DOCUMENT RESUME

ED 268 159 TM 860 182

AUTHOR McLean, Leslie D.
TITLE The Craft of Student Evaluation in Canada: A Report to the CEA on Policies, Practices and Uses of Assessments of Student Achievement.
INSTITUTION Canadian Education Association, Toronto (Ontario).
PUB DATE 85
NOTE 63p.
AVAILABLE FROM Canadian Education Association, Suite 8-200, 252 Bloor St. W., Toronto, Ontario, M5S 1V5 ($6.00).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Academic Achievement; *Achievement Tests; *Educational Testing; Educational Trends; Elementary Secondary Education; Employer Attitudes; Evaluation Methods; *Evaluation Utilization; Foreign Countries; Grades (Scholastic); Higher Education; National Surveys; School Community Relationship; Student Attitudes; *Student Evaluation; Teacher Attitudes; Testing Programs; *Test Use
IDENTIFIERS *Canada; Canadian Education Association

ABSTRACT

The Canadian Education Association sponsored a study to review and synthesize the literature on student evaluation; investigate changes in goals and methods; and make recommendations for improvement. Interviews were conducted in six provinces with administrative personnel, principals, and teachers. Employers and post-secondary institutions, as consumers of evaluation, were also surveyed. A model was developed for the relationships between evaluation, teaching, learning, time on task, students' self-concept and school attitudes, and teacher attitudes. Results indicated that achievement testing was common but little use was made of test results. There was increasing reliance on testing, and less on consultants, to assist in educational quality control. Employers were generally more concerned with attitudes and behavior than grades; banks and insurance companies were most interested in grades. Colleges were concerned about the comparability of schools' grades. The following recommendations were made: (1) raise the status of evaluation; (2) develop promotional materials discussing and explaining student evaluation; (3) establish a task force on evaluation and technological change within each province; (4) organize regional conferences on student evaluation; and (5) conduct a study of teachers' skills, attitudes, and feelings about evaluation. An annotated bibliography is appended. (GDC)

THE CRAFT OF STUDENT EVALUATION IN CANADA

BY LESLIE D. MC LEAN

A report to the CEA on policies, practices and uses of assessments of student achievement

Canadian Education Association
Association canadienne d'éducation
Suite 8-200, 252 Bloor St. W., Toronto, Ontario M5S 1V5
1985 $6.00

Suite 8-200, 252 Bloor St. West, Toronto, Ontario M5S 1V5
1985

ISBN: 0-920315-08-9

Cover by Fred Huffman
Printed in Canada by Twin Offset Limited

Publié en français sous le titre:
L'évaluation des élèves au Canada: un métier

CONTENTS

FOREWORD

IN THEIR PREFACE to the *Encyclopaedia of Educational Evaluation* (1973), Scarvia Anderson and her co-authors were able to write "that *almost* everything there is to say about the evaluation of education and training programs has already been said — or written — elsewhere." What mainly remained to be done with respect to program evaluation was "to make some order out of field and to bring its major concepts and techniques together in on place."

No one would have been so bold as to express a similarly rosy opinion on the status of student evaluation goals and practices in 1973 or even in 1982 when the CEA's Advisory Committee on Educational Research (ACER) turned its attention to that field. What the ACER saw in September 1982 was a considerable history of concern by students, parents, educators and others over the philosophy, methods and practices surrounding the matter of student evaluation.

Much of the public concern through the '70s and into the '80s had been focused on "declining standards" or on "accountability" or on related matters indicative of their fear that schools were not doing a good job of "quality control."

Meanwhile, many educators had been more concerned about evaluation's potential for negative impact on students' academic growth and self regard. That is to say, what they understood to be the current typically intrusive student evaluation practices gave cause for a concern that the processes of "weighing the baby" may have been interfering with "nourishing the child."

The ACER saw a pressing need to identify (or ultimately to create, if need be) evaluation practices that best promote student

achievement, positive attitudes to learning, and positive self-concept and that would also serve as vehicles for instructional program evaluation.

The ACER harboured no illusions that all that was needed was to prepare an encyclopaedia of existing literature and to fold in the corpus of underused research on the field. The committee speculated that, over time, an integrated series of studies might be undertaken that could lead to experimental implementation and validation of a developed "package of most desirable practices."

In the light of the confusion and inadequate data bases underlying many of the concerns, the ACER opted in late 1982 for a national research project to provide a review and synthesis of the literature on student evaluation, to identify (if possible) current exemplary practices, to search out pending changes in purposes or methods, and to recommend actions that would likely improve practice.

Initially, the committee also considered whether they could commission this national study to cover the pre-school to post-secondary span and to look at representative subject and skill areas. In addition, the ACER wanted to include "an assessment of the attitudes of students, teachers and parents toward various aspects of current practices." But these potentially worthwhile surveys were put aside after sober reflection on how much could be well done with the time and resources we could bring to the study's first phase.

During the winter of 1982-83, as the committee thought through what it hoped to evoke from a Canada-wide study, a consensus developed on the value of certain additional inquiries Among these was the need to clarify what evaluation information is wanted by those receiving results (e.g., parents, employers, post-secondary admissions officers) and to determine to what extent they actually used such data. The ACER also hoped to probe the relationships thought to exist among student evaluation philosophies or purposes; methods; educational accomplishment; and teachers' as well as students' purposes, attitudes and self-concept.

In the spring of 1983 a Request for Proposals, based on the ACER's outline of goals for a national study of student evaluation purposes, methods, accomplishments, problems, and prospects was circulated to 45 Canadian academics and other practitioners selected from nearly 100 known to be active in this field. A proposal by Dr. Leslie D. McLean was subsequently judged to meet most closely the spirit and substance of the committee's wishes for their first venture into the realm of student evaluation.

Professor McLean has worked in measurement and evaluation at The Ontario Institute for Studies in Education for many years and, more recently, has also served as the Head of OISE's Educational Evaluation Centre. His distinguished record is liberally sprinkled with well-received publications based on his developmental as well as research work in the field.

Perhaps it is Dr. McLean's training as a mathematician that is reflected in his concise as well as articulate reportage. The readers of his report of this study can judge for themselves. Certainly, Les's long experience in evaluation, which has been gained on both sides of the Canadian-U.S. border and in Asia as well as Europe, gives a rare breadth of perception.

