however, extensive research studies done by American universities in collaboration with their respective Departments of Education point toward patterns or sets of teaching skills rather than specific generic teaching behaviors as being the basis for establishing reliable evaluation criteria.

Following this, an overview was made of various research studies on present teacher evaluation practices here in Canada as well as in parts of the United States. These studies point toward similar problems in current teacher evaluation models. These shortcomings ranged from outdated evaluation criteria to lack of training and time on the part of administrators in accomplishing the task correctly. The attitudes and concerns of Canadian teachers were also reviewed in this study. Though the majority of teachers believe there is a necessity for teacher supervision, many have serious misgivings about its present practices. Unlike what is commonly believed, there presently exists strong support
among teachers for improved instructional supervision. However, educational researchers as well as teachers believe that positive steps cannot be undertaken without the support and collaboration of the teachers themselves.

Among the 32 districts in the United States that were found by a Rand Corporation study to have substantive evaluation systems, four were identified as having highly developed, somewhat unique, evaluation programs: Salt Lake City, Utah; Lake Washington, Washington; Greenwich, Connecticut; and Toledo, Ohio. Despite differences in context, closer examination of these four districts revealed that they followed certain common practices in implementing their teacher evaluation practices. All four districts recognize that the key obstacle is time, or more specifically the lack of it for observing, for conferencing with the teachers and for getting immediate expert assistance to those teachers in need of remediation. Each of the four districts has taken the proper safeguards so that remediation programs can procure the funds dictated for an adequate remediation process. Evaluator competence has been described in this study as probably the most difficult element of the process requiring two qualities: the ability to make sound judgements about teaching quality and the ability to appropriate, concrete recommendations for improvement of teaching performance. Lastly, all four district teacher organizations collaborated with the administration in the design and implementation of the teacher evaluation process.

An example of a teacher evaluation model is established in chapter 4. It utilizes as criteria a list of items pertaining to the five categories outlined in this study as a basis for formulating objectives for improving teaching skills. An alternate approach is also described which is based on evaluation through structured observation systems. Teachers may use this type of observation in conjunction with the forementioned approach or it may serve as an alternative for teachers who feel that the consensus approach of the established evaluation system is too subjective.

Research indicates that significant changes in the classroom performance of teachers occur when they are trained in the use of structured observation systems. Because of this, a certification program for beginning teachers in New Brunswick similar to that being implemented by the state of Georgia would not only benefit the beginning teacher but would also be an excellent starting point for improving both teacher evaluation methods and supervisor proficiencies in teacher evaluation throughout the province.
ACKNOWLEDGEMENTS

A study such as this cannot be undertaken alone and sincere appreciation must be expressed to those individuals who helped make the completion of this research possible. I am greatly indebted to my research director, Mr Hermann Girard, whose constant encouragement, guidance and thoughtful suggestions throughout this study served as sources of inspiration. To Barry Lydon, the P.D. director with the Department of Education in New Brunswick, I extend sincere thanks for making me realize that taking a year to study, though a difficult decision after years of teaching, is an experience to be savoured for years to come.

A special note of appreciation must go to the members of the Faculte des sciences de l'Education, as well as to Mr Real Gauthier and Mr Ulysse Roy at the Laval library for their special efforts on my behalf.

Finally, I wish to offer my deepest gratitude to my wife, Irma, without whose adept proof-reading skill, advice, and unwavering support throughout, this endeavour may never have reached its culmination. To our children Eric, Elisabeth, and Colleen, goes a special thanks for being a source of pride and motivation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>THE PURPOSE OF THE STUDY</td>
<td>6</td>
</tr>
<tr>
<td>ASSUMPTIONS</td>
<td>7</td>
</tr>
<tr>
<td>METHODOLOGY</td>
<td>8</td>
</tr>
<tr>
<td>LIMITATIONS</td>
<td>9</td>
</tr>
<tr>
<td>DEFINITION OF TERMS AND EXPLANATION OF ABBREVIATIONS</td>
<td>10</td>
</tr>
<tr>
<td>REVIEW OF LITERATURE</td>
<td></td>
</tr>
<tr>
<td>a) History of Teacher evaluation</td>
<td>13</td>
</tr>
<tr>
<td>b) Purposes of Teacher Evaluation</td>
<td>26</td>
</tr>
<tr>
<td>2. RESEARCH ON EVALUATION CRITERIA</td>
<td>30</td>
</tr>
<tr>
<td>Overlapping Categories of Evaluation Criteria</td>
<td>37</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3. CURRENT EVALUATION PRACTICES</td>
<td></td>
</tr>
<tr>
<td>a) An Overview of Teacher Evaluation in Canada</td>
<td>70</td>
</tr>
<tr>
<td>b) A Review of Some Evaluation Practices in the U.S.</td>
<td>91</td>
</tr>
<tr>
<td>4. ESTABLISHING A MODEL</td>
<td>106</td>
</tr>
<tr>
<td>5. CONCLUSION AND RECOMMENDATIONS</td>
<td>128</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>148</td>
</tr>
<tr>
<td>APPENDIX A - LETTER FROM NEW BRUNSWICK TEACHERS FEDERATION</td>
<td>156</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Basic Purposes of Teacher Evaluation</td>
<td>27</td>
</tr>
<tr>
<td>2.</td>
<td>Desired Role of Provincial Government and Ministry of Education Regarding Teacher Evaluation</td>
<td>85</td>
</tr>
<tr>
<td>3.</td>
<td>Desired Role of Local Board, Administration, and Teachers in Teacher Evaluation</td>
<td>87</td>
</tr>
<tr>
<td>4.</td>
<td>Type of Evaluation for Improving Instruction</td>
<td>88</td>
</tr>
<tr>
<td>5.</td>
<td>Type of Evaluation Perceived as Valid Judge of Current Teacher Performance</td>
<td>89</td>
</tr>
<tr>
<td>6.</td>
<td>Acceptable Criteria for Teacher Evaluation</td>
<td>112</td>
</tr>
<tr>
<td>7.</td>
<td>Categories for Flander’s Interaction Analysis</td>
<td>122</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Variables Involved in Determining Teacher Competencies</td>
<td>39</td>
</tr>
<tr>
<td>2.</td>
<td>Types of Test Questions From Which Teachers Can Choose in Making Their Own Tests</td>
<td>60</td>
</tr>
<tr>
<td>3.</td>
<td>An Example of a Flander's Interaction Analysis</td>
<td>123</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

The further I delve into literature on teacher evaluation and supervision the more I realize that the endeavours over the last few decades in New Brunswick are not unlike what has transpired throughout most of the communities in North America. Some years ago, most evaluation systems in education were inserted into the educational systems to show some form of accountability. The idea was generally initiated from some political sphere and thus the process proceeded in a rather inept manner - poorly planned, haphazardly staffed, widely misunderstood, and insufficiently funded. The results were dubious at best and likewise skeptically received and even ignored completely most of the time.

Over the last few years, we've seen an unprecedented wave of activity directed at improving classroom instruction with the latest impetus coming from the National Commission on Excellence in the United States. It published a report called "A Nation at Risk: The Imperative for Educational reform". Several of the commission's recommendations concerned with teaching would require teacher evaluation: "persons preparing to teach should be required to meet high educational standards, to demonstrate competence in the academic discipline(...) . Salary, promotion, tenure, and retention decisions should be tied to an effective evaluation system that includes peer review so that superior
teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated.¹

There is a growing realization that seems to reverse the educational policy trends of the past decades. Teacher-proof curriculum, test-based instructional management, and student achievement testing initiatives were all based on the premise that education could be improved without improving the quality of teaching. But increasingly, educational policymakers are beginning to perceive that better teachers and improved teaching techniques are the keys to better education.

Canada, with its forceful provincial teachers unions (N.B. not excepted) has been reluctant to pursue teacher evaluation as a viable concept for improving teacher performance in the classroom and as a means of elevating the quality of education provincially. Evaluation systems operating in New Brunswick, according to an N.B.T.A. deputy executive director, "have been designed for the specific purpose of documenting inappropriate teacher behaviours. Most models force evaluators to collect data and document activities that provide for a written record of a teacher's performance(...). It is not difficult to ascertain why many teachers view evaluation with trepidation."²

Alberta, however, is beginning to restructure its policies on teacher evaluation. Studies conducted in 1977 (Holdaway, 1977; Reikie, 1977) found that 83% of the 114 Alberta school jurisdictions surveyed evaluated teachers new to the system, whereas only 53% of tenured teachers were evaluated on a scheduled basis (e.g. once every three years). The implications of this study suggest that although policies may require teachers to be evaluated according to a fixed schedule, in practice this may not occur and thus it is entirely possible that a new teacher may be granted tenure without a formal evaluation, and a teacher with tenure may teach for many years without ever being formally evaluated. These findings are not the exception in Canada as most studies would indicate similar results. Alberta, as previously mentioned, is not maintaining the status quo. The Keegstra affair at Eckville High School in Dec. 1982, amply demonstrated to the people of that province what can happen when ineffective teacher evaluation systems persist. They implemented a new policy of teacher supervision and evaluation in five secondary schools in Lethbridge as a pilot in 1984. Already administrators and teachers perceive beneficial changes in improved conferencing and communication skills, data collection.

---


and feedback by supervisors.  

Unlike in Canada, the new concern for the quality of education and of teachers in the U.S. is being translated into merit pay, career ladder, and master teacher policies that presuppose the existence of effective teacher evaluation systems. In light of this, most American school districts will be reassessing their teacher evaluation practices. Investments in American universities' Faculties of Education have increased enormously in areas concerned with teacher competencies and measurement-based teacher evaluation. For example, Stanford university spent over 15 million last year alone on the Beginning Teacher Evaluation Study (BTES), a continuing research project that began fifteen years ago. There seems to be an infinite number of initiatives such as this being undertaken by American universities for improving teaching skills. Certainly American school districts will be closely scrutinizing any programs that lead to promising results.  

Already, they have excellent precedents to fall back on. There presently exist school districts south of the border that have fully implemented effective teacher evaluation systems which suit the educational goals and community values of the school districts. The teacher unions in these school districts were instrumental in developing the evaluation criteria and

procedures; and have said that "if the administration uses the procedures correctly, we are not going to be against them." 6

Within the near future, school districts and provincial sectors in education will be taking a new hard look at its teacher evaluation policies, not unlike what Alberta and the majority of American school districts are doing presently.

Educational perspectives and decisions look both forward and backward. (...) Understanding the supervisory process in education, therefore, requires knowledge of the way in which the process has evolved within the institution of American public education. This background is an indispensable element in analyzing present conditions and constructing plans for the future.

The necessity of this study could not be more aptly described. Surveying the present while attempting to implement changes for the future: it's within this context that I intend to pursue this research.

---


THE PURPOSE OF THIS STUDY

This study will attempt to present a general review of recent research on evaluation criteria; then we'll examine studies of teacher evaluation systems within different provinces in Canada as well as parts of the United States. By drawing on the experience of these research studies, a teacher evaluation model will be established. The reviewed evaluation criteria and literature on teacher evaluation systems presently in use within these two countries will be used as criteria for defending the different aspects of such a model. This teacher evaluation model, it is hoped, would be used for the improvement of instruction of teachers in School District No. 50 in New Brunswick.
ASSUMPTIONS

1. That review of the present state of evaluation systems in Canada and parts of the United States could elicit valuable methods and criteria for establishing an evaluation model adaptable to the school system in New Brunswick.

2. Similarly, by reviewing recent research in teacher evaluation data can be found for improvising a model selected for implementation in New Brunswick.

3. It is presumed that by surveying specific school districts in the United States that are knowledgeable in teacher evaluation useful techniques could be found for implementing an evaluation system in New Brunswick.
METHODOLOGY

The Major focus of this research will be on the development of an evaluation model for use in assessing teaching performance. Contrary to earlier practice, where evaluation was viewed as the sole responsibility of the administrator, this study seeks an alternative approach by studying research which involve teachers in the development of the evaluation instrument and anticipates their participation in the evaluation process.

In this study we will briefly review present teacher evaluation practices in Newfoundland, New Brunswick, Quebec, Ontario, and Alberta in an attempt to get a cross-section of evaluation systems presently being employed in Canada. Next a survey of recent Canadian and American research and innovations in teacher evaluation will be made in an attempt to establish the most valid, reliable, and adaptable method of teacher evaluation for New Brunswick. Finally, it is hoped by studying specific districts in the United States which are experienced in this type of teacher evaluation that practical methods for implementing a formative teacher evaluation system in New Brunswick can be found.
LIMITATIONS

The review of evaluation systems in Canada will consist of facts established from the most recent important publications from different provinces and review articles from Canadian periodicals and other available publications. The criteria established from the United States will be limited to research articles, periodicals, and publications which are made available by associations such as Phi Delta Kappa, National Education Ass. (N.E.A.), ERIC, as well as the Rand Corporation. More specific documents will be solicited from selected research centers as needed.
DEFINITION OF TERMS AND EXPLANATION OF ABBREVIATIONS

Collective agreement: A document containing the written agreements on the rights, recognition, and procedural aspects on pay, conditions of employment, types of teacher contracts, etc. between an employer and employees (teachers), the latter usually being represented by a labour union.

Evaluation: A formalized process of gathering data for the purpose of judging the value or worth of some product or process relative to some standard(s).

Feedback: Information on progress of teaching and learning provided through various methods of assessment.

Formative evaluation: assessment which seeks to improve the learning/teaching system by feedback of information from test results which can illustrate the effectiveness of teaching methods, or highlight learning difficulties.

Generic teaching behaviors The teaching behaviors which are considered to be effective regardless of the setting.

Master teacher: An experienced teacher recognized as qualified to give guidance to student teachers and interns.

Mastery learning: A school of thought which rest on the assumption that mastery of a topic, field of human knowledge or human behavior is theoretically possible for all individuals provided that each learner is given the optimum quality of
instruction which is appropriate for his/her particular make-up and that each learner is given the time that he/she needs.

NASSP: (National Association of Secondary School Principals) A national U.S. organization for secondary school principals, assistant principals, administrators, and college professors teaching courses in secondary education. Founded in 1916, the NASSP has approximately 30,000 members.

N.E.A.: (National Education Association, formerly National Teachers Ass.) An American professional organization and trade union for elementary and secondary school teachers, college and university professors, administrators, principals, counsellors and others interested in education.


Peer evaluation: In teaching refers to the use of professional colleagues for the assessment of teacher competence (for such summative evaluation purposes as promotion and tenure) and for the provision of feedback to the teacher in order to improve teacher performance (formative evaluation).

Pre-classroom conference: Process by which the teacher and supervisor mutually establish objectives for the lesson, thus providing the teacher input on the criteria by which he will be evaluated.
Student teaching: Practice teaching done as part of course work for a degree in education. It is usually undertaken in an affiliated school under the supervision of a cooperating teacher.

Summative evaluation: Evaluation at the conclusion of an educational plan or activity to determine the effectiveness of that activity.

Teacher: means a person who holds a certificate of qualification issued under the Department of Education Act of that province or state and is employed by a school whose jurisdiction falls within that province or state to instruct students.

Tenured teacher: Means a teacher who has been granted a continuous contract pursuant to the Schools Act and/or collective agreement of the province or state of which he is employed.
REVIEW OF LITERATURE

a) History of Teacher Evaluation

Teacher evaluation has evolved from the relentless probing in America for efficiency. As early as 1887, Joseph Rice, an American physician, conducted what is generally recognized as the first formal educational program evaluation in America. Using a list of spelling words as a testing instrument he carried out a comparative study on the value of drill in spelling instruction involving 10,000 students across a number of Indianapolis school districts. Using tests scores as his criteria measures he carried out this research in the hopes of verifying the inutility of spelling drill as a technique for learning. The results significantly verified his hypothesis. He found no significant learning gains between systems which spent 200 minutes a week studying spelling and those which spent as little as 10 minutes a week. His research was published in the 'forum' in 1896 under the title 'Obstacles to Rational Educational Reform'. At the 1897 meeting in Indianapolis of the Department of Superintendence of the National Education Association, Rice, conducting a round-table discussion, startled those present by asking "how any one could tell whether at the end of eight years' training children who had had forty minutes a day in spelling instruction were better spellers than those who had had only ten minutes a day." Vigorous opposition developed against the implications of Rice's

question and the testing movement in education lay dormant for more than another decade.

While Rice's articles were appearing in the 'Forum', however, a student in Columbia University, E. L. Thorndike, became impressed with the suggestions raised by Rice's work. Thinking along this line of evaluation, Thorndike tested the doctrine of formal discipline which upheld the value of studying certain subjects because they would provide skills or principles useful in other subjects. This doctrine of formal discipline was defended by teachers of Latin, Greek, and Mathematics on the grounds that their subjects had great value in training the mind. Thorndike, an expert in stimulus-response research, was well aware that the specification of similarities and differences between stimuli and responses, from setting to setting, was very difficult, and decided to study this transfer relationship in schools. Thorndike performed a series of experiments to see whether the practice on one test, such as crossing out all the E's on one page, would influence performance on other tests, such as crossing out the S's. The results revealed that except for shared perceptual abilities or motor behaviors, or whatever was common to the two tests, no general transfer was present. Thus, positive transfer, such as transfer from playing a piano to typing, takes place only insofar as eye-hand coordination skills are present for both tasks. Because of results like this Thorndike argued against the doctrine of formal discipline but

was not able to further substantiate the results until the 1920's with the help of I.Q. tests. He studied more than 13,000 school students from beginning to the end of the year to determine the general transfer effects of the various courses on their gains on intelligence tests. The differences in I.Q. gains for different courses were so small that it did not matter what was taken. Thorndike also concluded that those who have the most intellectual ability to begin with gain the most during the year. Whatever studies they take will always seem to produce large gains in intellect. The theory of formal discipline, for better or worse, slowly faded after these studies were presented.

Soon to follow were innovators such as G. Stanley Hall, who developed the use of questionnaires in educational research. Hall conducted research for the newly created Federal Bureau of Education (later the U.S. Office of Education), in curriculum revision and assistance to teachers. As a forerunner of formative evaluation, he investigated, in 1907, what children knew about common things, places, and objects in their everyday life. He found, for example, that his subjects knew that milk came from cows, but did not know that animals were the source of leather. Hall's investigations led to the idea that teachers must not take for granted what children know, or assume that children necessarily perceive relationships among the facts they already know. Relationships would also have to be taught.

10. ibid, p. 357
implications of this research in today's society would be even more profound since even fewer children live in rural areas.

Early Misuses of Educational Evaluation

Not unlike today, as educational services grew so did the demands to justify expenditures. Evaluation represented a response to these demands. What return are the taxpayers getting for their investment in the schools? Such early evaluations constituted little more than descriptions, countings and recountings of service rendered. For example, in 1909 Leonard Ayres surveyed school records and government statistics, counted numbers of children promoted per dollars of expenditures, and drew the conclusion that schools were filled with "retarded" children (i.e. children who were overage for their grade levels). He neglected to observe that the schools were crowded with newly arrived immigrant children and that the situation was hardly surprising and certainly no reflection on the schools or on the students. Such evaluations, used in the search for answers about the worth of schools, were frequently erroneous or misleading.

"Testing boom" period of the 1920's and 1930's

The dawning of World War I brought into focus new developments both in America and in France. Attempts to measure general intelligence originated with Binet and Simon in 1905. Their scale was originally devised for the purpose of selecting

12. ibid p. 142
mentally retarded pupils who required special instruction. However, other basic ideas of interpreting the relative intelligence of different children at any given chronological age by the number of tests of varied types and increasing difficulties were re-embodied in 1911 revisions of the Binet-Simon scale. American adaptations of these pioneer scales appeared in the "Group intelligence tests" in the form of "Army Alpha" used for the measurement and placement of army recruits and draftees during W.W.I. 13

The measurement of aptitude, or those potentialities for success in an area of performance, has been tied up with intelligence testing for many years. Aptitude tests for telephone girls and streetcar motormen in 1913 were followed by tests of mechanical aptitude, musical aptitude, art aptitude, clerical aptitude, and aptitude for various subjects of the high-school and college curricula prior to 1930. Splitting of total mental ability into a general factor and many specific factors accounted for the fact that aptitude tests are frequently called "specific intelligence tests". 14 It is worthy of noting that industry adapted intelligence and personality testing before its use was seen as appropriate for education; that precedent continues to this day.

The Tylerian Era

The work of Ralph W. Tyler during the 1930's laid the foundation for educational evaluation as we know it today. During the great depression, schools as well as other institutions stagnated. Tyler proposed an experiment in 1932 in which over 300 colleges agreed to waive their traditional entrance requirements for graduates from about 30 progressive high schools. The high school and college performance of students from these secondary schools would be compared to the high school and college performance of students from a group of traditional secondary schools.15

The Eight-Year Study, as it came to be known, was seen as having clear-cut advantages over previous approaches. A Tylerian evaluation involved comparisons of outcomes with objectives; it need not provide for costly and disruptive comparisons between experimental and control groups, as were required in the comparative experimental approach that Rice had used. It concentrated on learning outcomes instead of organizational and teaching inputs, thereby avoiding the subjectivity of the professional judgment or accreditation approach.16

As I began to work with the instructors at the Ohio State University, it was clear that they needed tests that would inform them about what students were learning and where they were having difficulties. Sorting students was not the purpose; instead, their concern was to improve the curriculum and the instructional program. I realized that test theory developed for purposes of sorting and based on measures

15. Madaus, op. cit. p. 9
16. Rosenthal, op. cit. p. 143-144
of individual differences would not produce the kinds of evaluation instruments needed. Hence I developed a procedure based on theories of instruction and learning. 17

Meanwhile with the dawning of another world war, expansion of evaluation systems was again in order. Novel instructional methods were devised and repeatedly evaluated in order to improve their effectiveness. The necessity for Tylerian "congruence between objectives and outcomes" blended well with the new developments and expansion of standardized testing necessitated by the war.

The late forties saw many new nationally standardized tests being published and schools purchased these tests in great quantities and also subscribed heavily to machine scoring and analysis services that the new technology made available. However, achievement tests had already been used in the 1913 School Survey of New York City, and by 1919 similar tests were used on an elaborate scale in state-wide surveys by the Virginia Educational Commission. Intelligence tests were also included in the Virginia State Surveys. At this time more extensive standardized tests of aptitude, scholastic achievement, and personality were being administered to students by the millions. Somewhere between thirty and forty million standardized tests had been sold during one of these years. 18 The testing movement

recieved another boost in 1947 when E. F. Lindquist, Ralph Tyler and others helped establish the Educational Testing Service in Princeton N. J. They began devising tests for admission to undergraduate, graduate and professional schools with the scholastic aptitude tests (SAT), National Teacher Examination (NTE), and the College Entrance Examination Board (CEEB). As external tests they served another important function as most states maintained one or more statewide testing program, the oldest and best known being the New York State Regents' examinations.19

Limitations of Evaluation Instruments Becomes Apparent

The growth of standardized achievement tests and the implementation of the Tyler approach to educational curriculum brought broad and innovative views to educational objectives and curriculum development but soon strategies required re-examination. For example, standardized tests had been designed to rank students of average ability; they were of little use in diagnosing needs and assessing any achievement gains of the disadvantaged whose educational development lagged far behind that of their middle class peers. Further, there was a dearth of information about the needs and achievement levels of disadvantaged children that could guide teachers in developing meaningful behavioural objectives for this population of

19. ibid, p. 6
learners. In response to this dilemma, evaluators continued to collect large amounts of data to verify "congruence between objectives and outcomes" and to justify the need for expansion or for broad, new programs. Yet there is little evidence that these efforts were useful in improving the areas in education that needed it the most. In fact the expansion of standardized testing outstripped any new testing approaches in evaluation. The late 1950s saw a growing disquiet with evaluation efforts in education due to consistent negative findings in attempts to respond effectively to new challenges. It was becoming apparent to leaders in educational evaluation that their work and their results were neither particularly helpful to curriculum developers nor responsive to those who wanted answers for reforming the social and educational inequalities in education.

Abrupt changes in the political sphere was soon to further expose the deficiencies in evaluation. It was in 1957 that Sputnik I went up and shocked the world. Immediately the Americans instituted new educational programs in mathematics, science, expanded counseling and guidance services. Soon to follow were funds to evaluate these curriculum development efforts. A few years later President Johnson's War on Poverty was launched with billions of dollars poured into reforms aimed at equalizing opportunities in health, social, and especially

20. Madaus, op. cit. p. 13
educational services. An expanding American economy at that time made these programs a priority.

The stagnation in evaluative techniques was further exposed as the same somewhat outmoded approaches to evaluation discussed previously were employed during this period. The Tylerian approach was attempted to define objectives for the new curricula and to assess the degree to which objectives were being realized. Secondly, new nationally standardized tests were created to better reflect the objectives and content of the new curricula. One of these was Project Talent. It consisted of a battery of 28 aptitude and achievement tests. They were administered to a carefully drawn sample of some 440,000 students in grades 9 through 12 in 1353 public, private, and parochial schools in all parts of the country. Each class tested was followed up one, five, ten, and twenty years after high school to obtain information on success and satisfaction in chosen careers and to determine the relationship of this success and satisfaction to test scores and background information.  

Fears that money invested might be wasted if appropriate accountability requirements were not imposed soon exposed the inadequacies of the tools and strategies employed by evaluators for this tasks. Tyler recognized and encouraged differences in objectives from place to place. Yet this model did not adapt well to nationwide standardized-testing programs which had to overlook particular objectives in favor of objectives stressed by the

---

22. V. Noll, op. cit. p. 6
majority of districts. The further apparent failures of attempts to isolate the effects of projects through the use of experimental/control group designs was due primarily to the inappropriateness of the evaluation criteria. Anyway when the findings of "no results" were reported, as was generally the case, the disheartened raised serious questions about evaluation in general and certain methodologies in particular.²³

Couple these events with the findings of Coleman's famous study "Equality of Educational Opportunity" in 1966. This report received particular notice because it concluded that "schools bring little influence to bear on a child's achievement that is independent of his background and general social context."²⁴ The growing disquiet with evaluation efforts and the growing negative findings called for the development of new theories and methods of evaluation as well as for new training programs for evaluators. But it would not suffice to conclude from all this that educational evaluation as an entity has failed to grow. It might be more appropriate to conclude that these negative events have simply set the stage for further growth.

Educational Evaluation: Strong Indication of Growth

Against such an unfavourable backdrop, progress seems to be remarkable indeed. Unlike 15 years ago, educational journals today deal consistently with evaluation, continually finding

²³ Madaus, op. cit. p12
research data pertinent to improving the educational field. The influx of books on educational evaluation especially from research centers can only make one wonder if the problem today is going to be keeping abreast with it. Popham describes this new era as such:

'Exciting' may be too strong a term to characterize the field of education these days, but perhaps not. For a variety of reasons, such as legislative mandates, public dissatisfaction with schools, and the influence of responsible educators, we find that educational evaluations are being carried out not only with increasing frequency, but also with increasing excellence. Many first-rate scholars are turning their attention to the study of educational evaluation as a field of disciplined inquiry. Many able people are choosing educational evaluation as their area of professional specialization. Many books and articles specifically addressed to the topic of educational evaluation are now being published. Many people, including legislators, lay citizens, and all sorts of educators, are beginning to believe that the conscientious application of evaluative procedures will actually enhance the quality of American schooling. And all of this sometimes furious activity has taken place in less than a decade. No, perhaps it is not too lavish to depict the current status of educational evaluation as being genuinely exciting.

As exciting as this period may be, the surging demand for accountability in education has shifted from broad issues of finance and program management to specific concerns about the quality of classroom teaching and teachers.

Improving teacher quality was the most frequent response to the 1979 Gallup poll's question on what public schools could do to earn an "A" grade, beating by large margins such highly touted

reforms as emphasizing basics, or improving school administration, lowering class size, or updating the curriculum. Does this statistic answer for the recent surge of texts as "Handbook of Teacher Evaluation" or closer still the title of my thesis for that matter?

Authors of many recent articles have cited the existence of tension that has come to bear because of public pressures in education. This tension may be a reflection of the different viewpoints among evaluation purposes. Knapp's (1982) articulation of various stakeholders' perspectives is useful.

Teachers have a stake in maintaining their jobs, their self-respect, and their sense of efficacy. They want a teacher evaluation system that encourages self-improvement, appreciates the complexity of their work, and protects their rights. Principals have a stake in maintaining stability in their organizations, allowing them to respond to parental and bureaucratic concerns for accountability while keeping staff morale intact. They want an evaluation system that is objective, not overly time-consuming, and feasible in the organizational context. Parents and public officials have a stake in the "bottomline"—the effects of teaching on student teacher effectiveness, and that guarantees appropriate treatment of children in classrooms.

These differing views make choices about teacher evaluation processes difficult. Before identifying basic issues in the area of teacher evaluation and developing ideas that might lead to improvement of teacher evaluation practices in the schools of New

b) Purposes of Teacher Evaluation

Recent studies in the United States and Canada relate to varying degrees the purposes of teacher evaluation. For example Jason Millman (1981) purports that:

Over a dozen reasonably distinct purposes for teacher evaluation have been suggested, such as improving teacher performance, aiding administrative decisions, guiding students in course selections, meeting state and institution mandates, promoting research on teaching and the like.28

Dale Bolton (1973) contends that if teacher evaluation is to be successful, it is crucial that teachers understand the purposes for which they are being evaluated. In support of this argument he identified the following six purposes:

1) Improvement of instruction
2) Rewarding superior performance
3) Modification of assignment
4) Protection of individuals and the organization
5) Validation of the selection process
6) Promotion of individual growth and self-evaluation29

---

29. A. Korshalla, op. cit. p. 33
Unlike during the 70's, the purposes of evaluation in the early 1980s appear to have shifted from an individualistic stance of teacher improvement to the accountability and staff development categories of public scrutiny. A. Wise in a 1985 article uses a matrix to establish a distinction between purposes and levels of evaluation.

As indicated in Table 1, teacher evaluation may serve four basic purposes. The Matrix artificially represents these purposes and levels of decision making as distinct. In fact, teacher evaluation may apply to small or large groups of teachers (rather than simply individuals or whole schools) and may represent degrees of combined improvement and accountability concerns (as when promotion decisions are linked to improvement efforts).

**Table 1**

<table>
<thead>
<tr>
<th>Purpose/Level</th>
<th>Individual</th>
<th>Organizational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement (formative information)</td>
<td>Individual staff development</td>
<td>School improvement</td>
</tr>
<tr>
<td>Accountability (summative information)</td>
<td>Individual personnel (job status) decisions</td>
<td>School status (e.g., certification) decisions</td>
</tr>
</tbody>
</table>

However we must consider another element which could help explain some of the dilemma facing evaluation systems. A study done by Wood and Pohland (1979) examined rating forms from a large number of randomly selected districts across the United States and found that,

Items related to the instructional role - that is, those teacher behaviors related directly to instruction - accounted for only 28% of the items on the checklists used. About 30% of the items related to personal characteristics such as appearance, voice, and the like, characteristics for which there is no direct relationship demonstrable in research in terms of student gains. Other items on rating forms were related to organization membership roles, professional roles, social roles, and administrative roles. 31

Wood and Pohland advance an interesting explanation for the gap between what is commonly held to be the purpose of teacher evaluation and the actual practice of it. They speak of the "hand of history". They analysed rating forms from the past and concluded:

...that emphases in the past on efficiency (scientific management) and personal characteristics (human relations) have 'trapped' school systems into a particular approach to evaluation, with resulting discrepancy between the rhetoric and the reality. 32

A study carried out in Ontario in 1972 by Palmer, Musella, and Lawton surveying 88 of 126 school districts in that province revealed a similar situation, "The process and the instruments used in teacher evaluation bear little relationship to the purpose of improvement of instruction." 33 However the results of this study showed strongly that the stated purposes of teacher evaluation "were to help the teacher improve classroom performance and to identify areas in the teaching activities needing improvement." 34

32. ibid, p. 4
33. ibid, p. 3-4
34. ibid, p. 3
Others such as Howsam (1983) remain true to the trend of recent years favoring the second purpose to teacher evaluation:

Society has passed to state and local authority responsibility for certain aspects of the formal education of children, youth and adults. This responsibility in turn is passed ultimately to teachers. Every delegation of responsibility carries with it the implication of accountability. Evaluation is the basis of accountability.

Though it appears that the literature on purposes of teacher evaluation may lack consensus as to the objectives of educational evaluation, the majority of research tends to lean toward at least two essential purposes for evaluating teacher performance:

1) to improve the teaching-learning environment
2) to facilitate accountability responsibilities

Much like one could not ignore the work of Jean Piaget while reviewing research in childhood development, neither can H. E. Mitzel be overlooked here. In 1957 he proposed a classic formulation of teacher evaluation consisting of three different classifications of criteria which he termed: (1) presage, (2) process and (3) product. By 'presage', Mitzel refers to teacher characteristics that are present before the teaching act begins: patience, intelligence, affection for children, achievement in university courses, and personal characteristics such as appearance and voice. 'Process' criteria relates to what the teacher actually does and says in the classroom. Emphasis here is on observable teacher behaviors such as method of instruction, interaction patterns among students and teacher, and verbal behaviors such as the amount of teacher time spent in presenting ideas as compared with that spent answering students' questions. 'Product' criteria refers to the results of teaching such as: pupil performance on achievement tests, students' attitude toward school and learning, or congruence between a set of objectives toward which teaching is directed. These effects are referred to as student gains, student growth, or simply student changes, but they all involve measurement of change in student behavior, some of which researchers logically attempt to attribute to the
influence of individual teachers. The distinctions made by Mitzel in his presage-process-product classifications of criteria has provided us with a useful method of examining teacher effectiveness in terms of before, during and result of teaching.

Methodological advances in educational research in recent years have brought into focus another subdivision. Dunkin and Bibble (1974) noted that classroom research involved four broad categories of variables, noting the three stipulated by Mitzel as well as another: context. Context was described as those elements of research which pertained to effects of grade level, subject matter, immediate instructional objectives, and other particulars of the situation and setting. Most literature on evaluation criteria have emphasized the presage or process criteria as opposed to product criteria partly because of the inability to measure precisely the gains which can be attributed to teaching performance. However, Popham (1975), for example, makes a strong case for evaluating teachers by their contribution to the performance of students as measured by standardized test scores rather than by the use of teacher process criteria.

Although the history of using norm-referenced achievement tests is bleak, we have yet to probe vigorously the utility of employing learner postinstruction performance on criterion-referenced measures as a measure of teachers' instructional skill. Surely this will be done in the next few years.

37. J. E. Brophy, "Teacher Behaviors and Its Effects", in Journal of Educational Psychology (Dec./79 Vol. 71 No. 6) p. 738
Millman (1981) also leans towards student achievement as a measure for assessing the performance of classroom teachers. "Teacher evaluation can be used either for improving instruction (formative evaluation) or for making decisions about teacher status (summative evaluation). Knowledge about student achievement, appropriately obtained, can be useful for either purpose." 39

Nevertheless, the appropriateness of student results for evaluating teachers is being hotly debated. The vast majority (89%) of the National Education Association members do not consider scores on achievement tests a valid measure of teacher effectiveness. 40 The dissatisfaction of these members seem well founded if one views some of the "inadequacies" stated by R. Stager of standardized tests for evaluation as cited by G. Sax in "Standardized Tests in Evaluation".

1. "Standardized tests ordinarily have a low degree of overlap with the actual objectives of instruction at any time or place." As Stager states, standardized tests do not measure the unique objectives of any specific program; they take special care not to do so.

2. "Standardized tests are not useful as aids in the planning of instruction, particularly the specification of instructional objectives." One does not scrounge through standardized tests in the hope of discovering what should be included in the...

40. A. Wise, op. cit., p. 66
3. "Standardized tests often require skills or aptitudes that may be influenced to only a limited degree by experiences in the classroom." Only teacher-made tests can effectively measure the content or effectiveness of a specific lesson.

4. "Standardized tests do not indicate the extent to which individuals or groups of students have mastered the spectrum of instructional objectives." Manuals that accompany these tests warn against interpreting raw scores because differences in means and standard deviations make such comparisons unintelligible.

5. "The use of standardized achievement tests in the sociopolitical context of 'evaluation' and 'accountability' has and will lead to corruption and dishonesty among educational professionals and to the further erosion of public trust in the schools and the people who run them." This accusation is more properly a reflection on the honesty and moral turpitude of educational professionals than it is a criticism of standardized tests. Any test, including those that are criterion-referenced, can be improperly and even dishonestly used. When teachers, administrators, or performance contractors are required to meet unreasonable goals or suffer serious consequences, the probability that tests will be misused increases.

Medley, Coker, and Soar (1984) also speak unfavorably about the use of achievement tests as a basis for evaluating teachers. They cite three problems: pupil variability, the regression effect, and the limitations of achievement tests presently available. The authors noted several studies which, "estimated stability coefficients of mean gains in a teacher’s class as a measure of the competence of the teacher (Rosenshine, 1970; Brophy, 1970). The median value of all such coefficients is around 0.30." Measurement experts, according to these authors, generally agree that any score that is to be used to make decisions about individuals needs a reliability of at least 0.90 to be usable. In concluding they state:

Deciding which teacher is the best — which to hire, which to fire — by a method that is little better than chance might seem worthwhile for other reasons. To be able to say you are using the best available method, however imperfect, may be politically advantageous. But before deciding to do so it is important to know that the use of the procedure cannot do great harm. If using pupil achievement as a basis for teacher evaluation worked successfully it would probably reduce the quality of teaching instead of improving it.

If teacher evaluation in the product domain seems plagued by its lack of appropriateness, research on "direct instruction" seems no less encouraging. Considerable research has been done in an effort to determine precisely those behaviors which characterize an effective interaction between teacher and student, an interaction which results in student gains and hence improvement of instruction. Efforts to link such specific teacher

43. ibid, p. 36
characteristics or teaching behaviors to 'discrete' sets of behaviors have encountered inconsistent and sometimes contradictory findings that tend to undermine faith in the outcomes of simple process-product research. The most extensive process-product study of teacher effectiveness, the Beginning Teacher Evaluation Study conducted for California's Commission for Teacher Preparation and Licensing, concluded "that linking precise and specific teacher behavior to precise and specific learning of pupils (the original goal of the inquiry) is not possible at this time." 44

These findings are consistent with conclusions obtained by Banks (1983) while studying the Beginning Teacher Evaluation Instrument for the Kentucky Department of Education:

Popular observational instruments used in teacher evaluation largely consist of specific teacher behaviors believed to represent "generic" teaching skills required of all teachers. The use of such instrumentation suggests that educators not only know what behaviors are effective but how much of such behaviors must be demonstrated to constitute "good teaching." However, this knowledge, that would seem prerequisite to the use of such techniques, has not of yet, been empirically generated and validated." 45

Both studies found little evidence that single teaching performance variables can be identified as essential for effective teaching. There are other interesting, though not any more applicable, research having to do with discrete teaching behavior variables. Peterson & Kauchak (1982) found that teaching

behaviors that have sometimes been deemed effective often bear a distinctly curvilinear relation to achievement: a behavior that is effective when used in moderation but can produce significant negative results when used too much or in the wrong circumstances. Furthermore, some recent research suggests that desirable 'affective' outcomes of education such as independence, curiosity, and positive attitudes toward school, teacher, and self seem to result from teaching behaviors that are different from those prescribed for increasing student achievement on standardized tests of cognitive skills.

As data accumulates, the influence of circumstance and/or environment is being recognized as more and more important. In fact, although there may be a few truly universal instructional principles (i.e., match level of instruction to student ability), there do not appear to be any universal teaching competencies (i.e., specific behaviors such as praising or asking higher level questions) that are appropriate in all teaching circumstances (Brophy & Evertson, 1978; Beginning Teacher Evaluation Study). Teachers must not only master particular skills but know when to use them.

46. A. Wise, op. cit. p. 296
After querying research on evaluation criteria, the question still remains: what constitutes effective teaching? What can we base ourselves on to evaluate it? Brophy (1977) titled a journal article "A Good Teacher is Harder to Define than to find". More than ever we appreciate his dilemma; but our goal still remains: to critically review evaluation criteria research in the hopes of upgrading teacher performance standards through the inclusion of criteria in an evaluation model that is research-based and defined by experts in the field.

Overlapping Categories of Evaluation Criteria

In spite of the inconsistency which pervades the evaluation criteria, there is some reason for optimism. Over the last ten years researchers have begun compiling research studies in attempts to organize the dimensions of teaching into broad categories. For example Overt (1975) critiqued a report in a NASSP Bulletin on a three year analysis of 39 research studies by Arvil Barr'. Barr's conclusions, as cited by Overt, were that no universally accepted criteria exist for determining teacher effectiveness. However, after analyzing 209 teacher rating scales, he determined that rated items fell into 10 categories for judging teacher competency. They were: instruction, classroom management, professional attitude, choice of subject matter, cooperation, personal habits, health, discipline, appearance of the room, and personal appearance. 49 C. R. Inglis (1970) reported

on teacher evaluation in 38 states and found that the rating scale items could be placed into three general categories: professional relations and attitudes, teaching techniques, and personal characteristics of the teacher. But each group included 7, 10, and 11 subgroups respectively. Another study working along similar lines was the joint research effort of Manatt, Palmer, and Hidlebaugh (1976) which identified 30 essential teaching behaviors. Each of the 30 behaviors was tested for reliability and validity when evaluating teacher performance. More importantly, they were able to group these behaviors into five general categories.