Such experience, plus the help of his research assistant, David Welch, enabled him to produce in little over a year what was admittedly a multi-year sprint for most who might have essayed the course

The ACER is grateful to Dr. McLean not only for his investigation, but also for the recommendations that complete his report. The committee hopes that he is as pleased as they are that the CEA has seen fit to begin to act on one of his recommendations, namely that

> The CEA should organize a series of regional conferences for officials, teachers and trustees to discuss student evaluation

Such a conference has in fact been organized to precede the 1985 CEA Convention in Quebec City and, contingent on its success, it should not be the only one of its kind.

For their continuing confidence in our venture, the ACER is grateful to the CEA Board of Directors. They have not only voted the funds to commission and to follow through on this study but have added personal and professional encouragement and support. The ACER also appreciates more than words can express the many ways in which CEA Executive Director, Bob Blair, and all his staff have assisted this committee since it came into being in 1982.

These first fruits are the products of many minds. The members of the ACER are:

Atlantic Region

Dr. Robert K. Crocker
(Director, Institute for Educational Research and Development, Memorial University of Newfoundland, St. John's, Nfld.)

Quebec Region

Jean-Marie Pépin
(Directeur général, Commission scolaire de la Jeune Lorette, Loretteville, Qué.)

Dr Robert Lavery
(Director General, Dawson College, Montreal, Que.)

Ontario Region

Brian Burnham, Chairman
(Chief Research Officer, York Region Board of Education, Aurora, Ont.)

Dr. Madeline I. Hardy
(Director of Education, London Board of Education, London, Ont.)

Duncan Green
(Assistant Deputy Minister, Education Programs, Ontario Ministry of Education, Toronto, Ont.)

Western Region
(Alberta, Saskatchewan, Manitoba and the Northwest Territories)

Dr. Robin H. Farquhar
(President, The University of Winnipeg, Winnipeg, Man.)

Dr. S.J. Thiessen
(10428-28th Ave., Edmonton, Alta.)

Eleanor Ingalls
(Superintendent of Education, Yellowknife Education District No. 1, Yellowknife, NWT.)

British Columbia and Yukon

Dr. John H.M. Andrews
(Professor, Faculty of Education, Department of Administrative, Adult and Higher Education, University of British Columbia, Vancouver, B.C.)

Audrey Sojonky
(Executive Director, Educational Research Institute of British Columbia, Vancouver, B.C.)

G.H. Wilkins
(Assistant Superintendent Instruction, School District No. 41, Burnaby, B.C.)

National

Dr. Stirling McDowell
(Secretary General, Canadian Teachers' Federation, Ottawa, Ont.)

C.H. Witney
(Executive Director, Canadian School Trustees' Association, Ottawa, Ont.)

From 1982 through 1984 the following also served as members of the ACEP and thus helped bring this study to birth.

Louise Nielsen
(then Chairman, Yellowknife Education District No. 1, Yellowknife, NWT)

Sarah Paltiel
(then Director General, Dawson College, Montreal, Que.)

Michel Paquet
(Président, Association des directeurs généraux des commissions scolaires, Qué.)

Dr. George Podrebarac
(then Assistant Deputy Minister, Education Programs, Ontario Ministry of Education, Toronto, Ont.)

Gérard Tousignant
(Directeur général, Commission scolaire régionale de l'Estrie, Sherbrooke, Qué.)

Dr. John H. Wormsbecker
(Deputy Superintendent, Vancouver School District No. 39, Vancouver, B.C.)

I believe that Dr. McLean would wish to join me in recognizing their role in conceiving this project and nourishing it through its long gestation. I trust that they will feel pleased with what has now issued forth.

Brian Burnham,
Chairman,
CEA Advisory Committee
on Educational Research

INTRODUCTION

EVALUATION of student achievement is an important, integral part of successful teaching at all levels. Observation, questions, exercises, quizzes, tests and examinations provide teachers and learners with feedback that shapes the amount of time they spend on teaching and learning and the ways they use that time. Summary evaluations in the form of marks or grades are used to inform parents and employers about students attainments and for promotion and graduation decisions. Other than diplomas, grades are often the only record of attainment available to a student after leaving the educational institution. Universities and colleges have consistently found school marks to be the best single predictor of success in post-secondary education, and post-graduate institutions rely heavily on undergraduate grades for admission and placement. It is fitting, therefore, that student evaluation be the object of study from time to time.

Origins of the Study

In 1983, the CEA Advisory Committee on Educational Research recommended that a study be commissioned of student evaluation in Canada. The preamble to the invitation for proposals described the committee's motivation.

> Perennial concerns of students, educators, parents, post-secondary admission officers, employers, and the general public have been the philosophy, methods, practices, purposes, and results surrounding the matter of student evaluation. Some of these concerns have been focused on "declining test scores" and "accountability," while others have

addressed the negative impact of some student evaluation practices on students' academic growth and self-regard.

In light of the confusion and inadequate data bases underlying many of these concerns, the Canadian Education Association (CEA) has decided to request proposals for a national research project to provide a critical review and synthesis of the literature on student evaluation, to identify current exemplary practices, to search out pending changes in purposes or methods, and to indicate further actions that would improve the practice of student evaluation.

The research project was seen as a modest beginning on what might become a series of integrated studies. The author's proposal was selected from among those submitted, and this is the report of his study.

Attainments of the Study

The traditional academic literature on student evaluation was reviewed; a bibliography is attached as Appendix A. Some references were found beyond North America, particularly from the United Kingdom, where formal evaluation systems are well developed, but resources did not permit a review of the wider and less accessible literature in trustees' journals, teacher federation publications and occasional publications such as the *Administrator's Notebook*. A questionnaire survey of employers was carried out to probe their use of school marks in hiring decisions.

Visits were made to six provinces, spending at least four days in each one. Interviews were arranged with relevant officials in the provincial ministry (department) of education, and in several districts (boards) that (a) were accessible in the time available and (b) had active student evaluation programs. A list of these visits is in Appendix B.

In each area, an effort was made to visit at least two schools and to interview the principal and some teachers. No visits were made to classrooms. In the interviews, attention was given to evaluation policies — how explicit they were at each level, how clearly they were perceived by district and school personnel and problems arising, if any. Testing programs and in-service training opportunities were discussed, and everyone was asked what his or her criteria were for good evaluation. Classroom teachers were asked about policies, especially school policies, about practice and about problems.

The study had as one of its objectives "the development of a model of the causal influences of evaluation on student accomplishment and self-concept and on teacher and student attitudes, and the relationships among such variables." Such a model is presented in the section entitled *A Model of Evaluation in the Teaching-Learning Process*

The effort to identify current exemplary practices did not succeed. Student evaluation is too complex and decentralized, and there is not sufficient consensus on quality criteria for this author to offer examples. There are exemplary programs, no doubt, but the design of the study was inadequate for detecting and documenting them. A survey would have to be augmented by case studies, and such an effort would have gone well beyond the resources of the present project.[1] A discussion of the issue is found in the section *Contemporary Quality Criteria and Future Trends*.

Thus, this report is in the nature of a synthesis of information from the literature and the interviews, inevitably filtered through the perspective brought to it by the author. The report begins with the views of consumers of student evaluation, and this is followed by the model of causal influences. A conception of teaching and evaluation is then discussed to bring out the salient features more clearly. This is followed by a discussion of quality criteria and future trends. The report concludes with a summary of the main points and with five recommendations.

[1]See, for example, the Science Council's Background Study 52, *Science Education in Canadian Schools*, Vol III, *Case Studies of Science Teaching* (Ottawa: Canadian Government Publishing Centre, 1984). Evaluation of student achievement gets scant mention, however

CONSUMERS OF STUDENT EVALUATION

THERE ARE two main groups of consumers—employers and post-secondary institutions. Parents are vitally interested consumers, of course, but they are a special group whose interests pervade most parts of this report. In this section we report on uses by more objective outsiders.

Employers

A sample of 100 companies was drawn from the *Financial Post Directory* in such a way that both small and large firms were represented, from many parts of Canada. A sample from this directory will be biased toward better established, mainstream companies, but given the project's limited resources it seemed a sensible choice. A two-page questionnaire was sent to each one, addressed to the "Personnel Department" (see Appendix C). They were asked what sorts of use they made of school marks in their hiring or promotion decisions and what other sorts of information they would like that they did not now have.

The questionnaire was short and open-ended for two reasons. First, from opinion polls we knew that companies were mainly concerned about the work habits of their employees. What we wanted to determine was whether they looked at marks at all in their initial hiring decisions, and, if so, what weight they gave to them. By providing an opening, we hoped to capture any other strong feelings personnel officers had about marks and schools in general.

A total of 50 replies was received, a not uncommon rate of return from a mailed questionnaire with no follow-up other than one postcard reminder. The 50 included a satisfying range of types and sizes of companies — retail, financial and manufacturing, employing under 10 to over 3000 employees. The sample cannot be regarded as a strictly scientific random sample, but it is more than a volunteer or expedient selection, and in view of the quality of the replies we feel that the information deserves to be taken seriously.

Use of grades in hiring. Just over half of the respondents checked "yes" when asked whether they "consider high school grades in the choice of candidates for employment," but they offered numerous qualifications. When the "no" replies and the comments were considered, about 80 per cent placed far less emphasis on grades than on other information. Those who did emphasize grades were trust companies, banks and insurance companies. With only one exception, companies put more emphasis on attitudes than on grades in employment decisions. This emphasis presumably explains why it is so hard to get a job without prior experience. With no reference to go on, an employer cannot tell whether the applicant has good attitudes toward work.

Satisfaction with present means of evaluation. Again, respondents were split evenly between those who were satisfied with the way things are done now and those who were not. Those who were not satisfied offered many comments, not all of which were directed to evaluation methods. One company representative, for example, used the opportunity to argue for more co-op programs. A small minority argued for more basics, a more practical orientation and standardized examinations in the final year. One respondent wrote that marks "are only indicating his ability to retain. They do not reveal his ability to sustain the job demands such as: pressure, work under supervision or without it, routine, adaptability, his interests." We will return to this comment later.

Comments of a general nature. At the end of the questionnaire the respondents were invited to add anything they wished by the phrase, "Any other comments?" Many did so with remarks such as:

> The educational system should, at some point, cover the principles of the industry regarding the performance, the productivity and the permanence that is expected of the work force.
>
> We require that individuals we hire have a desire to do a good job. This attitude is harder to find. Most graduates are clueless on work ethics.
>
> I feel a great many students can have excellent marks and be absolutely no good at all in the ordinary work force. They have no common sense at all. All they know how is what they memorize from a book.

. . . with more co-op programs a better evaluation of what the student has learned would be possible.

As we have moved from trade orientation to technology, we would expect a significant shift to more emphasis on marks.

Commentary. There certainly is not a strong, general interest in high school marks as a criterion for hiring. Neither is there great dissatisfaction with the way schools do their marking. Where there was unease, it tended to be as much with schooling in general as with evaluation. Companies in which the work is closest to school tasks (the banks and insurance companies, for example) valued marks more than others.

The respondent who felt that marks only indicated an ability to retain and nothing about sustaining job demands and working under pressure has certainly not visited the high schools the author has seen. In these schools, high marks cannot be earned by simple retention, and earning them certainly does require sustained work under some pressure. If a student is exceptionally brilliant *and* reasonably co-operative, the high marks come more easily, but such students are not a problem. Only in systems where marks are determined predominantly by examinations, not very good examinations, could the respondent's perception be true. Many do not know just what *is* demanded of students who do well, and as a result the students (and the school) do not receive credit where credit is due.

Post-secondary Institutions

No systematic survey of post-secondary institutions was attempted, but several ongoing developments contributed relevant information during the period of the study. Resumption of diploma examinations in British Columbia and Alberta prompted the universities to announce their policies with regard to the use of examination results in admissions decisions, and in Ontario the Commission on the Future Development of the Universities of Ontario (the Bovey Commission) stimulated debate on access to universities and the possible effect of examinations on accessibility. The use of marks and examinations by other post-secondary institutions (colleges, institutes, etc.) was not explored.

In British Columbia and Alberta, the only question was whether the universities would use the school mark, the examination mark or the "blended" (i.e., unweighted average) mark in admissions decisions. In both provinces, the decision was to look at all information in the first year of the new policy's implementation but to use the blended mark in subsequent years, unless persuaded to the contrary. It was not clear in the public statements what use would be made, if any, of marks in courses for which there was no provincial examination.

At the time of writing, Ontario had no provincial examinations at the end of secondary school, the last such examinations having been given in the late sixties. For a time, Ontario had a program of university entrance examinations which included the Ontario Scholastic Aptitude Test (a version of the College Board's SAT from the USA) and several subject-specific tests (mathematics, physics and English among them). This program was first expanded across Canada and then ended when the universities declined to provide any funding for its continuation. They cited disappointing predictive studies and the difficulty of getting results early enough in the year for use in admissions. Contributing to their decision, no doubt, was the onset of a period when universities accepted virtually every student who applied with the minimal high school graduation qualifications.[2]

Ontario universities use the average of the best six grade 13 marks in considering graduates of Ontario high schools; they require at least a 60 per cent average and as high as 80 per cent for limited enrolment programs. Some faculties, e.g., engineering, computer science, pay special attention to mathematics and science marks. Clearly post-secondary institutions, especially universities, are the largest consumers of high school marks. As a result, universities exert a larger influence on the curriculum and evaluation methods than the proportion of university-bound students would justify. At most, 50 per cent of high school students go on to any post-secondary institution, and perhaps 15-20 per cent to universities, but universities have been in the forefront of those advocating a return to common examinations and a restricted core curriculum.