1. Productive Teaching Techniques
2. Positive Personal Relations
3. Organized, Structured Class Management
4. Intellectual Stimulation
5. Desirable Out-of-Class Behaviors

50. Ovard, op. cit. p. 89-90
J. A. King (1981) proposed using five variables based on a continuum from direct to indirect measurement. The continuum extended from the most direct measures of teaching competence, product variables, to the most indirect measures, personal variables. (see fig. 1 below)

![Diagram](image)

**Figure 1.** Variables involved in determining teacher competence.


A 1982 National Symposium for Professionals in Evaluation and Research (NSPER) cited by Korshalla (1984) featured a workshop on teacher observation instruments. It highlighted as a model the evaluation instrument utilized in Sarasota County, Florida. This document emphasized seven categories of teacher evaluation: lesson planning, teaching methods, class control, evaluation procedures, recordkeeping, understanding of school programs and
policies, and use of materials. 52

Although numerous other studies exist with slightly different evaluation terminology, the essential attributes of successful teaching cited in Manatt's Teacher Performance Evaluation (TPE) model, overlap with categories suggested in Korshalla's thesis as well as those put forth by King. Many other diverse taxonomies of evaluation criteria could be reviewed here, however it appears less than coincidental that these taxonomies overlap significantly. California's extensive Beginning Teachers Evaluation Study, cited earlier, stated as a fourth major conclusion to its study:

that differences in patterns of teaching performances accounted for differences in pupil learning. In this study, 15 of 16 stepwise regressions at both grade levels in reading and mathematics yielded significant multiple correlations. (...)Therefore, it is reasonable to conclude that, in a study of teaching effects, one should be seeking patterns of performance that affect children's learning. (...)first, a pattern of teaching performances is more likely to be related to learning than is a single performance; second, effective teaching patterns differ by subject matter and by grade level. 53

Bearing these substantive findings in mind, it would seem more appropriate in this study to synthesize general categories which investigators have deemed essential along with their research

criterion in the hopes of emphasizing these areas in our teacher evaluation model.

1. Teacher Planning Skills

Teacher planning skills have long been recognized as prerequisites for effective implementation of teaching strategies. Lupone (1961), after studying 240 elementary level teachers, stated, "the permanently certified (vs. provisionally certified) teachers received higher ratings in: the ability to organize and plan effectively, the skill to translate subject matter into living experience, the proficiency in using effectively related materials in classroom instruction, an understanding and more sympathetic attitude toward the child,..." Research by Brophy (1978) brought to light the critical relationship between good planning and effective classroom managers and found that, in most cases, student inattention and misbehavior could be traced to discontinuity in the lesson due to inadequate teacher preparation. Brophy concluded that good planning eliminates needless interruptions, decreases students' frustrations, and permits more time for instructional activities.

To avoid these pitfalls, a good deal of time should be devoted to lesson planning, where decisions must be made on such matters as (a) textbooks and related curriculum materials,

55. Korshalla, op. cit. p. 69-70
setting objectives for short, medium and long range, time schedules for rates of progress, homework allotments, remediation or review time, testing procedures for evaluating appropriate lesson objectives and so on. The findings of the Beginning Teachers Evaluation Study, cited earlier, identified the following six planning skills as those of effective teachers:

1. Clear goals and objectives for students
2. Setting of well-defined standards for student performance
3. Evaluation procedures are built-in for corrective action
4. Builds on students’ previous experiences and skills
5. Provides sufficient reinforcement activities
6. Adapts instruction to individual needs

A presentation will be more meaningful if the teacher can relate the information to the interest, needs and experiences of the students. The teacher must consider such questions as: whether or not to review the previous lesson; when to state objectives; how much content should be included in each lesson; when to interrupt the presentation with short question and answer segments. Hough and Duncan (1970) have suggested that the teacher should:

1. Relate the frame of reference to the pupil’s experience
2. Make the essential ideas stand out by ordering, comparing and contrasting
3. Make major points clear
4. Carefully shift attention when new focus is needed
5. Use illustrations that have personal qualities and
concreteness before abstract and nonpersonal examples. 56

Not unsurprisingly, there is research that shows lesson planning can produce adverse effects on classroom teaching and this factor should be addressed. Zahorik (1970) selected 12 fourth-grade teachers from four suburban schools in Milwaukee. The teachers in each school were assigned at random to a "planning group" and a "nonplanning group". The "planning group" was given a partial lesson plan two weeks before it was to be used with their classes. The "nonplanning group" was given no such plan. The teachers in this group were merely visited and invited to carry out an unspecified teaching task for the researcher—a task in which they will teach their class a lesson in a given topic for about 30 minutes. The topic was not dealt with in any previous classes but was familiar enough to the nonplanning group for them to be able to teach on such a topic without preparation. Zahorik analyzed transcripts of tape recordings of the 30 minute lessons, using three categories of behaviors: (a) data initiation, narrow vs. broad; (b) data encouragement, nongenuine vs. genuine; (c) data extension, non-authentic vs. authentic.

The results showed the following: the nonplanning teachers used a larger percentage of authentic data-extending behaviors (32%) than did the planning teachers (22%). The planning teachers more frequently "summarized pupils' responses, repeated

original solicitations, and tried to shape pupils' responses to reflect the teacher's views.\textsuperscript{57} The nonplanning teachers, on the other hand, more frequently "asked pupils to expand... their ideas or gave some expansion to the pupils' ideas"\textsuperscript{58} than the planning teachers. Zahorik concluded that the two groups differed more than chance could explain in their use of the two types of data-extending behavior but not in their use of data initiators and encouragers.\textsuperscript{59}

From this one may surmise that planning can make teachers too rigid, too concerned about following their plan, and correspondingly less open to students' ideas. Teachers should beware of trying to achieve what they have planned at the expense of what students can contribute to the learning situation. Also, it should be noted that this study did not evaluate the effect of planning on what students learned. The criteria consisted mainly of the teacher's openness to and the use of students' ideas. In many lessons, the objectives are knowledge and understanding of complex concepts and principles. If students' ideas take a class away from the main objectives of the teaching and learning, then the greater rigidity associated with planning might be all to the good — the nonplanning teachers may have been more receptive to students' ideas simply because, owing to their lack of planning, they had too few ideas of their own.

\textsuperscript{57} J. Zahorik, "The Effects of Planning on Teaching", in The Elementary School Journal, Vol. 71, No. 3, p. 148
\textsuperscript{58} ibid, p.148
\textsuperscript{59} ibid, p.151
Another concern for educators is the aspect of bias which unintentionally seeps into the learning environment. Human beings as educators, we are told, simply do not act randomly; they act in accordance with established habits, attitudes, and expectations. 60 By simple calculations researchers were able to determine the number of interactions expected by chance, and from observation were able to determine the number of actual interactions. Researchers asked a group of new teachers to classify the students' involvement in the teacher-student interactions of experienced teachers according to the following categories:

* Seated in the front half of the class vs. seated in the rear half of the class
* Seated on the left-hand side of the class vs. seated on the right-hand side of the class (dependent on whether the teacher is left or righthanded)
* Girls vs. boys
* Nice-looking students vs. average-looking students
* More able students vs. less able students
* Minority group members vs. nonminority group members

They found that "every observed teacher showed some bias. Some worked predominantly with girls; others, with boys. Some worked mainly with students in the front of the room. Some worked almost exclusively with the more able students." 61

61. ibid, p. 611
Gage and Berliner also noted similar biases in pupil-pupil relationships:

Attractive children are more often chosen by other children for activities reflecting acceptance (Salvia, Sheare, & Algozzine, 1975). Relationships have also been found between pupils' acceptance by fellow-pupils and many other pupil characteristics, such as achievement and social class (Gronlund, 1959). These authors advocate planning as a means of overcoming such biases by making plans about what students you will ask to work on reports, by purposefully calling on some students to answer some questions, or by pairing your favorite student with the one you find most difficult, so that you can try to interact with both of them. In concluding Gage and Berliner added, "More than other persons, teachers have to face and compensate for their biases because their actions can harm the students they are responsible for."  

2. Classroom Organization and Management Skills

The first thing a teacher must accomplish before he/she can implement any plan is to create and maintain a classroom environment favorable to learning, and more importantly, to long-term learning. As long as the teacher maintains an environment favorable to learning, pupils in her class will learn something; and some of what they learn will result in progress toward the specific objectives the teacher has defined while some will not. It is worth noting that if we could help teachers improve their skills in this area alone we could probably produce a dramatic

62. ibid, p. 612
63. ibid, p 613
improvement in instruction with many teachers especially new teachers who tend to be less businesslike in the beginning.

When the classroom environment is favorable to learning, further acceleration of pupil progress toward the specific objectives of the lesson depends on implementation of the lesson plan. If the pupils perform the activities outlined in an appropriate plan, they may be expected to make more progress toward the specific objectives the teacher has defined than they would without these experiences. Observable behaviors of the teacher while implementing the plan constitutes a second basis for assessing teacher performance. Even when the plan is being implemented, there may be some pupils who, because of motivational factors, disabilities, or other individual characteristics, do not become involved in the appropriate activities. Behaviors of the teacher designed to adapt the plan to maintain involvement of such students provides another basis for evaluating teacher performance in the classroom.

Early literature dealing with classroom organization and management skills pointed toward the physical aspects of the learning environment: creation of an attractive and orderly classroom by the teacher was associated with gains in student achievement. But in recent years the creation of a warm and supportive environment by the teacher is being emphasized for effective teaching. The work of Good & Brophy (1977) stressed the critical role of classroom environment in the learning process. Disruptions and misbehaviors are more likely in classrooms where some students are seated out of sight of the teacher or where
classroom traffic patterns are such that students tend to bump into one another frequently. Education USA (1983) highlighted a Rutgers University report linking teachers of messy classrooms with a lack of stimulation, dependability, creativity, and enjoyment.

The general aspect of behavior management as it relates to teacher performance has received extensive attention in recent years, especially its relations to student achievement. Evertson and Emmer (1982) profiled an effective classroom manager as one who has a clear set of expectations about student behaviors, establishes procedures to guide student behaviors, gives prompt feedback to students, and develops 'clear' and 'consistent' consequences for student behavior. According to Brophy:

Teachers who produce substantial achievement gains in their students tend to have most of the characteristics identified by Kounin (1970) as keys to effective classroom management (Brophy & Evertson, 1976; Good & Grouws, 1975). These include "with-itness" (monitors the entire class continuously); overlapping (can do two or more things simultaneously without having to break the flow of classroom events); signal continuity, smoothness, and momentum (moves activities along at a good pace, without confusion or loss of focus); and variety and challenge in seatwork (provides seatwork that is at the right level of difficulty for students and is interesting enough to hold their attention).

---

65. Korshalla, op. cit. p. 72
This study provides additional evidence of the importance of the teacher's activities at the beginning of the year by stressing the necessity of an efficient system for organizing procedures, rules, and initial activities, and for communicating this system to the pupils as a major teaching task at the beginning of the year.

Though the authors suggest that further correlational and experimental research is needed for further verification and extension of these results, they feel that:

...it is reasonable to conclude that effective classroom organization and management during the year can be predicted from the first several weeks of the school year. The teaching characteristics and behaviors that appear to discriminate best among more effective and less effective managers include the quality of leadership exhibited by the teacher in managing behavior and instruction, in planning for student concerns, and in coping with constraints. The more effective managers had a workable system of rules and procedures which they taught to their students during the first several weeks. They monitored their students carefully and did not "turn them loose" without careful directions. They did not appear to treat inappropriate behavior differently than the less effective managers, but they stopped it sooner. Consequences of appropriate and inappropriate behavior were clearer in their classrooms and were applied consistently. Thus these teachers established their credibility early and they were predictable.

68. ibid, p. 230
3. Teacher Instructional Skills

As stated in the introduction, there has been a change in research emphasis away from variables extrinsic to the classroom. More and more investigators are concluding that the influence of the teacher is the variable of primary importance in maximizing student achievement. The corollary to this development follows: that teaching behaviors will correlate significantly with student achievement. Research on teaching behaviors has not elicited specific 'discrete' teaching behaviors consistent with student achievement simply because the art or science is too complex. However research has brought forth innumerable skills which enhance teaching methods.

Despite innovations in school architecture, according to Dunkin and Bibble (1974) classrooms are surprisingly similar. Most contain a desk for the teacher and smaller desks for each pupil, are surrounded by windows, chalkboards, and bulletin boards, and have tile or wooden floors. They also tend to have a waste basket, a pencil sharpener, supply cupboards and exhibits of work by pupils or of materials pertinent to the subjects taught. Likewise, classrooms are similar in the activities that take place in them. Citing Herbert (1967), Nuthall and Snook (1973) they suggest that the activities of lessons usually fall into three basic forms.

The first form (lecturing) is one in which the teacher is in control of the treatment of subject matter. The teacher is lecturing, performing, demonstrating or exhibiting materials. In the second form (teacher-pupil interaction) both teacher and pupil have some control over the treatment of subject matter. Usually this involves verbal interaction with variations in the
involves verbal interaction with variations in the degree to which teacher or students control the course of the interaction. In the third form (seat work) the students are displaced from the direct control of the teacher and are engaged in assigned or unassigned exercises, practical work, or study. In this last form of lesson the teacher's control is indirect.

From this one can infer that a great deal of teaching still takes the form of a solo performance. Whether it is lecturing, explaining, pointing out relations, or giving examples, teachers often find themselves before the class for periods as short as 5 minutes to as much as a half hour. Samuel Johnson, the great literary figure, was quoted in 1799 by his biographer James Boswell as saying "Lectures were once useful; but now, when all can read, and books are so numerous, lecturers are unnecessary. If your attention fails, and you miss a part of a lecture, it is lost; you cannot go back as you do upon a book". This point has since often been made. It is not difficult to say anything for or against the lecture method but the intriguing question is, why has the lecture method persisted despite its obvious shortcomings that have been evident for centuries?

McLeish, (1968) defends lecturing as "valuable in surveying a whole field of knowledge through the medium of a living personality, in relating this body of knowledge to the primary aims of life, and in arousing active interest that leads to

comprehension on the part of the student." It is true that many students do not learn effectively through reading, and the lecture can introduce them to the subject matter but part of the answer may also be found in administrative factors rather than learning and student factors. Gage & Berliner (1979) present the following reasons for lecturing's longevity. First is its cheapness because the ratio of students to teacher can be extremely large. Second, the lecture method is flexible since it can readily be adapted, on short notice, to a particular audience, subject matter, time limit, and set of equipment. It can also be adapted to the teacher's schedule to a degree that printed and programmed teaching materials do not allow. Finally, students may be reinforced by the warmth, humor, drama, enthusiasm, and attention - to say nothing of the knowledge and comprehension - that an effective lecturer bestows upon them.

Despite all these possible advantages, the lecture method would probably have disappeared if experience and research had indicated that it is completely ineffective. But such evidence has not been forthcoming. Scores of experiments have been performed and the overall verdict is that the lecture method is just about as effective, by and large as other methods. Dubin and Taveggia (1968), cited in Gage and Berliner, reviewed the data of nearly 100 studies over a 40 year period. Of these 51% favored the lecture method, and 49% the others. Dubin and Taveggia also

72. Gage & Berliner, op. cit. p 444
in these investigations and the difference turned out to be so close to zero as to be attributable to chance.

The traditional classroom has received much comment by researchers about the centralized role of the teacher. We learn that teachers spend most of their time as recitation or discussion leaders, supervisors of action, and informers, or are not directly involved in classroom events. About half the time teachers are emitters (suggesting that teachers are actually talking for several hours during the typical school day). It has been argued that teaching would be more effective if it were individualized, if individual interaction were encouraged between teacher and pupils. Centralization, according to Dunkin and Biddle (1974) is found to be associated with subject-matter differences and to be more prevalent at the lower grades. Younger teachers are more likely to exhibit centralization than older teachers, and also men rather than women. But surprisingly pupils are found to have a greater involvement when the teacher's role is central than when it is not. 73 It seems that when the teacher wanders off stage, pupil's attention wanders also.

Let us turn our attention now to variance in student ability and classroom pacing. Very important work in this area has been conducted by Lundgren (1972). He argued that the variance (spread) of the class ability, may effect upon the pacing of instruction and likewise affect class achievement. Lundgren hypothesized that teachers were likely to conduct their teaching so as to "reach"

73. Dunkin & Biddle, op. cit. p. 388-389
pupils toward the lower end of the ability continuum (students whose aptitudes fall in the 10th to 25th percentile range). These students he labelled the "steering group". His finding were that "Average I.Q. of "steering group" pupils varies inversely with amount of teacher informing and disciplining; co-varies directly with amount of teacher leading-discussions; is unrelated to amount of teacher motivating, planning, and evaluating. These data suggest that less able students place restrictions on how fast the more capable students can achieve (move through the curriculum). Arlin and Wesbury (1976) conducted an experiment to determine how teacher pacing and student pacing affected faster students. Two clear results were obtained. First, the variance of progress was much greater in the self-paced condition. Some students finished the material considerably faster than students led by a teacher, but some students finished considerably slower. Second, students in the teacher-paced group scored lower on the achievement test than did self-paced students. The mean score of the teacher-paced group was equal to the 23rd percentile of the self-paced group. This is interesting, in view of Lundgren's data suggesting that teachers aim their instruction between the 10th and 25th percentile in aptitude. These results are worth serious consideration in the design of classroom instruction. But other work must be studied in conjunction with these results to obtain a more global idea of effective classrooms.

74. ibid, p. 216
75. Good & Brophy, op. cit. p. 523-524
Jorgenson (1977) conducted a research and showed that the most successful teachers were those who moved students along at a pace that requires students to work continually up to their capacity. Also Barr (1975), and Good, Grouws & Beckerman (1979) concluded that the more successful teachers are those who are task oriented and businesslike in moving the class along at a brisk pace. Bear in mind, however, the distinction between inappropriately trying to teach students material that is too difficult for them and appropriately teaching material that is at the right level of difficulty but moving them at a brisk pace. Data on pacing support the latter but not the former. In fact other data stress the importance of presenting tasks that allow high levels of learner success. Brophy and Evertson found, that during recitations, when the teacher is present to provide feedback and help, about 75% of the questions that relatively successful teachers ask are answered correctly (the percentage is a little higher in low-ability classes). For seatwork assignments, when the teacher is not monitoring performance continually, success rates will have to approach 100% if students are to be expected to work continually on the assignments until completed.

In general, then, successful teachers move students briskly from step to step, but the steps themselves are kept small, easily within the grasp of most students. This is contrary to several trends of thought popular in the 1960's.

76. Brophy, op. cit. p. 736
77. ibid, p. 736
78. ibid, p. 736
One aspect that plays havoc with the studies we've just reviewed is class size, especially if it increases the variance (spread) of class ability. Teachers have to make instructional decisions: when to evaluate, whether to teach certain units by groups, individually, or as a class? These decisions become more complex as class size increases. When you consider Lundgren's study that teachers should aim their instruction between the 10\textsuperscript{th} and 25\textsuperscript{th} percentile, it would appear that the most able students are hurt by teacher pacing. Thus plans must be made to facilitate self-study materials for high-achievement students. However in the general class setting, where volunteers are called upon frequently,

...those who raise their hand tend to be the brighter, more attentive, and generally more successful in the class. Also by virtue of the fact that they are willing to volunteer frequently in public response situations, these students tend to be assertive, confident, and healthy in their self-concepts.\textsuperscript{79}

Hence general classroom instruction is inappropriate for students of the other end of the spectrum. It is in small groups that teachers concentrate on working with the slow and timid students whom they hesitate to put on the spot in the general classroom setting. A study by Arlin and Westbury (1976) noted that students working on mastering a self-paced science program while being monitored by teachers who frequently keep students on task did not generally move faster than a teacher-paced group and under less than ideal conditions it was conceivable that many students

would fall hopelessly behind. 

In concluding, it would seem that there are limits to the size and range of a group that a teacher can effectively instruct. In large classes, say greater than 25, research indicates that teachers who emphasize independent activities are more likely to benefit the high-aptitude students. Conversely, teachers who emphasize group instruction are more likely to benefit middle to low ability students. In large classes it is probably easier to accommodate differences in learner abilities by gearing general activities at low and middle students and supplement instruction with carefully planned independent work for high ability students.