The British Royal Society convened a group in 1982 to review the teaching and examination of science (including mathematics) in secondary schools in England and Wales, to consider the needs of potential employers and to consider how to meet those needs. In their comprehensive report, they referred to "the general consensus throughout higher education that mathematics and the traditional science subjects do have structures which make it possible in each to identify topics which are so central as to be indispensable at any particular stage," noting that "this is the basis of the argument in favour of a *common core* to examination syllabuses" (emphasis in original).[3]

---

[2]The Ontario story, with reference to other provinces, is treated at length in the reports of the "Interface" studies, e.g., H H Russell, C. Wolfe, P Evans, R. Wolfe, R. Traub, and A King, *Interface: Interproject Analysis* (Toronto. OISE Educational Evaluation Centre, 1976)

[3]*Science Education 11-18 in England and Wales. The Report of a Study Group* (London. The Royal Society, November 1982), p. 23.

The study group felt that through the Society and science teacher groups it would be possible to identify this common core (though there was not agreement at that time). The members of the Royal Society recognized, however, that:

> The real problem for schools lies in a different direction. It is that while no more than 20 per cent of the school population aim at entry to higher education, the schools have to provide for the equally pressing needs of the remaining 80 per cent. The charge is then made that the influence of higher education, even if it were acceptable for the top 20 per cent, pervades the whole school system to the detriment of the majority. It is certainly the case that nobody in higher education would want this to happen and there is enough evidence to show that many schools find ways of dealing with it which are sensible and humane and generally acceptable to their members.[4]

The predicted early consensus on a core has yet to be attained. A Joint Council of the examination boards met for more than a year without reaching agreement, though it must be added that they were considering more than science. In June 1984 the government announced a reorganization that would reduce the number of examination boards from 20 to 5, so perhaps the smaller group can agree. The Chairman of the Secondary Examinations Council said, however, that reaching agreement on a common syllabus was proving more difficult than expected.[5]

In contrast to the belief in centralized control so evident in the recent initiatives by the government in England and Wales, the Science Council of Canada's comprehensive study of science education had this in one of its 47 detailed recommendations:

> The major focus for the renewal of science education should be the school itself and it is at this level that most commitment and effort is required.[6]

University dissatisfaction with school marks is based on a perception that standards vary greatly from school to school, and that there has been uneven inflation of marks. Solid evidence for such variation is difficult to find, but the University of Waterloo (Ontario) engineering faculty did say openly that they had calculated a correction factor by monitoring the students admitted to Waterloo and relating their success to their high school marks. The correction averages about 14 per cent, with a range from 0 to

---

[4] *Ibid*, p. 23

[5] Sir Wilfred Cockcroft, personal communication, June 1984

[6] Science Council of Canada, *Science for Every Student: Educating Canadians for Tomorrow's World. Report 36* (Ottawa: Supply and Services Canada, 1984), p. 51.

30. A reduction in first year failure rate from 25 to 10 per cent was due to use of the correction, it was claimed.[7]

The University of Toronto faculty of applied science and engineering also adjusts marks based on ratings of schools. The ratings are available to the principals of the schools but not to boards, teachers or the public. In Newfoundland, the Department of Education adjusts school marks if the average school mark is too far above or below the provincial mean in comparison with the average achieved by students in that school on the provincial examinations.[8]

An ironic twist to the universities' laments over varying standards and absence of a core curriculum is that nowhere are such conditions more evident than in the universities! The same social trends that pushed high schools to offer a wide range of courses and to give students choices were felt to an even stronger extent at the post-secondary level. Two sections of the same course need not cover the same content, and rarely are examinations co-ordinated (except when all students attend the same lectures in huge halls). Students choose from many courses and strongly influence the structure of their own programs. Graduate schools complain of mark inflation at the undergraduate level.

Conveniently, these sorts of observations were published (and criticized) in a book just as this report was being written.[9] The book's authors provide an eloquent statement of the conservative's solution — matriculation and college entrance examinations should be required and students should take no-choice programs of language, literature, philosophy, science, mathematics and the arts. Higher fees, an end to student participation and regular reviews of tenured faculty are included in their polemic, which ends with a call for "the public" to don badge and holster and put an end to the robbery. So much for academic freedom. We will focus here only on recommendations for high school examinations and implications for accessibility to post-secondary education.

A more reasoned, scholarly view appeared in a "Discussion Paper" funded by the Commission on the Future Development of the Universities of Ontario. Analysis of economic factors showed that few students are deterred from enrolling in university by tuition and other fees, but that expectations of job opportunities can be important (comparing expected income with foregone income). As for examinations, the report says,

---

[7]Michael Tenszen, " 'Frunch Factor' Cited in Lower Failure Rate," June 16 *Globe and Mail* 1983.

[8]For a description and analysis of the adjustment process, see Philip Nagy, "An Examination of Differences in High School Graduation Standards," *Canadian Journal of Education* 9 (No 3, 1984), pp. 276-297

[9]David Bercuson, Robert Bothwell and J L Granatstein, *The Great Brain Robbery — The Decline of Canada's Universities* (Tororto: McClelland & Stewart, 1984).

University entrance examinations could lead toward more equitable admission decisions by standardizing the basis for comparing students — provided that cultural bias could be eliminated from such tests.[10]

It hardly seems possible to eliminate cultural bias from examinations as they are currently structured. The British Department of Education and Science reported that two-thirds to three-fourths of the variation in school examination results could be accounted for by the social composition of the area from which the students were drawn.[11] This is one reason that examination scores add very little to the prediction of university success afforded by high school grades alone. A stronger reason is, of course, that success in university is due in large part to other personal qualities not measurable by today's examinations.

In short, many of the benefits expected by university registrars from common examination scores have yet to be proven. Though it seems plausible, no one has demonstrated that scores can be used to make a fairer choice among applicants than is possible with current marks, with all their flaws. There are alternatives to distorting the school curriculum to suit the universities or investing the large sums required to develop tests specifically designed for admission purposes.