4. Teacher Skills in Pupil Evaluation

If tests are to be used for evaluating achievement as opposed to interviews, questionnaires, and anecdotal records, their form and content should, for the most part, be known in advance of an instructional sequence so that the test items reflect quite directly the objectives of the lesson plan. Testing is often the heart of an instructional series, and thus the tests items should not merely reflect retention but higher-achievement skills such as cognitive learning and the progress of the students' learning in grasping the lessons' objectives. Teachers must realize that tests sample only a small amount of

80. Good & Brophy, op. cit. p. 524
81. Gage & Berliner, op. cit. p. 685
what students have learned but that small amount must provide a meaningful indicator of what has been learned.

In a nutshell, testing has two major forms, "norm-referenced" and "criterion-referenced" and a good test must be "reliable" and "valid". Norm-referenced tests allows us to compare an individual with other individuals. For example, Johnny scores 55 on a reading comprehension test, is it an acceptable competence in reading? How it's evaluated depends upon the standards used to interpret the score. The norm could be based on students in the same class, students in the same school, students in the same district, or students from the same province. If a score of 55 places a student in the 74th percentile within the province, then that student's knowledge is exceeded in his reading ability by only 26 percent of the other students. If however a score of 55 places Johnny in the 47th in the nation other obvious deductions may be drawn about Johnny and his distant peers as well as his province's performance on this norm-referenced test. But as you may have already discerned norm-referenced measurements attempt to 'spread people out' by evaluating student performance differences among individuals and was never intended to be used for student evaluation in the classroom.

Criterion-referenced tests describe an individual performance to a fixed performance standard; the performance of others is irrelevant. Teacher made tests should evaluate students against the objectives laid out in the lesson plan, so their
tests are rightly called criterion-referenced tests. Although teachers are knowledgeable about their subject matter, they sometimes give insufficient attention to the preparation of tests. "Testing is often viewed as a disagreeable adjunct to teaching, rather than as an essential part of the instructional process."

However, according to Aiken (1971) when used in an efficient way,

Tests may serve not only to evaluate and motivate students but, more importantly, to inform the teacher of the extent to which her instructional objectives have been realized. Tests information, by acting as a source of feedback on the effectiveness of instruction, can be used to modify teaching procedure with a class or with individual students.

Once objectives for a lesson plan have been decided upon and an outline for proceeding with the class is prepared, test items can then be constructed in the hopes of measuring the extent to which students are attaining the objectives. However, one must be careful in choosing test items that are appropriate for measuring specific objectives. For example, in science, short-answer and completion items lend themselves well to measuring knowledge of terminology, but they are not appropriate for measuring higher order cognitive skills such as when a physics teacher must test for the quantitative relation between heat and work for example. Thus, before preparing a test teachers must seriously consider the variety that exists in test items and the appropriateness of each in relation to the objectives taught, cognitive levels to be tested and the overall validity of the test.

83. ibid, p. 26
Choices among types of test items can be divided into two major categories: long answer (essay questions) and short answer.

Types of test questions

- Essay
- Short-answer
  - Supply
  - Select
    - True-false
    - Multiple-choice
    - Matching

Figure 2. Types of test questions from which teachers can choose in making their own tests.

Source: N. L. Gage & D. Berliner (1979). *Educational Psychology*

As indicated by figure 2 above, the short-answer questions can in turn be divided into the "supply" and "select" types. These are respectively the types of short answer questions in which the student supplies the answer or selects it from a set of alternatives. Select types can be further divided into the well-known true-false, multiple-choice and matching categories. Initially, when evaluating students, teachers must decide between an objective or essay type test. The two types are drastically different in many ways and certain types of test items are more appropriate than others for measuring specific objectives.

84. Gage & Berliner, op. cit. p. 727
Essay questions are best used for objectives dealing with the improvement of written skills, interpretation and use of data, creative expression, organization of ideas, developing a position, and thematic development. However, according to Gage & Berliner (1977), teachers often write essay questions that only call for factual information. Knowledge and comprehension can be assessed more fully and more reliably by objective tests. If essay questions are to be asked, then teachers should make the question as clear and objective as possible. Gage and Berliner point out that teachers should select the verbs they use carefully and perhaps explain them to their students. These authors point out that verbs such as 'Discuss, define, justify, explain, tell about, summarize, classify, etc.' all imply more or less specific tasks, and teachers should choose among them judiciously. If a teacher wants an analysis or synthesis to have certain properties, then he/she should tell the students what he/she wants, unless the question was intentionally vague in order to require the students to learn to do structuring. Henson (1981) asserts that essay questions should force students to sharpen their understanding about similar or dissimilar concepts thus requiring students to do more than reproduce information. For example, if one asks, "Compare and contrast World War I with World War II." One can get more specific answers if the following limits are added: "Compare and contrast World

86. Gage & Berliner, *op. cit.* p. 733
War I with World War II according to their ground strategies, air strategies, number of causalities, and the number of countries involved.87

The main advantage of the essay question derives from the complexity of what it requires of the student. Another advantage is that it takes little time to make up. But this leads to corresponding disadvantages: essay tests not only take much more time to score, but they are much more difficult (and sometimes very tedious) to score. The complexity of the answers that they elicit leads to their major fault: they are hard to score reliably. It has been known since about 1910 that essay tests are often scored with too little agreement between different teachers.

In a dramatic study by Educational Testing Services, 300 essays written by college freshmen were all rated independently and anonymously by 53 "outstanding representatives" from several fields. More than one-third (34 percent) of the essays received all possible ratings on a nine-point scale! Another 37 percent received 8 ratings. Other research shows that even when experienced teachers expend their best efforts, inter-rater reliability is very low on essay tests. It seems that who marks the essay is more pertinent to the grade it will receive than the nature of the essay itself.88

Finally, essay tests have the disadvantage of inadequately covering, or sampling, the domain of achievement; its validity along with its reliability is questionably low.

So the question of whether you should use essay tests boils down to a trade-off between the advantage of testing for a specific, more complex kind of achievement and the disadvantages of low reliability of grading, relatively limited sampling of the subject matter, and time-consuming scoring of tests.

Many teachers compromise by using both essay questions and short-answer questions. This way you have the advantage of looking at the subject as a whole as well as the flexibility to pinpoint the particular fact, concept, or principle that you want to test. It also has the advantage of being relatively easy to score, since one needs to pay attention to only a limited range of ideas - a word, phrase, or at most a few sentences. Thus your judgement of an answer's correctness is made more simple and reliable.\textsuperscript{89} The disadvantage of the supply type is that the answer may be expressed in more than one form because a bewildering number of synonyms, arguable alternatives, etc. are bound to crop up. Furthermore, the test will likely fail to measure the abilities to make important distinctions and do subtle reasoning for which such questions are unsuited.\textsuperscript{90}

An important aspect of teaching is the necessity of students to perceive their teacher as fair and impartial. In this way objective questions are more likely to be more fairly graded since the answer is predetermined. Further, it is contended now that the multiple-choice item can measure the same aspects of an

\textsuperscript{89} ibid, p. 730
\textsuperscript{90} Aiken, op. cit. p. 33
educational objective as well as any pen and paper test excepting of course written expression and originality. Indeed, of all the select-type forms, multiple-choice is probably the most flexible. However true-false questions can be constructed much more quickly and as Brown (1976) noted, can be used to good advantage with young children, especially when only a general estimate of student performance is necessary (to quickly check if students read an assigned material?). However, the true-false question is relatively undesirable. First, because of its susceptibility to guessing, reliability is low and students tend to be less motivated to study for this type of item due to the "guessability" factor. Second, true-false items have been found to elicit the "response set" of acquiescence. This is a consistent complex tendency which people have in differing degrees, to say "true" when in doubt. 'Response sets' are tendencies on the part of the examinee to answer an item on the basis of its form rather than its content.

From time to time most teachers elect to use matching tests. This type of item enables the teacher to measure a student’s ability to make important associations and the value of such associations becomes apparent from the number of standardized tests that of late are including stimulus-response type of items more and more. But matching items are not easy to construct. Care must be taken to avoid using a stimulus that inadvertently

91. Good & Brophy, op. cit. p. 450  
92. ibid, p. 452  
93. Aiken, op. cit. p. 34
matches with more than one response. Further, according to Gage & Berliner, matching items "are usually appropriate only for measuring ability to match up relatively discrete facts and do not lend themselves to the measurement of more complex and subtle kinds of knowledge or comprehension." 94

The multiple-choice questions on tests and exams are becoming increasingly popular in schools today, mainly because of its merit and versatility. As noted by one researcher,

In expert hands, multiple-choice questions can be written so as to call for sophisticated and subtle mental processes of many kinds. Writing such questions can be intellectual fun of a high order, but it is time-consuming. 95

Just the opposite of essay questions, multiple-choice questions take a lot of time to make and little time to score. It is frequently stated that this type of tests favor average students but penalize poor students and bright or creative ones. Actually, the reverse is true. Bright students tend to do as well on properly constructed multiple-choice tests as they do on other measurements of performances such as essays and lab work. As stated by Violato (1985),

Properly constructed multiple-choice tests can certainly measure higher cognitive processes, such as application, analysis and even synthesis. By contrast most student essays are mere regurgitations of second- or third-hand facts. I have found that usually less than 10 percent of student essays demonstrate cognitive processes beyond the knowledge level, while multiple-choice tests can be specifically designed to measure complex learning outcomes. 96

94. Gage & Berliner, op. cit. p. 732
95. ibid, p. 732
96. C. Violato, op. cit. p. 19
Given these and other considerations which would be too lengthy to include, there are advantages and disadvantages of different kinds of tests, perhaps the best solution is to use a variety of tests and assignments to achieve a balance of the different kinds of learning strategies and types of content mastery that various types of tests foster. But as a rule, an essay item should not be used when it is possible to measure the same thing objectively. And if essay questions are to be asked, certain necessary precautions are necessary. Studies, showing the effects of handwriting on scoring and teachers biasing their scores on the expectation for particular students (halo effect), require that the paper be read in an anonymous fashion. This can be done by telling students to put their name on the backs of their papers. After doing one's best to score essay tests objectively, it must be remembered that feedback to students from subjective tests should always contain positive elements.

The most frequently committed sin of English teachers and others who grade essay questions is the failure to say something good about the student's answers. Especially in grading the long essays known as term papers, teachers should realize that students are sensitive about their writing. If students never receive anything but negative criticism and faultfinding, they develop a crippling anxiety about writing and, as soon as possible, stop writing altogether. Well-placed and sincere praise, along with moderate amounts of criticism, should be the policy in grading essay tests.

97. Gage & Berliner, op. cit. p. 734
5. Teacher Professional Growth

Traditionally, we have always associated education with societal change. Educational institutions have both been criticized for their lack of attention to the needs of tomorrow's learners but at the same time given credit (rightly or wrongly) for having the most potential for precipitating social change. This possibly is the underlying reason that authors of teacher evaluation models have begun to give increasing attention to teachers' professional development. An overview of the criticisms directed at educational institutions is provided by Kauffman (1976) in his discerning of their approaches to learning:

Futurism questions traditional education on two grounds. First, traditional methods of education tend to produce individuals who are psychologically ill-equipped to cope with a society undergoing continual rapid change. Second, the content of the traditional curriculum is designed to fit the student into existing society, in spite of the fact that the society - more accurately the successions of societies which the student will live in will be quite different - a case of trying to hit a moving target by aiming at it where it is now, rather than where it will be.

Extending this argument to teacher training programs in place today and realizing that a beginning teacher would practise until 2025 before being eligible for retirement certainly dictates the necessity of professional growth for teachers.

One of the most valuable resources available to a teacher for professional growth is that of other colleagues. Research suggests that the most effective teachers share materials and

communicate frequently about curriculum and instruction. Further, they keep abreast of other aspects of school change such as curriculum development and instructional modifications in the subject matter they teach. 99

It would be shortsighted indeed to argue that it is the total responsibility of teacher preparation programs to train teachers for tomorrow's classrooms and schools. To embark on a career expecting that today's programs will suffice for tomorrow's everchanging needs is unrealistic. Commenting on British Prime Minister Benjamin Disraeli's remark on education more than a century ago, "upon the education of the people of this country the fate of the country depends." Senator McMathais, speaking at a leadership institute offered this insight:

Those words were never more true than when we consider the 21st century. Young people now make up two-fifths of our population, but all of our future. Each one deserves the chance to prepare for a worthwhile career and the satisfaction that comes with making a productive contribution to society. That is our moral imperative for the turn of the century. Our practical imperative is that no modern nation can survive, much less prosper, without an educated citizenry capable of mastering the complexities of high technology. 100

These two imperatives, moral and practical, underline the task of the educators in the future. It is the duty of the evaluators to see that these obligations are met.

99. Henson, op. cit. p. 354-355
100. C. McMathais, "Winning the Race Between Education and Catastrophe", in Journal of Teacher Education, Jan-Feb/84, Vol. 35 No. 1, p. 12
Though only five areas of teacher evaluation criteria have been cited, it goes without saying that there are numerous other elements in existence that could help in establishing criteria for an evaluation model. However, the areas categorized above are considered essential in establishing a starting point for the development of criteria in judging acceptable teaching performances. Assuming that there is a general impetus for change, and we have identified some indications that educators are ready for some different approaches to evaluation, the remaining aspects of an evaluation model must be part of a collective effort by the interested parties involved both at the school level and at the administrative level. This aspect will be explored further later.
Current evaluation practices

a) An Overview of Teacher Evaluation in Canada

The purpose of this study is to propose a teacher evaluation model for use in New Brunswick, but as expressed in the introduction, it is essential to survey current practices of teacher evaluation before proceeding toward the establishment of a model for future years. It is hoped that by reviewing present policies and practices which are enforced by school districts across Canada and parts of the United States as well as obtaining the perceptions of teachers' attitudes towards these endeavours, that interesting conclusions might be elicited which could be helpful for implementing future evaluation systems.

Dormancy of Evaluation During Reorganization

During the early 1960s, school systems in New Brunswick and Quebec underwent dramatic changes. The Byrne Report which was implemented under Premier Louis Robichaud dictated sweeping changes for N.B. Similarly, the Parent Report foretold the drastic measures in store for Premier Jean Lesage of Quebec. Schools in both provinces, over a very short period of time, were made more accessible to the general public and with a more equitable financing system from their central governments and the creation of fewer, more regional school districts, education became an integral part of every family member in society. With
these changes came the disappearance of the traditional inspectoral system of teacher evaluation.

While provinces in Canada were struggling with what could be termed the global, physical aspects of organizational changes in their educational systems, the practice of formal teacher evaluation was, understandably, left behind. In a study by Richard and Michaud (1980) on "La pratique de l'évaluation des enseignants au Nouveau Brunswick, au Québec et en Ontario", they noted that

Vers 1960, les inspecteurs d'école ont cessé d'évaluer les enseignants. Puis, pendant plusieurs années, la pratique de l'évaluation formelle des enseignants fut négligée. Récemment la conjoncture a changé. 101

A similar history is made evident in research studies by Alberta Education. Reikie (1977) carried out a study on "Practices and Policies Involved in the Formal Evaluation of Teachers in Alberta". The study drew its data from questionnaires completed by superintendents. Some of the major findings were: (1) Tenured teachers were rarely, and sometimes never evaluated; probationary teachers were usually evaluated only once a year. (2) Most superintendents were not contemplating any changes in their formal evaluation practices. 102 The authors of other studies during this time were also cited by Duncan (1984) as noting "that few systems had clearly structured evaluation policies and

practices, that practices across the province were inconsistent, and that many jurisdictions had no teacher evaluation policies in place."103 In a study of teacher evaluation systems in Newfoundland and Labrador by George Hickman(1983), similar circumstances are brought to light in that part of the country:

Although the literature contains few references to the implementation of teacher evaluation systems in the Province, there is an indication that some school boards did attempt programs shortly after the reorganization and consolidation of school districts in 1969.104

Though teacher evaluation was in an apparent lull during these two decades in Canada, our colleagues south of the border were taking important strides toward redefining teacher evaluation. There, the inspectoral model with its listings of teacher characteristics, criteria, rating scales, or other types of procedures were rapidly being displaced by the emergence of several other models which were more teacher-centered as opposed to administrator-centered.

The Influence of Clinical Supervision

The multifaceted clinical supervision model achieved prominence in the early 1970s with Robert Goldhammer(1969). Emphasizing the more humanistic nature of teacher evaluation, Goldhammer saw the clinical supervision as a process that when

Given close observation, detailed observational data, face to face interaction between the supervisor and teacher, and an intensity of focus that binds the two

103. ibid, p. 31
together in an intimate professional relationship, the
meaning of "clinical" is pretty well filled out.

Following Goldhammer's death, Morris Cogan and other colleagues
at Harvard university continued the development of the many
aspects of this form of evaluation. His redefinition of the
essence of this method emerges as:

...the rationale and practice designed to improve the
teacher's classroom performance. It takes its principle
data from the events of the classroom. The analysis of
these data and the relationship between teacher and
supervisor form the basis of the program, procedures,
and strategies designed to improve the student's
learning by improving the teacher's classroom
behavior.

Cogan's work helped reorient evaluation away from the
administrative aspects of evaluation and toward its main
objectives of improving instruction and the continued development
and growth of the teacher. This was possible because the new
model molded by Cogan prescribed rather than defined roles:

The distinction between role definition and role
prescription is important: role definition rests on a
normative' standard expressed in terms of the central
tendency of a reference population, while role
prescription involves a value statement regarding a
desired state or goal. Therefore, because the roles of
the teacher and supervisor are prescriptive, they
reflect what the developers of clinical supervision
believed should exist and do not necessarily describe
any already existing norms.

105. Robert Goldhammer, Clinical Supervision: Special Methods
for the Supervision of Teachers, (Toronto: Holt, Rinehart and
106. Morris Cogan, Clinical Supervision, (Boston: Houghton-
Mifflin Company, 1973), p. 27
107. Cheryl Sullivan, Clinical Supervision: A State of the Art
Review (Alexandria, Va.: Association for Supervision and
Curriculum Development, 1980), ERIC ED 182 822, p. 17
Though the ideas of Goldhammer and Cogan helped develop a mode of teacher evaluation which was "interactive rather than directive, democratic rather than authoritarian, teacher centered rather than supervisory centered"\textsuperscript{108}, the complete application of the model in local school systems has not often been practical in spite of attempts to economize on methods of implementation:

Allowing existing administrators to function as clinical supervisors would help alleviate the cost associated with clinical supervision. These costs are high. Training of clinical supervisors is expensive. Further, because clinical supervision demands more time, energy, and skill than is required of a supervisor fewer teachers can be served in a given period than when traditional supervisory methods are used. If new personnel have to be hired to serve as supervisors, the costs of using clinical supervision becomes prohibitive for most schools and systems.\textsuperscript{109}

In spite of its obvious limitations, the ideals transmitted by clinical supervision contributed significantly in reorienting teacher evaluation away from its negative connotations. Traditionally teacher evaluation was used primarily to identify weak teachers so that they might be removed. As a result teachers feared evaluation, evaluators disliked the responsibilities, and the results of evaluation were substandard at best. The impetus of the clinical supervision process has modified the evaluation process by bringing to the forefront the requirement of pre and post conferences for improving communication skills, the necessity of developing a mutual trust between teachers and evaluator by having teachers become more involved in the development of the criteria to be used in the evaluation

\textsuperscript{109} C. Sullivan, op. cit. p. 39
instrument and in the establishment of the procedures to be followed in the evaluation process. This emphasis on analysis rather than inspection was not only oriented toward the improvement of instruction but was also expected to produce a continued development and growth of the teacher. In spite of its dismal fate, clinical supervision has left an enriching legacy.

As previously mentioned, other models emerged but striking similarities existed between the components of these models and the techniques created by Goldhammer and Cogan. One of these which involves the establishment and some level of teacher involvement in the evaluation process is the Manatt "Teacher Performance Evaluation" (TPE) model.\footnote{Richard Manatt, Developing a Teacher Performance Evaluation System as Mandated by Senate File 205 (Iowa General Assembly, 1976), (Association of School Boards, Des Moines Iowa, Sept. 1976), ERIC ED 134.222,} George Redfern (1980) developed a similar system based on the business model of "management by objectives".\footnote{George Redfern, Evaluating Teachers and Administrators, (Boulder Colorado: Westview Press, 1980)} Like the Manatt model, the criteria are set by the responsible school authority, but in the Redfern model, the teacher and evaluator jointly determine the individual objectives and the measurement process to be used.

There is a vast amount of literature available now describing many different school systems' approaches to teacher evaluation. As previously described in the research of evaluation criteria, the great majority of the models in use today show similarities at the general level. A striking characteristic of
these new approaches is the serious effort to relate teacher
evaluation practices to the improvement of instruction. First,
built into the models is the involvement of teachers in planning
the process: from establishment of criteria upon which one is to
be evaluated to deciding what aspect of the classroom process
will be included in the evaluation. Second, there is ample
provision for staff development leading to improvement once
problems in the teaching process have been identified.

If one couples this leap in evaluation effectiveness with
the public outcry for accountability which has occurred in recent
years it is not difficult to comprehend the difference in
attitude expressed by superintendents in Alberta toward teacher
evaluation in the last few years. A 1980 Alberta Education study
entitled "Certified Education Staff Evaluation", which according
to Duncan(1980) was essentially an abbreviated replication of the
1977 Reikie study indicated that a number of changes occurred
during the three intervening years in Alberta:

The Alberta Education study of 1980 showed a marked
change in the focus of teacher evaluation in a number
of areas. Certification and the awarding of a permanent
contract remained primary uses (90%), but use for the
improvement of instruction had risen from 18% in 1977
to 83% in 1980.112

A similar sentiment was expressed by Ryan & Hickcox(1980) in
their research commissioned by the Ontario Public School Men
Teachers’ Federation(OPSMTF).

112. Duncan, op. cit. p. 5
In the last decade or so, however, conditions have begun to change in several respects. Professors of educational administration have begun to make important distinctions in their training programs between teacher evaluation for dismissal/certification purposes and teacher evaluation for the improvement of instruction (i.e., the supervisory part of the role). Research and experimentation have brought forth scholarly frameworks for reexamining the process of teaching and learning. Developers have created a variety of evaluative and supervisory techniques to provide information and methods to improve the process. ...we have come a long way from reliance upon evaluation ratings based upon a supervisor's subjective impressions. Gradually, teacher evaluation is coming to be accepted by teachers, not as a threat, but as a means for personal and professional growth.

We have reached a point in Ontario educational history where conditions favor a major shift in attitudes and actions relative to teacher evaluation. 113

It is difficult to perceive this desire for a major shift in direction in teacher evaluation spoken of by Ryan and Hickcox without surveying present patterns being employed for teacher evaluation as well as the present perceptions of employers-employees toward the present approach.

Present Evaluation Practices

The Richard and Michaud (1982) study indicates that teacher evaluation was done in 100% of the districts of N.B. and Ontario as well as 68.4% of Quebec districts but similar to Alberta, these school districts are lacking structures and policies for carrying out these evaluations. For example when superintendents in the Richard and Michaud (1982) study were asked "A-t-on développé un formulaire d'évaluation qui répond aux besoins

particuliers de votre conseil scolaire?", only 62.5% of N.B., 23.4% of Quebec, and 68.4% of Ontario superintendents answered yes. In this area the most consistent province was Newfoundland where Hickman (1983) reported 76% of the boards specified the general procedures to be followed in the evaluation of teachers. The procedures specified were the following: composition of the evaluating team, number of required classroom visits, pre- and post-observation conferences, and types of reports to be used in the evaluation process.¹¹⁴

The "Survey of Teacher Evaluation Practices" instrument which Ryan and Hickcox (1980) used to survey 88 Ontario school districts revealed that only 66% of probationary teachers were observed 2 or more times during the year and 17% of the 231 respondents reported that the number of observations varied from year to year. This was interpreted by the researchers to mean that some probationary teachers in a school might be observed more than three times a year while others might be less, depending on the particular problems being experienced by the teacher and depending, perhaps, upon the time demands on supervisors. A somewhat different trend emerged for the permanent teachers as might be expected. For this group, 68% of these respondents reported being observed once or less a year.¹¹⁵

¹¹⁴. Hickman, op. cit. p. 72
¹¹⁵. Ryan & Hickcox, op. cit. p. 6
Similar patterns are found in the studies from other provinces. Duncan in his 1984 Alberta study reported the number of classroom visits made by evaluators during their last formal teacher evaluation. They were averaged out between superintendents, central office staff, and principals and were respectively 2.2, 2.7, and 3.4. A disproportionate 39.2% were non-tenured teachers. An interesting occurrence was that 31.4% of those evaluated were those teaching at grades 7-9 while only 14.7% of grade 10-12 teachers were observed, the lowest of all grades. \(^{116}\) The Richard and Michaud (1982) study of N.B., Quebec, and Ontario francophone school districts portrays a similar story with nonpermanent teachers being evaluated 2-3 times on average, while over 85% of permanent teachers being evaluated once or less during a school year. \(^{117}\) The findings in Hickman’s study indicate that "while many teachers do not object to being observed in their teaching, they are indeed skeptical about the validity of evaluators' reports based on one or two classroom observation." \(^{118}\)

Hickman also made a specific recommendation concerning the necessity of pre- and post-conferences:

Finally, the author recommends that formal classroom observations include a pre- and post-observation conference with the teacher. The objectives, criteria, and procedures should be clearly discussed with the teacher before the actual observation period. Pre- and post-conferences may also reduce the tension between evaluators and teachers and should be aimed at creating an atmosphere of mutual trust. The post-observation

\(^{116}\) Duncan, op. cit. p. 60-63  
\(^{117}\) Richard & Michaud, op. cit. p. 266  
\(^{118}\) G. Hickman, op. cit. p. 196
conference should be held with the teacher immediately after the lesson, because it gives instant feedback. It is in this conference that plans should be made for follow-up procedures and, if necessary, suggestions for improvement. 119

Ryan and Hickcox summarized the frequency of both pre- and post-observation conferences between teachers and their supervisors in the following manner. For probationary teachers, less than half report pre-observation conferences, meetings that ordinarily would establish between the teacher and the supervisor the objectives of the lesson, to discern what problems or concerns the teacher has thus producing greater evaluatee input, and for communicating the criteria upon which such judgements will be made. Even less impressive was the finding that only 33% of permanent teachers reported pre-observation conferences. 120

Certainly questions arise from these patterns. If observations are the major aspect of evaluation, what kinds of judgements can be made on the basis of such few observations? Surely, with classroom observation being an essential part of evaluation, the pressure is on the new teacher to perform well during these two or three visits by his/her superior. Medley, Soar and Soar (1984) in their text "Measurement-Based Evaluation of Teacher Performance" cast serious doubts about this practice:

It seems likely that, because of the instability of human behavior in general, and of teaching behavior in particular, one visit will not usually be enough; scores based on records of single visits to each teacher will probably not be reliable enough to be

119. ibid, p. 196
120. Ryan & Hickcox, op. cit. p. 9
useful."  

Though these authors believe that there should be some flexibility to the guidelines concerning the number and length of visits that are necessary for objective, reliable and valid data collecting, they report that "highly reliable scores can be obtained with six visits." Users of the 'Coker' instrument for evaluating, "report satisfactory results from four visits". Regardless of the methods or instruments one chooses to use, these authors recommend that,

...if at all possible, you schedule one trial visit to each classroom before the first visit for evaluation purposes. This will give both the teachers and the class an opportunity to get accustomed to being observed, and at the same time will give your observers a chance to get used to observing.

While current practices seem highly questionable to probationary teachers, the pattern of one evaluation or less for permanent teachers along with the possibility of no pre-observation conference would seem even less just. Hickman in his 1983 study of evaluation practices in Newfoundland found that

The heavy teaching loads of the vast majority of principals...combined with the usual routine administrative tasks, often leave little time for the evaluation of teachers. Therefore, many readily admitted frustration, claimed that they pay little more than lip service to teacher evaluation, and that their efforts in this regard are often little more than a form of bureaucratic routine which rarely results in the improvement of instruction.

122. ibid, p. 135
123. ibid, p. 135
124. G. Hickman, op. cit. p. 193
This feeling of frustration expressed by Newfoundland principals toward teacher evaluation might be comprehensible if one considers the findings by Duncan (1984) on the number of university credit courses on teacher evaluation. Of 288 principals surveyed in Alberta, 73.6% had one course or less in this complex field. Nonetheless, the Ryan and Hickcox study indicated that principals were involved in over 90% of the observations carried out in the 88 school districts surveyed in Ontario and principals in Newfoundland were involved in 97% of the evaluations according to Hickman. While this is not a surprising finding, it does serve to emphasize the crucial role principals play in current evaluation practices. Furthermore, the Ontario study also showed that 65% of the teachers "agreed to some extent that the principal is the person qualified to evaluate teachers".

Robert Howsam in a 1963 educational article titled "Teacher Evaluation: Fact & Folklore" states quite succinctly the school principal's problems of status and role:

Tradition has it that the principal possesses the competence to evaluate teacher effectiveness. After all, the principal has always been looked at as the principal teacher. Almost universally the principal's role has been defined in this way - he is expected to supervise and evaluate the teachers under his direction. Parents, school board members, superiors and, indeed, most if not all of the teachers themselves share this expectation. Besides, out of his experience and training he has developed this expectation of himself.

125. A. Duncan, op. cit. p. 52
126. ibid, p. 108
On the job, however, the expectation gets rude jolts. In the first place, there isn't time to do the job as most believe it should be done. Secondly, the principal soon finds that he doesn't know enough to justify immediate and direct supervision and evaluation in every area and at all levels. Some never get further with the disenchantment than this. Others, however, begin to ask themselves searching questions about the adequacy of the definitions of the status and role of the principal vis-a-vis the teacher. They find much to be wanting.  

Today it can be seen that Mr. Howsam had insight far beyond his time.

Teacher and Evaluator Misperceptions of the Evaluation Process

There is a distinct discrepancy between how administrators and teachers see the development of evaluation procedures within a school or district. In the extensive Ryan and Hickcox (1980) study of teacher perception of evaluation, only 31% of the administrators reported that procedures were developed entirely by supervisors, as opposed to 71% for the teachers. There are also some misperceptions about how the evaluator and evaluatee see the evaluation process itself. David Townsend in one section of his 1984 Alberta thesis drew comparisons between teachers and evaluators "perceptions of usual and ideal supervisor behaviors". His findings show that there are some significant misconceptions of the process among both groups. Evaluators perceived themselves as "meeting with their teachers prior to formal observations much more frequently than did their teachers" and evaluators reported


128. Ryan & Hickcox, op. cit. p. 12
"spending more time before observations finding out about teachers' problems and concerns than their teachers indicated they were."\textsuperscript{129}

Concerning the "Ideal Supervisor Behavior" aspect of the research, evaluators saw the Ideal Supervisor as "providing less direct advice for improvement and fewer opinions regarding the teacher's teaching than was seen as ideal by teachers." Also, evaluators "saw the Ideal Supervisor as one who would listen more than talk during conferences while teachers felt the Ideal Supervisor should talk more."\textsuperscript{130} Hickman, in his Newfoundland thesis, noted, "Whereas administrators preferred to give written reports to teachers only at the end of the evaluation period, teachers preferred to have a written report at the end of each classroom observation."\textsuperscript{131}

Increased in-service training in conferencing and communication skills could go a long way in alleviating these problems.

\textsuperscript{130} ibid, p. 55
\textsuperscript{131} G. Hickman, op. cit. p. 189
The Need for Restructuring the Teacher Evaluation process

In the "Implications about Practice" aspect of Hickman's thesis it is noted that

...if evaluation is to become mainly formative in nature, evaluators must somehow gain more time to evaluate teacher performance and to work directly with teachers to bring about needed improvements.

Teachers also point out the need for evaluators to spend more time on all stated purposes and criteria of evaluation if maximum benefits are to be achieved.\(^{132}\)

The Ontario study also showed a positive teacher attitude toward the need for restructuring the evaluation process;

Table 2 reveals that a majority of respondents agreed with each statement. Indeed, the responses seem to indicate a desire by a large part of the membership for the government to provide more active leadership in the area of teacher evaluation than is currently the case. This is especially true for the establishment of a province-wide evaluation advisory committee representing all interested groups.\(^{133}\)

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired Role of Provincial Government and Ministry of Education Regarding Teacher Evaluation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Provincial government and its Ministry of Education</th>
<th>Agree</th>
<th>Disagree</th>
<th>Unsure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. should require each school board to develop a plan for the conduct of teacher evaluation</td>
<td>62</td>
<td>29</td>
<td>.9</td>
</tr>
<tr>
<td>2. should establish common criteria for teacher evaluation</td>
<td>61</td>
<td>33</td>
<td>5</td>
</tr>
<tr>
<td>3. should establish common procedures for teacher evaluation</td>
<td>57</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>4. should establish an advisory committee including representatives from the federation,</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

132. ibid, p. 181
133. Ryan & Hickcox, op. cit. p. 90
trustees, administrators, etc. to develop common guidelines for the conduct of teacher evaluation 80 15 5

5. should establish an advisory committee (as above) to develop common criteria and procedures for the conduct of teacher evaluation 66 23 10

6. if it requires evaluation of teachers, should provide special funds to boards for training of supervisory personnel, observation teams, etc. 66 25 9

7. if it requires evaluation of teachers, should provide special funds to boards for professional development and in-service activities which are justified by evaluation findings 74 16 10

Note: where % does not total 100, balance indicates non-response to item

Further questions by Ryan and Hickcox to Ontario’s teachers dealt with the roles at the local level of the boards, administrators, and teachers in the evaluation process. The sample provided some stimulating responses. As shown in table 3, while a whopping 89% of the respondents believed that every school board in Ontario should have a written policy on teacher evaluation, only 38% agreed that the school boards should have the power to establish criteria or standards for evaluation. The need for mutual cooperation in the evaluation process is very much in evidence as 91% of the respondents to item 3 believe that evaluation policies should be established cooperatively, and another large majority (75%) believe that a board-wide evaluation steering committee should design, develop, institute, and monitor the evaluation program. 134

134. ibid, p. 94
### TABLE 3

**Desired Role of Local Board, Administration, and Teachers in Teacher Evaluation**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Agree %</th>
<th>Disagree %</th>
<th>Unsure or no opinion %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Every school board in Ontario have a written policy on teacher evaluation</td>
<td>89</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>2. The school board should have the power to establish criteria or standards of evaluation</td>
<td>38</td>
<td>50</td>
<td>12</td>
</tr>
<tr>
<td>3. Details of operation or procedures to implement board policy re teacher evaluation should be established cooperatively by the administration and teachers</td>
<td>91</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. A board-wide evaluation steering committee should be formed as the means of designing, developing, instituting, and monitoring a teacher evaluation program</td>
<td>75</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

Another essential area surveyed by Ryan and Hickcox was the type of evaluation system that best served the "Purpose of Improving Instruction". As can be seen by table 4, administrator evaluation of teachers was ranked first, followed closely by teacher self-evaluation. Peer evaluation was most frequently ranked third while student evaluation of teachers was seen by 50% of the teachers as the least likely to improve instruction.  

---

135. ibid, p. 100
TABLE 4

Types of Evaluation for Improving Instruction

<table>
<thead>
<tr>
<th>Evaluation type</th>
<th>1 %</th>
<th>2 %</th>
<th>3 %</th>
<th>4 %</th>
<th>Other response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher self-evaluation</td>
<td>28</td>
<td>26</td>
<td>15</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>Administrator evaluation of teachers</td>
<td>30</td>
<td>27</td>
<td>12</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>Student evaluation of teachers</td>
<td>4</td>
<td>11</td>
<td>15</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Peer evaluation (Teachers' evaluation of fellow teachers)</td>
<td>19</td>
<td>16</td>
<td>35</td>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

*—most likely to improve instruction; 4—least likely to improve instruction.

However, when asked which type would most validly evaluate a teacher's current performance, (Table 5) administrator evaluation was selected by 41%, and peer evaluation replaced self-evaluation as the second choice. Again student evaluation was the least favored. 136

136. ibid, p 101
<table>
<thead>
<tr>
<th>Evaluation type</th>
<th>Which of these would most validly evaluate a teacher's current performance? %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher self-evaluation</td>
<td>17</td>
</tr>
<tr>
<td>Administrator evaluation of teachers</td>
<td>41</td>
</tr>
<tr>
<td>Student evaluation of teachers</td>
<td>8</td>
</tr>
<tr>
<td>Peer evaluation</td>
<td>21</td>
</tr>
<tr>
<td>Other (combination)</td>
<td>13</td>
</tr>
</tbody>
</table>

Hickman believed that teacher evaluation required a cooperative effort and proposed the following changes to the Newfoundland school boards:

It is recommended that specific roles be defined in the school boards adopted policies for teacher evaluation with respect to all members of the evaluating team. While most of the current practices involve mainly the assistant superintendent and principal, many respondents suggested that integral roles should be developed for the vice-principal, department heads, and subject-area coordinators at the central office level.137

While writing about their overall impression of the responses on teacher evaluation, Ryan and Hickcox noted:

...our sample also felt that optimum professional growth comes from teachers' observing and assisting fellow teachers. Thus, the respondents appear to be receptive to both peer and administrative supervision (as opposed, perhaps, to evaluation). Some 65% agreed to some extent that the principal is the person

137. Hickman, op. cit. p. 194
qualified to evaluate teachers, but 72% agreed that
evaluation should be done by a team (including at least
one peer). 138

Summary of Research on Teacher Evaluation Practices in Canada

An analysis of the data viewed in this section reveals a
degree of consensus among teachers about most of the issues
believed absent in teacher evaluation. Teachers appear to want a
stronger leadership role from their local school board as well as
from school administrators in the formation of cooperative
structures pertaining to the procedures, guidelines and criteria
for teacher evaluation. Teachers are largely supportive of
administrator and self-evaluation, and reluctant although not
necessarily opposed to peer evaluation. Both pre- and post-
observation conferences between evaluator and evaluatee are very
derious, though misperceptions exist about exactly what each
person's role should be in these conferences. In-service training
is warranted in this area to increase communication skills and
improve evaluator competency.

The views expressed by most researchers about present
teacher evaluation practices generally reinforces the overall
sentiment that emerges from their data. Teachers seem to be open
to a revised teacher evaluation system that would be directed
toward the improvement of instruction, and while some differences
exist among them, they seem to share this openness to new ideas.

138. Ryan & Hickcox, op. cit. p 108
b) A Review of Some Evaluation Practices in the U.S.

As previously mentioned in the introduction of this study, the new concerns for the quality of education and of teachers in the United States has been translated into merit-pay, career-ladder, and master-teacher policies that presuppose the existence of effective teacher evaluation systems. As a result, many school districts in that country have reassessed their teacher evaluation policies. In an attempt to evaluate processes and produce information that other school districts could use for helping their teachers to improve and/or for making personnel decisions, the Rand Corporation of California in cooperation with the National Institute of Education undertook a study in 1983 to identify districts in the U.S. having highly developed teacher evaluation systems. In a review of school districts across the United States, 32 districts were identified as having substantive evaluation systems.

Assessment of these 32 districts systems led researchers to conclude that school authorities seldom if ever agree on what constitutes the best practices with regard to frequency of evaluation, role of the teacher in the process, criteria to be used in the instrument for evaluating teachers or how the information obtained could best be put to use.
Major Problems of Teacher Evaluation in the U.S.

The study revealed that there were major problems associated with their evaluation practices and that these difficulties were similar to all the 32 districts studied. Two important problems were inferred from respondents perception of their district's teacher evaluation practices. Almost all respondents sensed that "Principals lacked sufficient resolve and competence to evaluate accurately." The other major problem was a "lack of uniformity and consistency within a school system." In many districts, teachers believe that the present system still depends too much on the sole judgement or predisposition of the principal and leads to different ratings for similar teacher practices in different schools. Many respondents said that principals considered evaluation a necessary evil or a time-consuming chore. According to the researchers, in most districts teacher evaluation has been added to a principal's responsibilities (formerly the responsibility of central office personnel) without taking other functions away or providing additional assistance. Many Local Educational Agencies (LEAs) felt that staff responsibilities for evaluation did not receive enough training and that the training received provided insufficient guidance in the conduct of evaluation.

140. ibid, p. 22
141. ibid, p. 22-23
The major problems cited in this study are not unlike the deficiencies associated with the evaluation systems in Canada. However, there were other troublesome, though less blatant, problems within their systems that were not evident in the research in Canada. For example, teacher resistance or apathy was a frequently cited problem. Teachers reportedly fully supported their evaluation program in less than half of the districts.\textsuperscript{142} Much of this lack of support stems from teacher anxiety associated with the evaluation itself. Unlike in Canada, our colleagues do not have strong statewide teacher unions to offer support so an inadequate evaluation, for whatever the reason, could spell disaster for the teacher.