> It appears at present, however, that the issue is not so much a matter of developing tests to replace or complement school grades, but to identify other criteria that can be used to overcome any bias in using meritocratic measures. Such alternative criteria would be related to motivation and other personal characteristics and would include interviews, work experience, and assessment of social-cultural background. This approach may also break through a possible circularity in relying only on grades, namely that while admission depends on grades, these in turn may depend in part on a student's expectation of admission.
>
> These alternative criteria are being pursued not only for equality or equity concerns, but also to identify vocational aptitudes relevant to certain professions and to develop more diversified membership in professions.[12]

Commentary

High school marks play an important role in university admission decisions in every province, contributing from 50 to 100 per cent of

---

[10] David Stager, *Accessibility and the Demand for University Education* (Toronto: Commission on the Future Development of the Universities of Ontario, June 1984).

[11] *Statistical Bulletin 16/83 School Standards and Spending: Statistical Analysis* (London DES Statistics Branch, December 1983).

[12] David Stager, *op cit*, p 19

the information on which the decision is based. A majority of other post-secondary institutions (Cégep, CAAT, . . .) accept the usual high school diploma as a qualifying credential, but many technical institutes and colleges have requirements as strict as any university. No survey was attempted of ways marks may be used to assign students to programs after admission, but it was said by some interviewees that marks were often used in this way.

Many of the principal consumers of school marks (especially universities) want some form of achievement measurement that is validated outside the student's classroom or school. Their hesitation to accept school marks as valid indicators of achievement does not seem to be based on informed judgement of evaluation methods as much as on lack of congruence between student marks and student performance in individual instances. This attitude is not confined to those outside the schools. High school teachers interviewed during the study reported that they hesitated to accept elementary school reports when placing students, and in several schools they had implemented tests that were open to the same criticisms teachers made of end-of-high-school examinations.

Many teachers expressed a desire for an externally validated measure of achievement, for a way to know where they (and their students) stood. When provided with such scores, however, they accepted them as valid only when the results agreed with their own assessments. The willingness of those outside the schools to accept examination scores as equal or better measures of achievement than marks derived by teachers in constant contact with students should be a matter of concern to all educators. When we portray evaluation as a craft, it does not imply that the evaluations that teachers provide are of little use. Quite the contrary; the craft of evaluation produces indices of achievement that are reasonably accurate and extremely useful. In the next section, a model is presented in an effort to see where and how evaluation of student achievement fits in the processes and outcomes of schooling.

A MODEL OF EVALUATION IN THE TEACHING-LEARNING PROCESS

SCHOOLS, especially the publicly supported heterogeneous schools open to every student, take a form that is the result of thousands of compromises. In this large undertaking, evaluation of student achievement is an integral part, and its contributions to the outcome, "student learning," cannot be separated cleanly from the rest. When any model that tries to do justice to the complexity of schooling is written down, it has so many components that the contribution of any one is likely quite small. Even *teaching* is accorded a small influence by researchers.[13] Since it is reasonable to consider much of evaluation as a part of teaching, it will come as no surprise that the effect of evaluation is small compared to the sum of effects of other variables.

Research on teaching and learning has documented over and over in recent years that the key, visible, manipulable element in enhancing learning is time — learning time, time on task. Therefore, the heuristic model that will be used to portray the place

[13] Teacher effects are likely to be small when compared with the totality of the effects of the other variables affecting student achievement," in J. A. Centra and D.A Potter, "School and Teacher Effects. An Interrelational Model," *Review of Educational Research* 50 (No. 2, 1980), p. 287

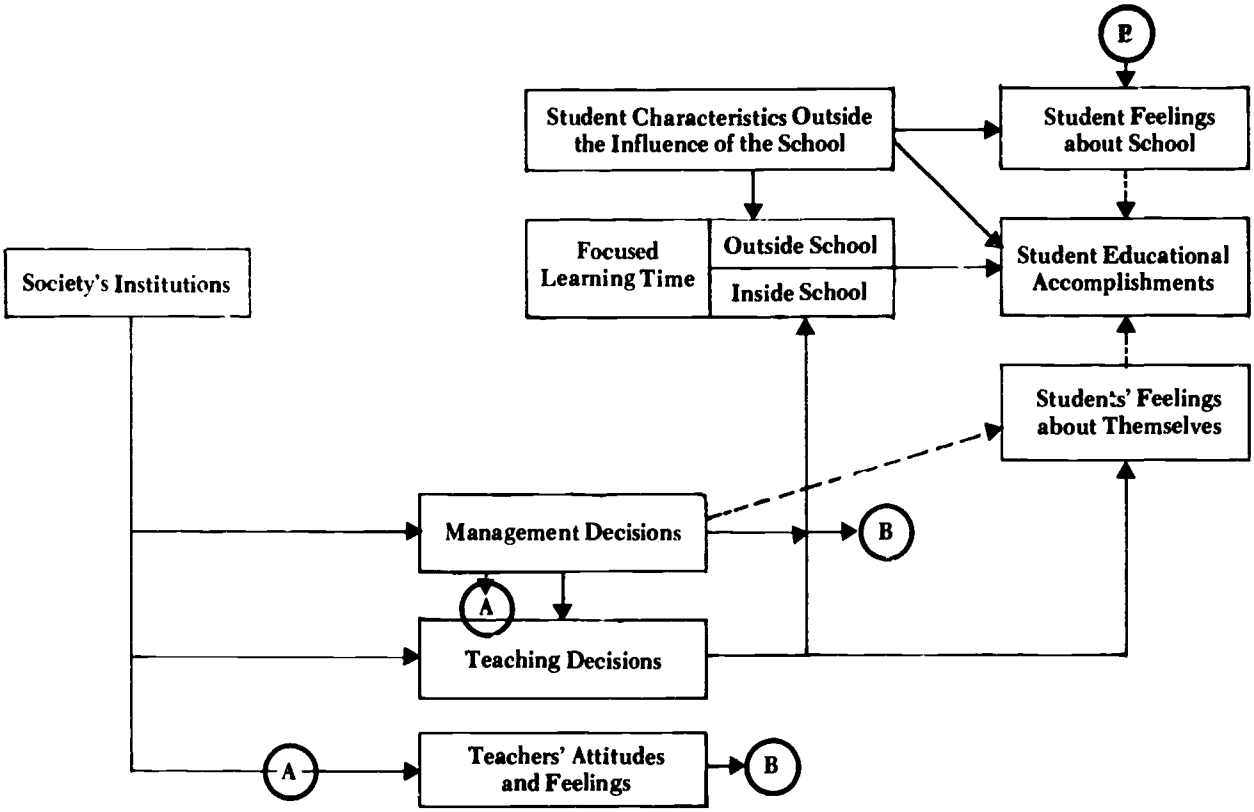

Figure 1. Simplified Heuristic Model of the Strongest Forces Influencing Student Accomplishments in Education
See Figure 2 for a more detailed model. Arrows indicate strong causal influence. Letters in circles are used where arrows would have to cross too many boxes.

of evaluation in teaching and learning is centred on time. The term "Focused Learning Time" will be used to convey that not all minutes are of equal value. Some models[14] use terms such as "Teaching Performance (behaviour)" and "Student Behaviour," but the meaning is the same: what teachers and students do, and how much time they spend at it, are the important determinants of outcomes.