Teacher Evaluation Can Contribute Positively

Respondents involved in this study consistently reported two major results of teacher evaluation, "improved teacher-administrator communication" and "increased teacher awareness of instructional goals and classroom practices."\textsuperscript{143} Even in the less developed systems, teachers noted frequently that the process of pre-observation conferences, observation, and post-conferences substantially improved teacher-principal relationships and helped establish common goals. Others reported that the school climate had improved since evaluation responsibilities brought principals into the classrooms regularly. Effective evaluation systems

\textsuperscript{142} ibid, p. 22
\textsuperscript{143} ibid, p. 23
according to some respondents has created a sense of team effort and has contributed significantly to staff development. These responses are strikingly similar to those expressed by teachers in Canadian research. Barnabé (1981), when summarizing recent literature on "l'évaluation des ressources humaines" remarked, "L'évaluation du personnel constitue selon nous un outil de gestion et un moyen de communication entre un subordonné et un supérieur." In spite of its limitations, evaluation contributes positively because by increasing communications greater direction, feedback, and assistance follows.

Evaluation Systems Deemed Successful in the U. S.

Among the 32 districts found to have substantive evaluation systems, four were identified as having highly developed, somewhat unique, evaluation systems: Salt Lake City, Utah; Lake Washington, Washington; Greenwich, Connecticut; and Toledo, Ohio.

The evaluation system in Salt Lake City occurs in a 25 000 student population district lacking a teacher tenure clause or state-mandated teacher evaluation. The urban community is relatively homogeneous, dominated by a Morman culture which emphasizes education, conformity, and cooperative efforts.


The Salt Lake Teachers Association collaborated with the district board in designing the teacher evaluation system. In negotiations the association gained a promise of job security in return for accountability-based remediation and dismissal procedures. The Salt Lake teacher evaluation system relies on an annual goal-setting exercise in which the principal and teacher confer on which system, school, or personal goals the teacher will pursue for the coming year. Observations may focus on either the adopted goals or a list of teaching criteria included in the collective bargaining agreement. The SLTA developed the list of "teacher expectancies" that provide the basis for evaluation decisions. The evaluation system does not begin to operate in a highly formalized manner unless a teacher is performing poorly. Prior to formal remediation, a principal may initiate informal remediation, at which point observed deficiencies and a specified plan of action are put in writing, and the teacher is given additional supervision and assistance. If informal remediation succeeds, no record of the process enters the teachers personnel file. If it fails, the teacher receives formal remediation.  

Though the principal is responsible for evaluating teachers and for instigating procedures for those who are performing poorly, once the remediation begins, however, expert teachers who are chosen for their outstanding teaching ability in the appropriate teaching area assume a large portion of the

assistance function. The two to five month remediation process has resulted in the removal of 37 teachers judged inadequate over the past nine years and the reinstatement of nearly an equal number. Although originally the entire remediation team had to agree to the dismissal of a teacher who failed remediation, more recently the SLTA has asked the principal to make the decision after confering with the team.  

Salt Lake City achieves efficiency through shared decisionmaking rather than central policymaking. This communal philosophy, more distinctly known as 'shared governance' lends itself well to the supportive role teachers must play in a teacher remediation process. Further evidence of this shared decisionmaking is exhibited by the following policy decisions. The president of the SLTA is invited to attend the superintendent's staff meetings. Teachers have an equal vote on instructional committees dealing with in-service training, administrator hiring, class size and salaries. Further, teachers have primary responsibility for curriculum development and for assisting new teachers in classroom improvement efforts. Experienced teachers receive renumeration and release time to help and counsel new teachers.  

In addition, a variety of mechanisms make a teacher's classroom performance a legitimate domain of interest for all parents of school children. Teachers are required to provide a written statement to parents of what they plan to do in each  

147. ibid, p. 30
school year. Also a "review-of-services" process allows anyone to raise a complaint about any practice for investigation by a third party. About one-third of all teachers placed on remediation in Salt Lake were identified through the review-of-services process. Because of the openness of the system, poor performance is usually noticed and addressed. A system couldn't have more to offer in accountability.

Lake Washington: An Engineering Approach

At the hub of the aerospace industry, this district's clientele understands a problem-solving approach to educational reform. Unlike in Salt Lake City, Lake Washington's teacher evaluation system is highly structured from beginning to end. Developed in 1976 because of a state mandate, the evaluation system employs the state criteria in a checklist that the principal uses in observation of each teacher twice a year. Each observation includes a pre- and post-conference with each classroom visit. If a less than satisfactory evaluation is given, the principal outlines a detailed personal plan which includes assistance from a master teacher, in-service classes, and specific reading assignments. If this does not bring improvement then the teacher is placed on probation. The principal meets weekly with a probationary teacher to monitor progress toward specific performance goals. At the end of the semester, the principal along with central office supervisors, decide if that

148. ibid, p. 27-29
teacher will continue with the school district.\textsuperscript{149}

Staff development is tightly linked to teacher evaluation. In 1983 the district of about 18,000 students spent over $1 million or about 2% of the district's budget on staff development. A large portion of this is spent on Madeline Hunter's "Instruction Theory Into Practice" (ITIP) approach. Skilled teachers designated as ITIP trainers help maintain a uniform approach in the system. Each teacher was initially given a 30 hour ITIP training course and are expected to follow-up with nine credit courses from in-service each year.\textsuperscript{150}

Principals are evaluated on their staff development skills and on how well they evaluate teachers. When a principal feels a teacher needs assistance, he can call on a full-time ITIP trainer who receives release time to provide this assistance. In spite of this available assistance district administrators state that the evaluation system has resulted in the 'counselling out' of about 40 teachers over a four year period, a figure representing about 5 percent of the total teaching staff. Nonetheless, the combined approach of centrally determined goals and performance standards together with a commitment to staff development has brought a 20-percentile gain in pupil achievement over the same period.\textsuperscript{151}

\textsuperscript{150} ibid, p. 80
\textsuperscript{151} ibid, p. 81
Greenwich: Performance Goal Approach to School Management

Greenwich, in the state of Connecticut, is a wealthy, suburban, district of 7500 students. The school representatives in this community recognize that evaluation takes time if it is to be done well. It has set a ratio of 1 evaluator to 20 teachers and has designated teacher leaders who spend half their time teaching and the other half on evaluation to perform these duties. The assistant superintendent reads and critiques each teacher evaluation report for its thoroughness and specificity as well as how well the evaluations match up with lists of marginal and outstanding teachers that the principal include in their annual school assessment report. Each year teachers, in consultation with their principal or teacher-leader (a teacher with part-time administrator status), set their own individual goals, make plans for achieving these goals, and prepare means for measuring whether the goals have been accomplished. Although teachers may choose system goals, the evaluation process is intended to foster individual growth. 152

The evaluation process includes at least one observation and three conferences per year. Teachers complete a self-evaluation report and the evaluator completes an open-ended report based on both the specific goals and the general teaching guidelines of the collective agreement. A poor rating may result in a teacher

being placed on marginal status; however this action seldom occurs. The reasons stated in the study are, "possibly because of the evaluation process or, perhaps because the district's teaching force is highly experienced and highly educated." 153

The Greenwich Education Association played a central role in developing not only the district's evaluation system but the state's as well. This approach instituted in Greenwich in 1971 was adopted in 1974 as part of Connecticut's teacher evaluation requirements. 154

Toledo: A Strong Teachers Union Approach to Evaluation

The last district, Toledo, Ohio, is a working class town with a strong teacher's union. As elsewhere, Toledo's evaluation system was created in response to public demand for quality education. The difference in this system however is that it is the Toledo teachers organization that "took the lead in defining and enforcing standards of professional conduct and competency." 155 In the Toledo teachers evaluation system the resources are targeted on the first-year teachers (interns) and tenured teachers assigned to an 'intervention' program. Skilled 'consulting teachers' evaluate new teachers and the experienced teachers having difficulty. Then the consulting teachers observe

---

153. ibid, p. 34
and confer with these teachers at least once every two weeks for the period of the internship or intervention. The consultants are released from classroom teaching full or part-time for up to three years. An annual allocation of $80,000 supports the costs of substitute teachers and other materials needed in assisting the interns and the intervention teachers.\textsuperscript{156}

Even though the stated purpose of this evaluation program is to promote professional growth among its teachers, the program also serves as a basis for decisions regarding a teacher's contract status and continued tenure in the district. For example, over a two year period (1982-83), 4 of 66 intern teachers were not rehired and surprisingly 4 of 10 intervention teachers were removed from teaching.\textsuperscript{157}

Principals evaluate teachers annually until the teacher reaches tenure, and once every four years thereafter. The principal and the union's building committee jointly decide the assignment of a teacher to intervention. The assistant superintendent and the president of the Toledo Federation of Teachers must concur with the decision.

The prominent role played by Toledo teachers in the district's evaluation program has increased teacher input into the shaping of other educational policies such as curriculum development. Thus teacher collaboration has helped to maintain a balance in educational policymaking in that district.

\textsuperscript{156} ibid, p. 85
\textsuperscript{157} ibid, p86
Similarities and Differences of Successful Evaluation Systems

These four districts address the task of teacher evaluation from somewhat different modes. For example, the processes by which judgements are made, the instrument used, the number of observations required, and the criteria upon which such decisions are made differ considerably. Greenwich and Lake Washington abide by state mandated standardized criteria while Salt Lake City and Toledo developed 'minimum competence' criteria from the expertise of its teacher organizations. As well, the relation between the teacher evaluation process and other district personnel activities such as staff development and instructional management leadership have strikingly different orientations. The Greenwich 'management by objectives' approach is diametrically opposite to the Mormon 'shared governance' approach to decisionmaking. Finally, these four districts were purposely selected on criteria which differed markedly in terms of student population, financial circumstances, and political environment.

Despite these differences in context, the four districts follow certain common practices in implementing their teacher evaluation systems. All four districts recognize that the key obstacle to successful evaluation is time, or more specifically the lack of it for observing, conferencing with the evaluatee and for getting immediate expert assistance to those teachers who are in need. As described, each of the four districts has taken the proper safeguards to insure that funds are available to procure
the necessary evaluators and master teachers so that evaluation occurs as dictated and that the remediation process is adequate.

Evaluator competence, has been described in this study as probably the most difficult element of the process, requiring two qualities: "the ability to make sound judgements about teaching quality and the ability to appropriate, concrete recommendations for improvement of teaching performance." All four districts, recognizing the complexity of this process, divide the function between principals and expert teachers. The study concluded that this was essential for a successful teacher evaluation system. They cited several reasons, the first being time. Principals, even the most conscientious, have other equally pressing duties. They lack the time to give intensive daily supervision and remediation help. The second aspect in the division of responsibilities is that principals are not always chosen for their evaluation ability or, unfortunately, for their outstanding teaching ability. Though they may recognize the presence or absence of teaching competence, the task of providing concrete assistance to teachers requiring it often demands intimate knowledge of a particular area, something that, more often than not, a principal is not likely to possess. The logical solution to this dilemma is to assign this task to one who has previously excelled in these areas of pedagogy. In Salt Lake City, Lake Washington, Greenwich, and Toledo they are known as expert teachers, ITIP trainers, Teacher-leaders, and expert consulting

158. A. Wise et al., op. cit. p. 40
159. L. Darling-Hammond, op. cit. p. 87
teachers respectively. In all cases committees and administrators together carefully choose the variously titled master teachers on the basis of teaching competence and interpersonal skills.

Other considerations in dividing these responsibilities are the often cited role conflict which precludes principal's serving as both judge and counsellor and the 'halo effect'. According to Barnabé,

Plusieurs évaluateurs sont victimes de l'effet de halo dans leurs évaluations. Ils ont tendance à penser qu'une personne en général est, ou bonne, ou mauvaise; celle-ci obtiendra la même évaluation sur tous les traits évalués. 160

The extensive nature of the collaboration between teachers and administrators on evaluation techniques, procedures and goals minimizes these effects. Principals also attend a two week workshop every summer which includes study of ITIP techniques, clinical supervision skills, and evaluation methods.

In all four districts the teacher organizations collaborated with the administration in the design and implementation of teacher evaluation process. The evaluation processes had, as their primary purposes, improvement of instruction and accountability. However, in the element of fairness, real opportunities for improvement were enacted before a teacher was dismissed.

160. C. Barnabé, op. cit. p. 182
Finally, what made these systems work was the attention given to four factors: a) each district was totally committed to teacher evaluation, especially the top levels of the organization which instituted concrete mechanisms for translating that insistence into action; b) procedures for ensuring evaluator competence; c) collaboration with teacher organizations in the development of the evaluation program; d) time - the key obstacle to successful evaluation. These districts create time for evaluation. Attention to these four factors has elevated the teacher evaluation process in these districts from what is often a superficial lip-service exercise to a meaningful process that produces meaningful student achievements.

In the next section an attempt will be made to propose a practical model that incorporates these important elements. However, consideration must be given to the constraints within which such a model must operate. In New Brunswick, unlike our American counterparts, teacher unions are extremely reluctant to participate in an evaluation process that could ultimately lead to the dismissal of a tenured teacher. However, attempts have been initiated to move away from unilateral practices where the teacher is the passive subject of observation and assessment. However, as has been made abundantly evident here, the development and implementation of any successful evaluation system resides with the two different levels of hierarchy, the teacher organization and provincial government.
CHAPTER 4

Establishing a model

For a teacher evaluation system to be useful to a district and at the same time be perceived as credible to the teachers, administrators, and the community, it must begin by addressing four basic elements: a) the school district must be totally committed to the teacher evaluation system; b) there must be collaboration between the teachers' organization and the district in the development of the teacher evaluation procedures and criteria; c) the system must include procedures for ensuring evaluator competence; d) the school district must involve expert teachers in the supervision and assistance of their colleagues, particularly beginning teachers and those in need of remedial assistance.

The School District Must be Totally Committed

In a school district, a teacher evaluation system will work as intended only if it matches the fundamental operating assumptions of the district's educational goals, conception of teaching, and community expectations. For example, a highly decentralized, district (loosely coupled system) should not adopt a teacher evaluation process that stresses adherence to centrally determined goals and uniformity of instructional objectives. Though at first glance this conclusion may appear obvious, some studies indicate otherwise:
the educational landscape is littered with remnants of unsuccessful procedures produced by bygone fads, administrators, and policies. The procedures failed—that is, lost their relevance and ceased to be faithfully implemented—in part because they did not serve the school system's more fundamental operating assumptions.  

This is a fundamental decision that most probably should be made by the district superintendent and school administrators, in conjunction with the district teachers organization before any attempt is made to establish a teacher evaluation system. If the district wants teachers to take responsibility for their own professional growth, a likelihood that one cannot assume, then the district and teacher organization must commit themselves to using teachers as well as administrators for evaluators and the school district must translate this commitment into resources. Without the resolution of these philosophical differences, and without the top-level commitment of school district leadership and resources as needed, teacher evaluation will surely deteriorate to its usual formal, meaningless exercise.

For School District No. 50 in New Brunswick, obtaining the necessary resources for implementing a truly effective teacher evaluation system will be a formidable task, regardless of the commitment by school district officials. The district budget is tightly tied to provincial priorities and any deviation from these 'norms' could only occur with a similar commitment by the provincial department of education toward such an endeavour. This

endeavour would only be attempted (piloted) if provincial personnel felt there was a possibility of eventually implementing a similar program province-wide. Nonetheless, there are advantages to piloting this project here. The district is small, requiring less funds to initiate such a project and unlike other districts in the province, there are many young teachers in the district who could greatly benefit from such an undertaking.

Teacher Involvement Improves Quality of Evaluation System

Communication and mutual understanding are vital in the development of an evaluation system. Not only administrators but teachers alike should be aware of all of the aspects and implications of the teacher evaluation system being used in their school district. There is no better way to bring all these elements to the forefront than to involve teachers in the development of the process. The idea that teachers participate in the formation of policies which affect them is not new. Ella Flagg Young, in an address to the National Education Association (NEA) in 1907, suggested that teachers will be "stronger in their work when they have a voice in the planning of the great issues committed to their hands." While it would seem that the NEA has had ongoing resolutions on teacher evaluation, the opening sentence still reads: "The NEA believes that it is the major

responsibility of educators to participate in the evaluation of the quality of their services." As early as 1974, the American National School Public Relations Association (NSPRA) reported a trend in teacher participation in evaluation:

The growing practice is to involve teachers in the establishment of evaluation programs. The unilateral imposition of evaluative standards on the teacher by the administrator is going the way of the dinosaur.

There are many studies, reports, and professional articles which call for teacher participation in evaluation procedures from the planning stages to their implementation. Dale Bolton, however, has dealt extensively with research and practice in developing evaluation procedures. His summary gives overwhelming support to the concept of involving teachers in the development of an evaluation process:

Who will decide what is important in teaching? Certainly, teachers should participate in this decision and in the total design of evaluation procedures. When they do, a better plan develops, teachers are committed to the procedures, they know what they are to do, and they know what will be evaluated. School districts that have involved teachers in planning have found that teachers output and procedural goals are better understood and attained when they are developed cooperatively and written in precise terms than when they are determined unilaterally or written in very general language. There is more commitment to goals and procedures when the goals are specific and attainable and when the people who are to accomplish them participate in establishing them.

As well, other studies and reports have noted "that involving teachers in developing a teacher evaluation process is related strongly and positively to teacher job satisfaction and satisfaction with the new teacher evaluation process."  

Collaborating with administrators and board officials toward development of an evaluation system can be a two edged sword for teachers. As cited by numerous authors and experts, it's important for teachers to give input to the establishment of procedures and criteria for such a process, but after the proceedings are in operation they, unlike others, have to live with them. With such overwhelming evidence favoring a cooperative approach to establishing a teacher evaluation system, one would be foolhardy to do otherwise.

Though close collaboration with teachers is fundamental in developing an evaluation model, it is equally important that the cooperative effort be oriented in the proper direction. Along with having to work together on basic questions of philosophy and general objectives of the evaluation system, administrators and teachers will encounter a labyrinth of literature on evaluation criteria. As outlined earlier, it is important that the criteria focus on general patterns of teaching rather than attempting to narrow in on specific generic teaching behaviors because as explored earlier, research to date has failed to reveal significant findings in that domain.

165. E. Tobia, op. cit. p. 31
Research reviewed in this paper on the establishment of evaluation criteria indicates that significant benefit could be obtained by concentrating on criteria that is oriented toward these specific categories:

1) teacher planning skills
2) classroom organization and management skills
3) teacher instructional skills
4) teacher skills in pupil evaluation
5) teacher professional growth

A committee of teachers and administrators attempting to establish an evaluation model would be well advised to begin with these dimensions of performance. These domains are researched extensively and, as noted in the section "research on evaluation criteria", these overlapping categories appear to establish patterns of behavior that are essential for successful teaching.

What we must now ask ourselves is what kinds of pedagogical behavior determine a teacher’s skills in each of the phases or aspects of classroom teaching. Evaluation of these skills would need to be based on the identification of specific items of behavior relevant to that particular skill. We need to identify these broad characteristics or patterns of behavior through a set of identification items referred to by Medley, Soar, and Soar (1984) as 'summary measures'.

In this instance, we may decide to develop an evaluation system based on the consensus of the classroom teachers in single school district in which it is to be used. A committee made up of
teachers from the system along with their administrators would have to meet regularly to study list items of teaching competencies in an attempt to derive from them a list of summary measures which could be used as an instrument for teacher evaluation in their school. Before proceeding to that step, however, the committee might find it wise to survey the teachers and administrators in order to ascertain what are desirable or acceptable evaluation criteria. A model of 22 possible criteria is found in the Ontario research by Ryan and Hickcox (1980) and reproduced here could serve as a guide for such a task.

**TABLE 6**

<table>
<thead>
<tr>
<th>Acceptable Criteria for Teacher Evaluation</th>
<th>Definitely should be a criteria</th>
<th>Probably acceptable</th>
<th>Doubtful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional skills, strategies</td>
<td>87</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Skills in diagnosis of student learning difficulties</td>
<td>68</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td>Professional ethics</td>
<td>60</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Attention to duties</td>
<td>76</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>Subject matter competencies</td>
<td>71</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>Skills in planning - curriculum planning, teaching techniques, classroom organization, assignments</td>
<td>80</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Student evaluation techniques(tests, grading)</td>
<td>68</td>
<td>27</td>
<td>5</td>
</tr>
<tr>
<td>Classroom atmosphere</td>
<td>83</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>Personal qualities</td>
<td>51</td>
<td>35</td>
<td>14</td>
</tr>
<tr>
<td>Student cognitive development and/or growth(Pre- and post-testing)</td>
<td>34</td>
<td>35</td>
<td>31</td>
</tr>
<tr>
<td>Student attitudes and affective development</td>
<td>35</td>
<td>39</td>
<td>26</td>
</tr>
<tr>
<td>Category</td>
<td>28</td>
<td>51</td>
<td>21</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Relationship with community/parents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom management</td>
<td>60</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Interpersonal relations</td>
<td>44</td>
<td>43</td>
<td>13</td>
</tr>
<tr>
<td>Contribution to the total school effort</td>
<td>63</td>
<td>33</td>
<td>4</td>
</tr>
<tr>
<td>Curriculum development and implementation</td>
<td>35</td>
<td>49</td>
<td>15</td>
</tr>
<tr>
<td>Selection and use of assignments</td>
<td>40</td>
<td>49</td>
<td>11</td>
</tr>
<tr>
<td>Organization of field experiences</td>
<td>20</td>
<td>44</td>
<td>36</td>
</tr>
<tr>
<td>In-class student interactions</td>
<td>35</td>
<td>38</td>
<td>27</td>
</tr>
<tr>
<td>Classroom appearance</td>
<td>37</td>
<td>39</td>
<td>25</td>
</tr>
<tr>
<td>Personal appearance</td>
<td>33</td>
<td>41</td>
<td>26</td>
</tr>
</tbody>
</table>

More than half of the respondents identified 11 of the 22 possibilities as components that definitely should be criteria for teacher evaluation. These items have been asterisked in the table. There were some interesting responses by the Ontario teachers. They viewed skills in diagnoses of student learning difficulties and student evaluation techniques as essential criteria. Yet these criteria are not currently used widely. Research on evaluation criteria also concurs with the teachers' assessment. Some other criteria that are used widely were not ranked nearly as high: personal appearance, classroom appearance, and additional formal training. Even for districts which are not establishing new evaluation systems, it would seem advisable to review present evaluation criteria if they plan on adopting a more consensual approach to teacher evaluation.
An example of a teacher evaluation model for a N.B. district school which utilizes a list of items pertaining to the 5 categories outlined in this paper as a basis for formulating objectives and which also respects the acceptable criteria designated by the Ryan and Hickcox (1980) survey could read as follows:

**JOHN CALDWELL SCHOOL**

Grand Falls, N.B.

**TEACHER EVALUATION FORM**

Teacher ______________________________ Date ____________

Grade _______ Subject _________________ Period 1 2 3 4 5 6

Descriptive statement:

This evaluation form consists of a sample of items from five sets or patterns of teaching behaviors. The items included in this evaluation were chosen by a committee consisting of school teachers and administrators. Evaluation comments must be collected/document over a two week or one teaching unit period and consist of 4-6 classroom observations. In order to maintain minimal subjectivity, evaluators will simply indicate with an "A" for each teacher performance that meets or exceeds a criteria requirement; or an "NI" for needs improvement if the evaluator feels that the collected or documented data does not satisfactorily meet a criteria requirement.

**PART 1**

1) **Teacher Planning Skills**

a. Teacher plans well-organized lessons which provide instruction relevant to the stated objectives. _________

b. Teacher gathers ideas from a variety of sources in planning instructional activities. _________

c. Teacher plans lesson to the interest level, needs and experiences of the students. _________
d. Teacher plans sufficient reinforcement activities so that students can meet predefined standards of student performance.

e. Teacher organizes a variety of teaching strategies and materials which stimulate pupil attention and response.

2) Classroom Organization and Management Skills

a. Teacher maintains a learning environment which provides optimum conditions (pleasant, attractive, stimulating, and orderly) for the comfort, health, safety and welfare of the students.

b. Teacher maintains discipline in a consistent, fair, and positive manner which fosters the development of self-control.

c. Teacher promotes a positive self-concept in pupils by providing opportunities for all students of participating in learning activities.

d. Is highly enthusiastic and supportive of student contribution and encourages students to examine and to respect differing points of view.

e. Engenders self-pride in students by eliciting order in assignments, worksheets, tests, and care of materials and equipment.

3) Teacher Instructional Skills

a. Teacher speaks clearly and fluently, and communicates ideas effectively.

b. Teacher exhibits a thorough knowledge of his/her subject area and exhibits an understanding of the principles of learning for the grade level being taught.

c. Teacher keeps the "difficulty level of instruction" appropriate to the level of student achievement.

d. The teacher presents material in a well organized fashion and leads the instructional activities in order that class time is used efficiently.

e. Teacher is able to question skillfully by eliciting responses indicative of high level cognitive learning or divergent thinking as opposed to students merely regurgitating facts.
4) Teacher Skills in Pupil Evaluation

   a. Teacher maintains a comprehensive system of recordkeeping that provides meaningful student progress to students, administrators and parents.  

   b. Teacher uses sufficient quantities of objective data to arrive at indicators of student progress.  

   c. Teacher uses test information as a source of feedback on the effectiveness of instruction and uses such information to modify teaching strategies with a class or with individual students.  

   d. Well-placed and sincere praise, along with moderate amounts of corrective criticism, is used in grading assignments and subjective tests.  

   e. Teacher provides prompt feedback to students from tests and provides students an opportunity to review tests results. For feedback purposes, frequent and shorter tests are preferable.  

5) Teacher Professional Growth

   a. Teacher strives for improvement through positive participation in local district inservice and curriculum activities.  

   b. Teacher is cooperative with colleagues and administrators in planning and implementing school and/or instructional activities and shares ideas, materials, and methods of instruction.  

   c. Teacher assumes responsibilities outside of the classroom as they relate to school and utilizes community resources in instruction.  

   d. The teacher is committed to the primary goal of assisting pupil growth and is available to assist students outside of classroom instructional time.  

   e. Implements a plan for self-improvement by identifying strengths and weaknesses in his/her teaching and seeks assistance from peers or administrators for particular instructional problems.
Descriptive statement:

The preceding appraisal criteria are to be used as an inventory to extract information about a teacher's present performance and to serve as a point of departure for initiating goals or objectives for ameliorating future teaching efforts. In this phase, an administrator along with the teacher being evaluated will jointly determine specific objectives the teacher will be expected to achieve for the ensuing school year. A brief description of the goals or objectives to be accomplished followed by a description of the responsibilities to be taken toward the accomplishment of these objectives will be included. The criteria for assessing the attainment of set objectives must also be mutually agreed upon and stated below.

1. Goals or objectives that are to be attained.

2. Criteria for assessing degree of goal attainment.
3. Modifications made during interim conferences. (Include dates)

4. Additional comments:

Signed: ___________________     ___________________     Date __________
Evaluatee                        Evaluator

Signature indicates only that both parties have read this completed form.
PART 3

ASSESSMENT OF ACCOMPLISHMENT (to be completed April-June)

5. Have the objectives been achieved?

6. Other Comments:

signed: ___________________ ___________________ Date _____
   Evaluatee                    Evaluator
The items for part 1 were derived from work done by R. Manatt, A. Korshalla, the Georgia TPAI instrument, along with some of my own. The part 2 and 3 sections were developed from ideas contained in the Fairfield Connecticut Teacher Evaluation Program. But as remarked by Medley, Soar & Soar (1984) "If you are contemplating using the consensual approach to the development of a set of behavioral indicators, there is a considerable body of relevant literature." 166

An evaluation should be done over two weeks or one teaching-unit period. In the pre-evaluation conference the lesson plans and objectives for the evaluation period along with other mutual concerns should be previewed. Next, 4 to 6 classroom observations are necessary over the two week period, as explained earlier, to insure validity. Following this, a structured post-conference would allow both parties to discuss the achievements of the students in relation to the objectives set out in the lesson plans. The discussions should also involve the manner of testing, test scores as well as an analysis of future presentations in light of present student achievements. An evaluation done in this manner during the fall can easily lend itself to using the 5 categories of evaluation criteria as an inventory to extract information about a teacher's present performance and serve as a point of departure to initiate goals for ameliorating future teaching efforts.

Precautions have also been taken in establishing this consensus approach model to limit the amount of subjectivity and biases. For example, the criteria in part 1 uses little presage criteria and is rated by the evaluator simply with an "A" for each teacher performance that meets or exceeds a criteria requirement, or an "NI" for needs improvement. This eliminates having the evaluator decide whether the teacher's performance is excellent, very good, good, fair, etc. which can be highly subjective. In spite of these precautions however, teachers might, with understandable concern, find that the evaluations still remain highly subjective. In such instances, there remains another approach that may be used as an alternative or in conjunction with the consensus approach, that being the structured observation systems.

Types of Structured Observation Systems

Over the years three basic types of structured observation systems have evolved: category systems, sign systems, and multiple coding systems. Of these the most familiar is the 'category system' developed by Flanders (1970). It consists of 10 items or categories which have two different types of influences as the central distinction for the verbal interaction between the teacher and pupils. 'Direct teacher interaction' refers to a one-way style of communication in which the teacher gives information and directions and expects pupils to follow them. 'Indirect teacher influence' refers to a style which permits pupils to have input into the direction and content of the discussion, and in
which the teacher accepts or uses at least some of the ideas they suggest. The teacher exerts her influence by deciding which ideas will be accepted and which will be encouraged by extension or expansion.

An observer, using the Flander's system, records the category number that describes what is happening every three seconds. An example of an observer's task is shown below by the short segment of interaction presented in figure 3 along with the codes in table 7.

**TABLE 7**

<table>
<thead>
<tr>
<th>Summary of Categories for Flander's Interaction Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TEACHER TALK</strong></td>
</tr>
<tr>
<td>1. ACCEPTS FEELING: Accepts and clarifies the feeling tone of the students in a nonthreatening manner. Feelings may be positive or negative. Predicting or recalling feelings is included.</td>
</tr>
<tr>
<td>2. PRAISES OR ENCOURAGES: Praises or encourages student action or behavior. Jokes that release tension, but not at the expense of another individual; nodding head, or saying &quot;um hum?&quot; or &quot;go on&quot; are included.</td>
</tr>
<tr>
<td><strong>INDIRECT INFLUENCE</strong></td>
</tr>
<tr>
<td>3. ACCEPTS OR USES IDEAS OF STUDENTS: Clarifies, builds, or develops ideas suggested by a student. As teacher brings more of his own ideas into play, shift to category 5.</td>
</tr>
<tr>
<td>4. ASKS QUESTIONS: Asks a question about content or procedure with the intent that a student answer.</td>
</tr>
<tr>
<td>5. LECTURING: Gives facts or opinions about content or procedures; expresses her own ideas, asks rhetorical questions</td>
</tr>
<tr>
<td>6. GIVING DIRECTIONS: directions, commands, or orders with which a student is expected to comply.</td>
</tr>
<tr>
<td><strong>DIRECT INFLUENCE</strong></td>
</tr>
<tr>
<td>7. CRITICIZING OR JUSTIFYING AUTHORITY: Statements intended to change student behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what she is doing; extreme self-reference.</td>
</tr>
</tbody>
</table>
8. **STUDENT TALK - RESPONSE:** Talk by students in response to teacher. Teacher initiates the contact or solicits student statement.

9. **STUDENT TALK - INITIATION:** Talk by students which they initiate. If "calling on" student is only to indicate who may talk next, observer must decide whether student wanted to talk. If he did use this category.

10. **SILENCE OR CONFUSION:** Pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

---

Summary of categories for Flander's interaction analysis.

Source: Soar and Soar (1982), Climate and Control System.

---

<table>
<thead>
<tr>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

*Source: Soar and Soar (1982), Climate and Control System.*

*Fig. 3* An example of a Flander's interaction analysis matrix

*Source: Soar and Soar (1982), Climate and Control System.*
To facilitate interpretation of the interaction, a sequence of numbers which represent the interaction under the heading 'Pairs of Codes' is shown at the left of the matrix. We can infer a number of things about this brief segment of interaction from the matrix. First, the teacher asked a question (her own idea) which lasted more than three seconds (the 4/4 cell). Following a student's response, she again asked another by expanding on the student's idea and again a student responded (3/9). The second student's response lead to a temporary pause (9, 10), then another student followed by giving a somewhat lengthy response (9/9). These, again, were expanded upon by the teacher eliciting another response from a student (3/9). Rather than continuing with ideas which the pupils had introduced, the teacher introduced an idea of her own which probably had the potential of taking the discussion in a different direction (5/5).

What can be deduced about the teacher from such a short recording segment? Surprisingly, a great deal. The teacher initiated student interest in the topic by allowing students to contribute to the development of the lecture. More information is also revealed by the absence of certain categories. For instance, the matrix does not indicate a teacher with management problems or one that is required to give directions to correct behavior. A build up of cells represented by the intersection of rows and columns 6 and 7 would represent these difficulties, and a build up in the 7/7 cell would indicate extended criticism by the teacher.
It should be understood that inferences drawn from such a brief matrix only suggest a certain kind of teacher. A set of more extensive samples of interaction would have to be recorded in order to establish that the pattern is stable and reliable.

A second type of structured observation that has been widely used, though perhaps not as widely as the category system is the 'sign system'. Instead of a set of categories into which any behavior must be classified, a sign system is a list of narrowly defined behaviors called signs, each of which is considered relevant to a dimension of behavior that is to be measured. If the observer sees one of the listed behaviors occur, he records it. Behavior not on the list is ignored. Observation sessions are of short durations of between two to five minutes. During each period an observer records the listed behaviors he observes but not how often. Frequency of listed behavior is dependent upon number of times it occurs within the set of observations. In this way an observer's task is easier, making him able to cover a greater range of behaviors such as nonverbal as well as verbal interaction, groupings, movement, etc. Thus sign systems contain detail not possible on a category system.

There are many sign systems in use today. An example of a system capable of deciphering information effectively on a teacher's management skills is the Soar & Soar (1982) 'Climate and Control System'. As implied by its name, it records teacher management behavior, pupil responses to it, and the expressions of affect by both teachers and pupils. Numerous other sign
systems exist which might be considered for measuring any number of other dimensions of instructional skills.

The last type of structured observation system is the multiple coding system. This system was originally designed to produce objective records of the experiences individual students have in a school. The observer, using a tape recorder with a cassette that emits a signal at 20 second intervals into an earplug, records what is happening with a pupil by marking the appropriate spaces on previously prepared recording forms. This process is repeated until five events have been recorded, at which time the observer records "context" on the back of the recording form. These experiences are described in terms of 11 items called 'words'. These clarify relationships between child and teacher, classify communication according to medium, indicate level of attention of pupil to task as well as other dimensions depending on the instrument used. By the time the recorder finishes all 11 words the 14 seconds allowed for this would have almost expired and the observer would have to begin watching the child again for the next of five recordings. The context side will indicate what was going on elsewhere in the classroom while the child was having the five experiences recorded. The role of the teacher in reference to the class would be included as well as the learning material used, the general climate of the class during this period along with other statements again depending on the type of information sought by that particular instrument. Once the cycle for this child is completed, the recorder then turns to a second child and repeats the process.
A typical multiple coding system yields much more detail about each experience than a simple category system. For example, more than two million different statements can be generated using the 11 item word STARS (Spaulding Teacher Activity Schedule) instrument spoken of above. It is a 1983 revision of earlier versions reviewed in 'Evaluating Classroom Instruction' by Borich & Madden, 1977. According to the information found in the appendix of Medley, Coker & Soar (1984) a study of the generalizability of STARS show correlations reaching 0.80 after four observations were aggregated and 0.90 after eight observations. In training observers, coefficients of agreement of 0.80 and above are indicated. Most of these type of instruments are easily scored. For example the STARS instrument is coded on opscan sheets, similar to test answer sheets, and all scoring is accomplished using programs written for the Apple II microcomputer.

There are many appealing aspects to this type of instrument. Assuming that the primary purpose for which one uses these measurements is for the improvement of instruction, and that the strategy is to use staff development rather than personnel decisions as the primary purpose of such an endeavor, an observation instrument like the COKER (Classroom Observation Keyed for Effective Research) can yield not only a profile of a teacher's strength and weaknesses across a spectrum of

competencies but could profile a whole staff of teachers on any one competency. All this, done on a microcomputer which can be procured in any school; the implications are mind boggling! For example, one could plan professional development around results obtained from particular competency deficiencies found; then follow-up the workshop by using the pre and post-test approach to see if the professional development workshop was effective.

Another possibility is using such a system to monitor curriculum changes. When a new program is introduced, before and after observations could be done to see if the nature of the new curriculum is resulting in the types of teaching behaviors desired, or whether unintended side effects are occurring. The possibilities are limitless.

However, what was originally anticipated by using structured observation systems for evaluation was objectivity. According to Medley, Soar & Soar (1984),

> Because the observer's role is restricted to observing and recording, not evaluating teacher performance, it is neither necessary or desirable to employ highly trained professionals as observers. (...) Our recommendation is that you look for people who have had some classroom experience but have not been trained or employed as administrators or supervisors. (...) Promising prospects are people employed as substitute teachers or as teacher aides and graduate students in nearby schools of education.  

As can be seen by this, objectivity is maximized in this type of evaluation and cost can be kept to a minimum because administrators are not involved in the classroom observations.

168. ibid, p. 128
Not only would it be an inefficient use of an administrator's time to have him do something a less highly trained person can do as well, but it is also an enlightening experience for aspiring teachers. It is confirmed by research that:

significant changes in the classroom performance of teachers occur when the teachers themselves are trained to use a low-inference observation system. This seems to happen because the categories in the observation system provide a new "language" which the teachers can use in thinking about and monitoring their own behavior and in discussing it with supervisors. 169

However, a great deal of leadership is required of administrators if feedback to the staff is to be utilized in a positive manner. In order to accomplish this task, an administrator responsible for improving instruction must acquire a good working knowledge of any observation system he proposes to use for teacher evaluation. If the intended use of this type of observation system is with all the staff, then the appropriate role of an administrator in improving instruction would involve interpreting records, and in collaboration with his personnel, planning strategy to counterbalance the weaknesses exposed by the observational data.

At this point other elements must also be taken into consideration. Is this a beginning teacher? If so, then a second evaluation session would be required within a short period to assess the progress being made toward the predetermined goals outlined from the first evaluation session. However, if a tenured teacher was evaluated and favorable results were obtained, which

169. ibid, p. 127
should generally be the rule, then there is little need to delve further. Simply having surveyed the situation with that teacher should surely improve communication and reveal any areas of concern.

**Expert Teachers Must Provide Remedial Assistance**

A less desirable, though not to be totally unexpected dilemma that will present itself is the unsatisfactory evaluation. This aspect of evaluation must be dealt with and agreed upon by both the district and the teachers before a mutual consensus approach can be attempted toward evaluation. Quite simply, teachers in New Brunswick can only benefit by the implementation of a remediation program. At the moment, a teacher can be dismissed for "just cause" provided he/she is given a 10 day prior notice to dismissal. A letter in response to a request for clarification on this matter from a NBTA assistant executive director stipulated:

> In all the cases we have brought before the adjudicator involving qualifications or teacher evaluation, we have found your statement to be true, that is that the adjudicators would not substitute their judgement for that of the school administrators if decisions of the latter are not arbitrary or capricious. A general statement by the employer that the teacher cannot satisfactorily fulfill a given position is usually sufficient for an adjudicator.\(^{170}\)

It would be inconceivable to expect collaboration in establishing an effective teacher evaluation system without providing an extensive remediation process for those teachers experiencing

\(^{170}\) Edouard Allain, New Brunswick Teacher Association letter in response to a request for clarification, appended to thesis.
difficulty. Likewise, children need equal protection from the wrath that may befall them from unbridled pedagogy. The type of remediation program implemented again would have to be mutually acceptable to the district and the teachers. Release time must be made available for master teachers chosen for their outstanding teaching ability in the appropriate teaching area and they would assume a large portion of the assistance function.

Initiating such a process would probably require that a compromise be struck between the district and the board. By applying the process only to beginning teachers initially, the procedure would seem less threatening. Because of the lull in the hiring of new teachers over the last decade a disproportionate number of our experienced teachers are nearing retirement. If such a program was implemented now, nearly all teachers would be participants within as short a time frame as ten years. Also, an evaluation program of this sort would help diminish the gap which will soon occur in teaching by increasing communication between new and experienced teachers and creating more uniformity of instruction.

The remediation process is costly and requires that the province make allocations for such resources available in the district budget. For the past 5-8 years in N.B., we've seen the implementation of subject co-ordinators who specialize in one particular program within a district. They make sure that teachers under their jurisdiction are teaching the contents of the program correctly, and that these same teachers know how to
use the learning material correctly. It might be time to add another responsibility to their job descriptions. They, in cooperation with other well-qualified, experienced and respected teachers could coordinate such a remediation program.

Remediation is probably one of the most important elements of the teacher evaluation process. Nonetheless, if the absence of available research on remediation for ineffective teachers is an indicator, it must be the least understood. In an attempt to understand this complex task we shall focus on the causes of unsatisfactory performances; otherwise, the types of remediation may be directed at the wrong target. Steinmetz (1969) purports that unsatisfactory performances could stem from any one of three major causes: 1) managerial and/or organizational shortcomings; 2) a problem with the employee; and 3) outside or non-job-related influences affecting the employee.  