The basic building blocks of the model and their links are shown in Figure 1. Evaluation is not a basic block itself but a component of at least two blocks, so a detailed version of the model is shown in Figure 2. Before discussing the details, the term "heuristic model" should be explained. By definition, a model is a representation, usually in miniature, of an object or process. Some physical models are exact representations, including all working parts. Motion in A causes B to move, then C and so on. In social science, models are far less sophisticated because we don't even know for sure what the working parts are, and we have to guess what many of the causal links are.

Social scientists have therefore had to rely on rough approximations, tentative models of complex social processes. The most likely "working parts" are put in a diagram in boxes and arrows are drawn from box A to box B to represent the present understanding (or a best guess) that change in A causes change in B. For example, there is a strong link between the social composition of a school's catchment area and the examination marks of the students. A model at such a macro level will show an arrow *from* "social composition" *to* "examination marks," since the social composition existed prior to the marks. A longer-term model might show some influence of marks on social composition by also including an arrow going the other way.

The tentative nature of knowledge represented by such models is emphasized here by the qualifier "heuristic," a term that has come to mean "for illustrative purposes only — should be close but is not to be taken literally." In Figure 2 the heuristic quality is stretched to its extreme by suggesting roughly what *proportion* of the change in B is caused by A. These proportions are based on the author's reading of the literature and are not to be confused with estimates derived from specific empirical research.[15]

As mentioned at the beginning, one of the author's tasks was

> the development of a model of the causal influences of evaluation on student accomplishment and self-concept and on teacher and student attitudes, and the relationships among such variables.

---

[14] E.g., J. A. Centra and D. A. Potter, *ibid.*

[15] For example, Maribeth Gettinger, "Achievement as a Function of Time Spent in Learning and Time Needed for Learning," *American Educational Research Journal* 21 (No. 3, 1984), pp. 617-628.

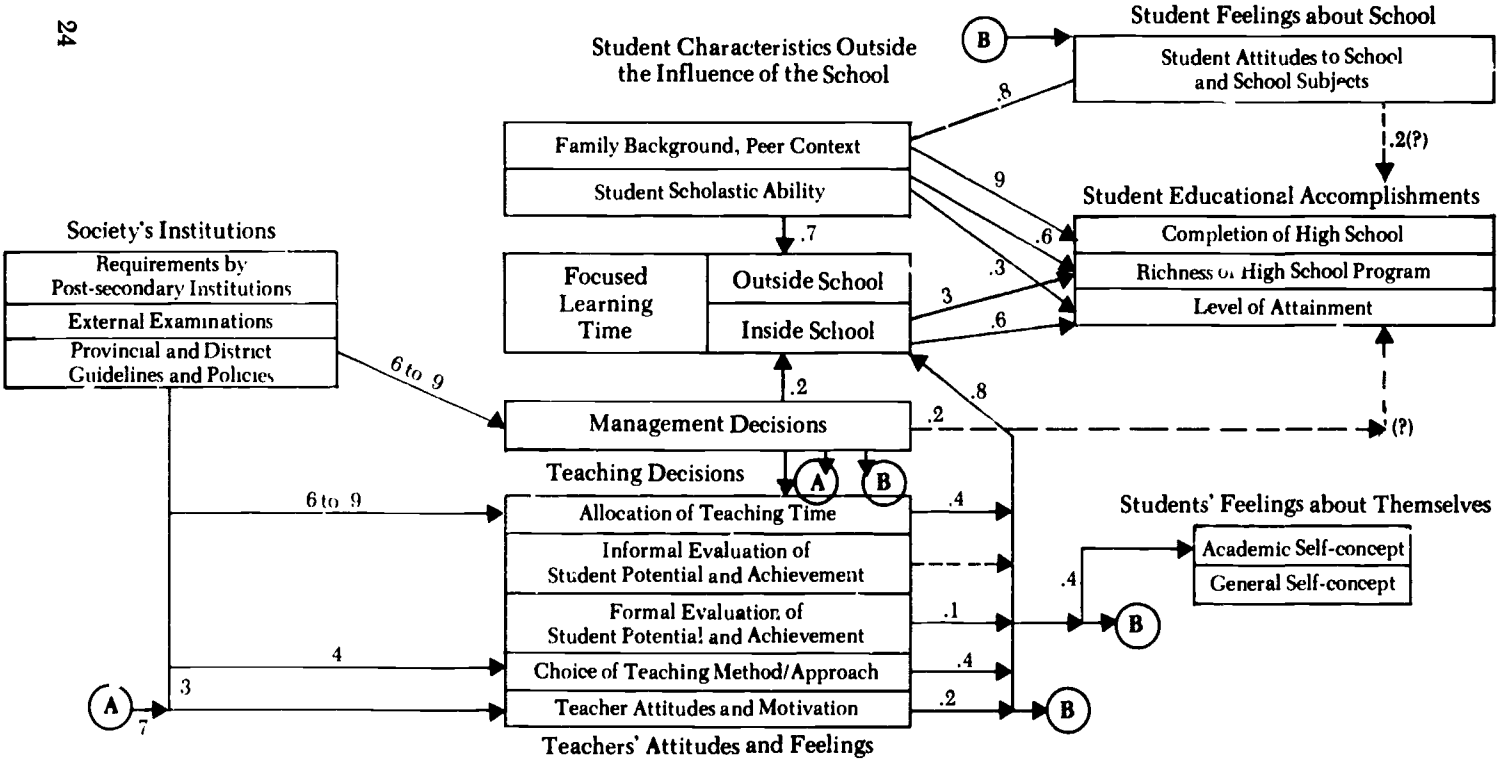

Figure 2. Elaboration of the Model of Figure 1.

The numbers represent rough estimates of relative impact of various factors on eventual student achievement.