The managerial and/or organizational inefficiencies may precipitate ineffectiveness in employees simply by exerting too much supervisory control. Often superiors don’t trust subordinates feeling that they are too incapable or uncaring to accomplish difficult tasks adequately. Rigid job procedures and other control devices, designed to augment productivity, tend to stifle employee creativity and frustrate those who want to obtain feelings of achievement and responsibility. Isherwood and Hoy (1973) researched the effects such "authoritarian"  

administered schools have on employees as opposed to the more "collegial" approach of other similar sized educational institutions. This study concluded that "Teachers with professional work values and teachers with mixed work values had a significantly greater sense of powerlessness in Authoritarian schools than their peers in Collegial schools." As can be seen by this research, a good remediation program can not only serve a vital function in teacher evaluation but if implemented efficiently can act as a weathervane for what's ailing the administrative aspects of schools. If teachers that should be more productive feel frustrated because of a lack of trust, appreciation, or responsibility in the direction of the educational facility then perhaps the bureaucracy is precipitating these difficulties.

The second cause of unsatisfactory performance could simply be a failure to expend the necessary effort at being a competent teacher or it may stem from the inability of the teacher to carry out the teaching task in a satisfactory manner. This ineffectiveness could be related to organizational shortcomings because the teacher has been changed to a teaching assignment he's not properly trained for; his teaching load requires too many preparations, or there are too many "difficult" students in his courses. These causes could all be alleviated by administrative changes. However, if the problem is skill related,

173. E. Bridges, op. cit. p. 31
then Steinmetz suggests the following remedial procedures. Goal setting: apparently leads to improved performance if it clarifies exactly what is expected of the employee. Modeling: allows the teacher to observe examples of a teaching performance that exemplifies key behaviors and skills. There are several ways in which the teacher can be supplied with models—opportunities to visit and observe the classrooms of exceptional teachers, model teaching in the classroom of the teacher experiencing difficulty by the remediating expert teacher, and team teaching assignments with gifted teachers.\footnote{174} Feedback: according to Miller (1978), direct feedback is more effective than feedback that is the result of another person's judgement of performance.\footnote{175} A teacher can listen to an audiotape or view a videotape of their own classroom performance to obtain information of their verbal interaction. If the teacher feels uncomfortable with such technology then second best would be using the teacher assigned to the remediation to provide a written record of a classroom episode for analysis later.

The third major cause outlined by Steinmetz is unsatisfactory performance due to non-job-related influences. The teacher may be experiencing marital difficulties, serious emotional distress, alcohol or drug addiction, or other personal

\footnote{174. ibid, p. 32}  
pathologies. There already exists an Employee Assistance Program in N.B. which offers counselling and therapy to any provincial government employee to help them cope with their own personal difficulties. But sometimes these situations are allowed to linger on to the detriment of both parties because of an ineffective evaluation program.

From the descriptions of the three major causes outlined by Stienmetz concerning unsatisfactory performances, a teacher remediation program would be of utmost importance in dealing with the second type of unsatisfactory performance. Nonetheless, an effective evaluation system could also provide some insight into aspects of the organizational structure that are elements of tension and frustration; thus having a detrimental effect upon the global viability of the education system.

Insuring evaluator competence

According to the Rand study (1984) on teacher evaluation, evaluator competence requires two qualities, "the ability to make sound judgements about teaching quality and the ability to make appropriate, concrete recommendations for improvement of teaching performance." In the four American districts reviewed extensively in the preceding section, all recognized this dual function of evaluation and divided the function between administrators and expert teachers. In Lake Washington and Salt

176. E. Bridges, op. cit. p. 31
Lake City, principals evaluated teachers and initiated the remediation process. Expert teachers - ITIP trainers in Lake Washington and teaching specialists in Salt Lake - provided most of the help to teachers needing improvement. In Toledo, Expert teachers both evaluated and assisted first-year teachers. The principal assumed the evaluation role after the first year. In Greenwich, both principals and teacher-leaders evaluated and offered recommendations for improvement. However, their evaluations were in kind reviewed by the assistant superintendent and critiqued on how well they performed their evaluation function. Each teacher evaluation report is assessed for its thoroughness and specificity. Since teacher evaluation is the major responsibility of teacher-leaders, their continuation in that position is tied to their performance as evaluators. Greenwich has set a ratio of 1 evaluator (teacher-leader) to 20 evaluatees and the teacher-leader spends half his time teaching and the other half evaluating.

According to Bridges, evaluators need to possess special knowledge and skills which according to him are frequently overlooked in university programs for preparing school administrators. These are: 1) ability to describe and analyze what is happening in a teacher’s classroom; 2) ability to conduct appropriate pre- and post-conferences with teachers regarding their instructional evaluation; 3) ability to diagnose causes for a teacher’s poor performance; 4) ability to prescribe appropriate remediation for a classroom teacher’s deficiencies; 5) knowledge of legal aspects involved in documenting matters related to 1
through 4 in cases where dismissal and grievance procedures are invoked. 178

School districts need to take steps to insure that their evaluators possess these skills. The Lake Washington School District developed an elaborate program to teach their principals skills in analyzing instruction and conferencing as well as finding support budget for staff development. These activities are centered on Madeline Hunter's 'Instructional Theory Into Practice' (ITIP) approach. It relies heavily on videotapes and printed materials but is quite expensive. Townsend (1984), reporting on the first year implementation of a new policy of teacher supervision and evaluation in the Lethbridge Alberta school district, noted that "when questioned in November, nine supervisors were confident they had the necessary skills to conduct supervision and evaluation of teachers." 179 However by April, 13 of the 16 supervisors who participated in the project indicated that "they had limited knowledge of and experience with the techniques of supervision and evaluation of teachers." These are sixteen school-based administrators of five secondary schools in the Lethbridge district. From this, Townsend concluded;

It is possible that as supervisors became more involved in the supervisory and evaluation practices in their schools there was a corresponding increase in their awareness of the professional responsibility associated with such tasks, their complexity, and the extent to which they influence other aspects of school life; hence the reluctance of supervisors to overestimate

178. E. Bridges, op. cit. p. 37
179. D. Townsend, op. cit. p. 107
their skills and knowledge. It is not difficult to see that fulfilling the
responsibilities in this domain is a difficult and demanding task
requiring extensive training. In Lethbridge they not only did in-
service with the supervisors but also reviewed in detail what the
new evaluation process would consist of and what was expected of
the supervisors and teachers.

In-service is useful for improvising present or introducing
new techniques. However only closer collaboration with
universities will lead to profound and enduring changes. In N.B.
education students spent long periods of time with cooperating
teachers who are asked to evaluate their efforts during the
student’s practicum. Having been a cooperating teacher myself in
1984 for half a year, I must say the experience has left me with
feelings similar to those experienced by Alberta supervisors:
confident in the beginning but increasingly aware of my
shortcomings as an evaluator and pedagogical counsellor when the
process ended.

Georgia’s Department of Education in cooperation with the
University of Georgia has completed a mammoth research project
aimed at improving the quality of new teachers in Georgia. The
certification program has two elements, a Teacher Certification
Test and the Teacher Performance Assessment Instrument (TPAI)
which are used for on-the-job assessment. The program has been

180. ibid, p. 108
operating efficiently since 1980. Trainees in The Beginning Teacher Staff Development Program conduct sessions in the use of the instrument through the network of Regional Assessment Centers located throughout the state. There are two dissertation studies done on the TPAI that are noteworthy. Tobin (1980) developed and validated the 'Test of Integrated Science Processes' and used it in all of the classrooms in his study and found a TPAI total score to correlate ($r = .6$) with pupil achievement. McGarrity (1981) partitioned his sample of teachers into two groups depending on whether or not they had demonstrated mastery according to TPAI criteria. He found significantly higher achievement in classrooms taught by teachers who had met the TPAI criteria than in those classrooms were the teacher had not. Interestingly, these patterns were consistent even after statistical correction had been made for the ability of the learners.

A similar certification program for beginning teachers in N.B. would be an excellent starting point for improving both teacher evaluation methods as well as supervisor proficiencies in teacher evaluation. It would create dialogue in teacher evaluation between supervisors, provincial universities and the Department of Education. Certainly one can argue that money, expertise, and personnel are all factors that make such an undertaking seem prohibitive. However, considering the embryonic stage of teacher evaluation in N.B. few requirements other than resourcefulness, initiative, and energy are necessary to initiate a program. Of course, once a perceiveable influence has been made
a valid argument can be made for additional resources for dealing with the complexities of future efforts.
Conclusion and Recommendations

This study was undertaken in the hopes of acquiring a better understanding of the fundamental elements involved in teacher evaluation. By studying recent endeavours in this field it was hoped that maybe some new ideas might help me with my administrative duties. After completing the historical review of educational evaluation, it became abundantly clear that original intentions might have been somewhat naive. The research on evaluation criteria began to appear like a quagmire yielding no positive stances in generic teaching behaviors. The work of Elias and MacDonald (1974) on the extensive Beginning Teacher Evaluation Study (BTES) done at Stanford seemed like the only light beckoning on teacher evaluation. It pointed toward patterns rather than specific generic behaviors as being the essence to teacher evaluation.

The TPAI instrument was developed almost coincidentally from the same standpoint. More than five thousand Georgia educators worked on identifying sets of teaching behaviors which they found essential for the certification of beginning teachers. Again it was not specific generic teaching behaviors that they based their research on, but sets or patterns of behavior. The efforts which began in 1976 resulted in an instrument formatted to measure as many as fourteen sets or patterns of teaching behaviors when first piloted in 1979. With further refinement to
make it a better and simpler diagnostic tool, the number of sets or patterns of teaching behaviors has been reduced to eight. Analyses which have been done to assess the generalizability of the scores have produced rho-squared values between .7 and .85 for the eight competencies.

There are other interesting developments that though not directly part of teacher evaluation are related to the field, namely selection of candidates for the profession. Since the process of teacher certification is not based on a body of empirical knowledge, faculties of education are unable to identify the skills and attitudes required of potential teachers leaving the program. Yet the development of such a research base appears to be an antecedent for public acceptance of teacher education programs. Consequently, the curriculum for aspiring teachers is often criticized as vague, nebulous and generalized. Nonetheless, the welfare of children cannot be left to the entrustment of a teacher who may possibly possess the intellectual and social abilities necessary to be a good educator. Faculties of education must forge ahead with the major responsibility for this refinement utilizing the best available research.

Likewise, school boards are expected to utilize the utmost care in selecting candidates for the classroom. Most educators see nothing wrong with competency testing for new teachers and advancements in testing technology has made available achievement tests in any one of 35 different subject areas. These scores
could be available for school board personnel committees to help assure that the candidate they chose actually possesses the basic knowledge that he is expected to dissipate.

Research has also shown that the personnel interview is an invalid and unreliable procedure for selecting competent teachers. Recognizing this, researchers have developed the Teacher Perceiver Interview in an attempt to further refine interviews as a selection device. Though empirical research support for such a technique is still pending, a structured system of interviewing could only help provide additional feedback to board committees in making these decisions.

The teachers who have completed their practicum within the parameters of the Georgian TPAI Certification Program would certainly feel comfortable working within the collaboration process that is coming to the forefront as an established procedure for implementing a teacher evaluation program in schools today. Furthermore, they would be able to contribute significantly because of their graduate training. Sergiovanni and Craver purport that

The advent of the 1980’s has brought with it a new breed of teacher who is more assertive, and more autonomous while being less respectful of authority, less conforming and less malleable. This trend is causing and will continue to cause considerable difficulty for school executives unless present assumptions about administration, control and organization change.

Part of the organizational change the authors speak about has to do with shared decision-making, allowing those who are capable of providing input to do so.

There are some basic assumptions being made by these "less conforming and less malleable" teachers in suggesting or demanding increased participation. First, it is assumed that, when evaluation is treated as a cooperative effort among educators to improve instruction, teachers will feel that the process is worthwhile. A second assumption which closely follows the first is that a positive attitude will prevail. This attitude is vital for developing significant results towards the improvement of instruction, the primary purpose of teacher evaluation.

This paper was written in an attempt to better understand the elements involved in teacher evaluation. In the review of literature, it was deducted that teacher evaluation has primarily two basic purposes: improvement of instruction and accountability. The research on evaluation criteria did little to accommodate any preconceived ideas that there must be a firm foundation to any practice for it to be incorporated in nearly all educational systems. This paper embraces the premise that, though there seem to exist few generic teaching behaviors, sets or patterns of teaching behaviors can be established for evaluating teaching personnel.
The advent of the 1980’s has brought with it the need for a cooperative approach to teacher evaluation. A cooperative effort in establishing evaluation criteria, procedures, and instituting a remediation process can transform an evaluation practice from a meaningless ritual to a responsible endeavor responsive to the public’s needs. This new trend of the 1980’s will profoundly change the methods and techniques of administrative duties not only in teacher evaluation but in all other aspects of educational administration. In the eyes of many authors and researchers, these new initiatives were long overdue.
Based on the literature and research reviewed in this study, the author makes the following recommendations:

1. Because the literature reviewed in this study overwhelmingly designates that the primary purpose of teacher evaluation should be formative in nature, that is, aimed at the improvement of instruction, it is recommended that school boards in New Brunswick reorient their philosophies and policies for teacher evaluation programs in that direction.

2. It is recommended that the classroom observations include a pre and post-conference. These are necessary because they help create a positive attitude and a reassuring sense of trust in the evaluation process and also they reduce tension. The post-conference provides the feedback necessary for the establishment of objectives for improving instruction.

3. Though studies indicate that the majority of teachers understand the need for teacher supervision, many are indeed skeptical of evaluation reports based on one or two observations. This study embraces the premise that if supervision is to lead to improvement of instruction, it must be done thoroughly. Four to six classroom observations over a two-week or one teaching unit period are recommended and at least 5 sets or patterns of teaching behaviors, as described in this study, should be assessed during these observations.
4. Recent research indicates that a positive attitudinal change occurs among teachers when they are allowed to collaborate in the establishment of teacher evaluation programs. It is therefore recommended that a collaborating committee of school board personnel, school administrators, and teachers jointly establish the evaluation criteria, policies, and remediation processes within New Brunswick school districts' teacher evaluation programs.

5. Research indicates that significant changes in the classroom performance of teachers occur when they are trained in the use of structured observation systems. This happens because the categories in the observation system provide a new "language" which teachers can use in thinking about and monitoring their own behavior and in discussing it with supervisors. In light of this, and because training for objective observation is now being considered as a primary domain of pedagogical training, it is recommended that Faculties of Education within New Brunswick universities begin training cooperating teachers as well as their aspiring student teachers in this field.
BIBLIOGRAPHY


Brophy, Jere L. "Teacher Behavior and Its Effects". In Journal of Educational Psychology. Vol. 71 No. 6

Bush, Ronald. "Implications of the BTES". In The Generator. 1979 Vol. 9 No. 1


Dunkin, Michael J. and Biddle, Bruce J. The Study of Teaching. Montreal: Holt, Rinehart and Winston 1974


Evertson, Carolyn, Emmer, E. "Preventive Classroom Management". In Helping Teachers Manage Classrooms. D. Duke (Ed.)


Gage, N. L. and Berliner, David C. Educational Psychology. Chicago: Rand McNally College Publishing Co. 1979

Gallup, G. H. "The Eleventh Annual Gallup Poll of the Public's Attitude Toward the Public schools." Phi Delta Kappan. 1979 No. 60 p. 33-40


Good, Thomas L. and Brophy, Jere E. Educational Psychology: A Realistic Approach Toronto: Holt, Rinehart and Winston 1977


Horwitz, R. A. "Effects of the Open classroom". In 
Educational Environment and Effects: Evaluation, Policy, and 
Productivity, H. J. Walberg (Ed.) Berkeley Calif: McCutchan 
Publishings 1979

Howsam, Robert "Teacher evaluation: Facts & Folklore". In 
National Elementary Principal. November 1963 Vol.44 No. 2

Husen, Torsten, and Postlethwarter, Neville, Editors-in-chief. 
International Encyclopedia of Education. Pergamon Press, 
Toronto 1985.

Isherwood, G. & Hoy, W. "Bureaucracy, Powerlessness, and 
Teacher Work Values. In The Journal of Educational 
Administration. Vol. IX No. 1

Jones, Anthony A., Bagford, Lawrence W., and Wallen, Edward A. 
Strategies For Teaching. Metuchen, N. J.: The Scarecrow 
Press, Inc. 1979

Kaffmann, D. Futurism and Future Studies. Washington D. C.: 
National Education Association 1976

Knapp, M. S. Toward the Study of Teacher Evaluation as an 
Organizational Process: A Review of Current Research and 
Practices. Menlo Park, California: Educational and Human 
Services Research Center, SRI International 1982.

Korshalla, Andrew F. A Comparative Analysis: Tenured Teacher 
Evaluation Criteria in New Jersey School Districts and 
Research-Based Teacher Evaluation Criteria. Doctoral Thesis: 
Rutger University, New Brunswick N. J. 1984

Lewis, Anne C. Evaluating Educational Personnel. Arlington 
Va: American Association of School Administrators. ED 212 055

Lupone, O. "A Comparison of Provisionally Certified and 
Permanently Certified Elementary School Teachers in Selected 
School Districts in New York State". In Journal of 
Educational Research. 1961 Vol. 55

Madaus, George F., Stufflebeam, Daniel L. and Scriven, Michael 
S. "Program Evaluation: A Historical Overview." In 
Evaluation Models: Viewpoints on Educational and Human 

Manatt, Richard, Developing a Teacher Performance Evaluation 
System as Mandated by Senate File 205 (Iowa General 
Assembly, 1976). Association of School Boards, Des Moines 
Iowa, Sept, 1976, ERIC ED 134 222

Manatt, Richard, Palmer, K. L., and Hildengaugh, E. "Evaluating 
Teacher Performance With Improved Rating Scales". In NASSP 
Bulletin. September 1976 Vol. 60 No. 401


Zahorik, J. "The Effects of Planning on Teaching". In *The Elementary School Journal*. Vol. 71


APPENDIX A

A letter from the New Brunswick Teachers Federation
April 11, 1986

Mr. Earl Toner
8-2215, rue Banville
Les Saules, Quebec
Canada
G1P 1H6

Dear Mr. Toner

Your letter to Dwain McLean dated February 12, 1986 has recently been brought to my attention. I must apologize first of all for the delay in answering your letter; as I understand it, it was referred to several other members of the Association and Federation staff before it was determined that most of the questions were related to the Collective Agreement. For that reason, as we say in French, "le sort tomba sur le plus jeune"!

In all the cases we have brought before the adjudicator involving qualifications or teacher evaluation, we have found your statement to be true, that is that the adjudicators would not substitute their judgment for that of the school administrators if decisions of the latter are not arbitrary or capricious. A general statement by the employer that the teacher cannot satisfactorily fulfill a given position is usually sufficient for an adjudicator although the employer usually will argue that the teacher is lacking either the courses or the experience in the given subject.

In cases of dismissal for incompetence or as a result of summative evaluation, Articles 55 and 56 of the Collective Agreement specify that there must be just cause. In these cases, the onus is on the employer to show that there has been some objective type of evaluation, some attempt at improving the situation and in cases of discipline, some type of progressive discipline prior to dismissal. Generally, the five steps of progressive discipline are a verbal warning, a written warning, a suspension with pay, a suspension without pay and finally dismissal. In certain cases, however, depending on the severity of the situation, one or all of these steps may be omitted.

In Article 56, the onus would be on the employer to show that the teacher is unsuitable and that the termination is due to that unsuitability and not for disciplinary reasons. In these cases, there would still have to be some evidence of repeated incidents or situations showing unsuitability.
As far as a reasonable period of time is concerned, the only provision in the Collective Agreement in this respect is the one in Paragraph 55.05 (b) which states that any unfavourable report placed in the teacher's personal record file must be filed within 12 months of the event or situation upon which the report is based. There is no provision for a reasonable time for dismissal although any termination under Article 56 has to be indicated prior to April 30 and would normally take effect on July 1. In all other cases of dismissal under Article 55, Paragraph 55.03 provides for 10 teaching days notice prior to dismissal.

Our Agreement makes no distinction between a probationary teacher or one holding a continuing contract since both Articles 55 and 56 refer to just cause. One adjudicator has however suggested that the holder of a probationary contract may be subject to termination or dismissal for reasons that might not be considered sufficient for the holder of a continuing contract. We still maintain that the term just cause would apply to any teacher regardless of status.

I am enclosing a copy of our Collective Agreement; as far as I know, no modifications are being negotiated for Articles 55 and 56 which are the pertinent articles. I am also enclosing a copy of the section dealing with just cause from the "Labour Arbitration in Canada" textbook by Palmer which is widely used in Labour Arbitration.

I hope this will be some use to you and again our apologies for the long delay as we have been rather busy these last few months.

Yours truly

Edouard Allain
Assistant Executive Director

Encs.

EA/kc