The resulting model was so complicated that it seemed useful to show the overall structure first. This is Figure 1 The flow of influence is from left to right. A plausible argument exists for some influences in the opposite direction (feedback), but these are much weaker and have been omitted for simplicity.

For present purposes, the essential message is that society influences student educational accomplishments through schools where the key element is the amount of focused learning time the school succeeds in organizing. "Student Characteristics" is used as a very general term that includes social class. These influences, over which the schools have no control, are also important as the strongest influence on learning time outside school. They include students' peer context and that elusive and controversial characteristic, intelligence (or its close relation, scholastic ability).

Through its elected and appointed officials, society exerts a very powerful influence on the management of schools, the decisions teachers make and the way teachers feel about themselves and their work. Management in turn is a potent factor in virtually every aspect of schooling. It has sometimes been fashionable to downgrade the importance of administrators, but directors, superintendents, principals and heads of departments make decisions that, both directly and through teachers, affect learning time — every facet of schooling, in fact.

Teachers influence every facet as well, of course. In Canada, even the "external" examinations are constructed, marked and interpreted by teachers not far removed from any school. As will be seen in Figure 2, external examinations are given a place with society's institutions because the initiative has come most often from outside the school system and the examinations are functionally administered outside any district.

Separate boxes have been provided for attitudes and feelings, partly because these were singled out by the CEA Advisory Committee on Educational Research and partly because they are of a different character. Teacher and student allocations of teaching and study time can be directly observed and evaluated to some extent, but attitudes are psychological constructs that must be quite indirectly inferred. Our knowledge of such constructs is even more tentative than that of other facets of education, but as time passes we have more, not less, respect for the importance of how people feel about what they do.

The discussion of the model will be presented in two parts, Student Educational Accomplishments (with their direct and secondary causes) and Society's Institutions (which affect accomplishments only through the other factors). The objective throughout will be to put evaluation of student achievement in an appropriate perspective, neither exalting nor denigrating it. In Figure 2, the arrows are labelled with decimals (.2, .5, .9) that are meant to suggest roughly what proportion (out of 1) of the outcome

at the arrow head is due to the *cause* at the arrow's origin.

Readers will quickly see that an accuracy of estimate is suggested that goes beyond our understanding of the processes at this time. The numbers should be read for what they are — rough estimates that allow the author to distinguish among many plausible causes and put them in rough rank order. Some, the overall figure of .2 between Teacher Attitudes and Motivation and Students' Feelings about Themselves, for example, are based on considerable correlation evidence. The correlation rarely exceeds .4, suggesting that about 16 per cent of the variance in student feelings could be caused by teacher decisions. This was rounded to .2.

Student Educational Accomplishments

The first point to be made is that there are at least three important sub-facets of student accomplishment, each with its own combination of causes.

Completion of high school. A high school diploma, at whatever level of distinction, is required in so many parts of society that it deserves its own category. The decision to finish school is essentially determined by family background and peer context; evaluation plays little or no role.[16] Students often cite low evaluations when withdrawing, but this is not accepted by sociologists as the *cause* because other students (often those from middle and upper-middle class families) who receive the same evaluations either persevere as is or increase their efforts and eventually graduate.

The strong influence of family and society is indicated in Figure 2 by the single arrow and the .9, suggesting that finishing high school is about 90 per cent determined by family background and peer context. The *nature* of the courses and the diploma are separately important and get their own category — richness.

Richness. Rich high school programs are available nearly everywhere, but as enrolments fall educators are concerned that small schools cannot offer programs rich enough. The opposite of rich is bland, as in "the sauce was bland, almost tasteless." Courses should and do vary in difficulty and challenge because students vary in their capabilities to profit from courses. Science and mathematics offer different sorts of challenges from languages and literature. A rich program has some of each, at the highest level the student can possibly attain. The school's part in determining richness is recognized by the arrow from the Learning Time, Inside School box and the relative importance by the .3 (as compared with the .6 from Student Characteristics).

More and more attention is being paid to the richness of school programs as technological developments in society continue to

[16] An excellent discussion of the evidence can be found in A.H. Halsey, A.F. Heath, and J.M. Ridge, *Origins and Destinations: Family, Class and Education in Modern Britain* (Oxford: Claredon Press, 1980).

outstrip expectations, but there are more mundane concerns as well. The study of science education conducted for the Science Council of Canada came to some conclusions that suggest a need for more richness in classrooms:

> . . . Most children from kindergarten to the end of elementary school receive only a token education in science . . .
>
> . . . Some students need more challenge to reach their full potential in science educatic .
>
> . . . Research has sh wn that "textbook science" tends to be overly standardized and simplified in order to present a smooth road to scientific knowledge. But if s ience itself is a search for explanation, then surely science education must give students an authentic explanation of the way science works.[17]

Evaluation received attention in the science study and, after a recommendation that assessment techniques must be developed and implemented for *all* obje. ves [emphasis in original], the researche.s offered this general observation:

> When achievement of educational goals is not measured, those goals are not valued by students, teachers or the public; this fact has been well documented.[18]

In Figure 2, this observation is recognized by the arrow from Formal Evaluation out to the symbol (B), indicating influence on Student Attitudes (upper right-hand corner). The proportion of .1 represents a judgement on very little evidence that the influence is small relative to that of family and peers.

Discussion of richness would not be complete without mention of the rapidly growing number of computers in society and in schools. Teachers who have yet to come to terms with calculators face computers today that can carry out all the operations taught in high school algebra (factoring, expanding, simplifying, . . .) and can display in seconds the graph of much more complex functions than normally attempted in high school. As if that weren't enough, programs on university computers can do all the calculus operations! As one mathematics professor put it, "Mathematics is getting easier. We will not be able to keep this secret from our students forever."[19]

There is still plenty to teach, of course, but the content has to change. Discrete mathematics and computer science are now recommended for all high school mathematics teachers. As evidence of the importance scientists attach to these developments, the editorial in *Science*, the journal of the American Association for the Advancement of Science, was recently devoted to them. Evaluation was accorded its central place.

---

[17]Science Council of Can a, *op. cit* , p 33, 36, 37.

[18]*Ibid* , p. 43

[19]John Poland, "Computers and the Impending Revolution in Mathematics Education," *Ontario Mathematics Gazette* 23 (No. 1, 1984), pp. 26-29.

Reform of school mathematics must reflect this new mathematics. Requiring more tests is of no use if the tests examine only the old mathematics; increasing time in class is of no benefit if it only reinforces old traditions. Standardized tests must be changed, new textbooks must be written, and teachers must be provided with opportunities in substantive workshops to learn this discrete, computer-oriented mathematics.[20]

Level of attainment. Finally, we come to the sub-facet that many equate with the notion of educational accomplishment, the "How much?" that teachers try to capture with grades and marks. The model suggests that amount of focused learning time is the dominant influence on level of attainment; scholastic ability is also important. Student attitudes and academic self-concept are also linked to it by arrows. Who can believe they are not important? The arrows are dashed and their proportions are questioned because the research is inconsistent and inconclusive. The influence of .1 left for these factors may be due more to the weakness of the research than to the influence of the factors.

Note that the word "motivation" has not yet appeared, or rather has been by-passed. This model gives primacy to the amount of focused learning time the student spends. Students are led to spend time by family and ability (especially outside school) and by teachers and administrators (especially inside school). The most important influences on learning time inside school are the teacher's decisions on allocation of teaching time and choice of teaching method or approach. Evaluation does influence student learning time, but its influence is small relative to the other sources. Some influence on attainment probably results from either a positive or a negative influence on attitudes and self-concept. Separate reference to motivation does not seem especially useful.

Teachers evaluate formally, with quizzes, tests, assignments and the like, but they also make informal assessments. They size up a class at the beginning of the year by questions and perhaps with a few assignments, and these informal assessments can influence their allocation of teaching time in profound ways.[21] Some teachers who perceive that they have a weak class redouble their efforts and cover the prescribed content with special thoroughness. These are a minority. A majority of teachers who perceive they have a weak class reduce the amount of material they cover but do not appear to cover it more intensively.

Some experimental studies in the USA have been able to document a direct, positive influence of frequent curriculum-based

---

[20] *Science*, 7 September 1984, p. 981.

[21] These inferences are derived from data gathered during Ontario's participation in the Second International Mathematics Study, of which the author was Principal Investigator. The final report will be published by the Ontario Ministry of Education.

testing on school achievement. In Pittsburgh, 2500-3000 students in each of grades 2, 5 and 8 (virtually all students in those three grades in one district) took the appropriate form of the California Achievement Test (CAT) before the Monitoring Achievement in Pittsburgh (MAP) program was started. After the program had been operating for two years, the testing was repeated at the same grade levels (different students, of course). The content of the CAT and the content of the MAP testing program were compared, revealing some areas of high overlap and some of low overlap.

Important gains were observed in the areas of highest overlap, and one conclusion was: "There can be no question but that the monitoring program is a powerful tool in enhancing the achievement of students." Also, however, "there is legitimate cause for concern when considering the long-term effects of this phenomenon." The concern arose because there was evidence that the program resulted in instruction being made routine and the content domain being narrowed to that of the tests.[22]

More positive results were reported from a study involving 18 experimental and 21 comparison teachers in special education classes in New York City. Half of the 64 students were emotionally handicapped, a third were brain-damaged and the rest were in "resource programs." The data-based program modification system (DBMP) was implemented over a school year. Teachers wrote detailed objectives and monitored progress at least twice a week. Students with teachers who employed the DBMP recorded higher achievement and showed greater involvement in and awareness of their own learning.[23]

Society's Institutions

In every province and territory, the government has legal responsibility for the provision of schooling. An Education Act charges the Minister of Education with this responsibility and gives the Minister sweeping authority. Legally, everything is crystal clear.

In practice, Ministers have increasingly delegated this authority to local boards of trustees who employ staff and in turn delegate the responsibility to them. Only 20 years ago every province employed inspectors who visited schools, certified teachers and approved curricula. High school diplomas were granted by the province entirely on the basis of provincial examinations, and high school entrance examinations had only recently been discontinued. Today, although approval by the Minister may be required as a formality,

---

[22]Paul G. LeMahieu, "The Effects on Achievement and Instructional Content of a Program of Student Monitoring through Frequent Testing," *Educational Evaluation and Policy Analysis* 6 (No 2, 1984), pp. 175-187.

[23]Lynn S. Fuchs, Stanley L. Deno, and Phyllis K Mirkin, "The Effects of Frequent Curriculum-based Measurement and Evaluation on Pedagogy, Student Achievement, and Student Awareness of Learning," *American Educational Research Journal* 21 (No. 2, 1984), pp 449-460

local districts in six provinces grant their own diplomas. In Newfoundland, Quebec, Alberta and British Columbia, the secondary school graduation diploma is granted by the province and almost all students have to take at least one provincial examination. Where examination marks exist, the final mark is an equal blending of the school and examination mark.

It is difficult to describe such a heterogeneous system accurately in a few words, but the pattern is certainly clear — a very rapid decentralization of control has taken place since 1960 and the recent reinstatement of a few examinations in Alberta and British Columbia has reversed that trend only slightly. Quebec, the most examining province, has reduced the number of different examinations administered per year from over 400 to just over 100.

Examinations and control. The question of control is relevant in a report on student evaluation because it appears that elected officials everywhere see common examinations as the only effective means left to them to exercise some control over public education. This is evident in the USA, the UK and France (at least) as well as in Canada. State legislators in the USA, especially in the south, opted for "minimal competency" examinations when it seemed some students were graduating without the most basic literacy and numeracy skills, *and when they apparently could not find any other way to change the situation*. Several states are requiring teachers to take competency tests.

In England and Wales, the Thatcher government has had running battles with local authorities over proposed education reforms and it finally decided that the examinations system was the most practical route to exercise more control over the system. First, the Schools Council was disbanded and replaced by two bodies, the very powerful Secondary Examinations Council (members are appointed by the government) and a Curriculum Council. In June 1984, it was announced that the number of examination boards would be reduced to five and the Certificate of Secondary Education and O level examinations would be combined. The advanced certificate examinations (the A levels) will not be changed.

Having gained effective control over the examination syllabuses, the government could exercise more effective control over the secondary schools, since all but the lowest 40 per cent of students take at least one examination. The examination syllabus is the effective curriculum guideline for secondary schools. The Minister would extend this influence throughout the system. Here's how two British authors reported it in a postscript to their book on secondary school examinations:

> On 6 January 1984, Sir Keith Joseph, Secretary of State for Education and Science, proposed a number of reforms to raise school standards and pupils' achievements in a speech at the North of England Education Conference in Sheffield. Much of

the speech was concerned with the quality of teaching and the need for an agreed curriculum from 5 to 16. The principal thrust, however, was directed towards examinations.[24]

Paths of influence. Provinces typically legislate (or issue regulations for) precise minimal times for the school year and for high school courses. In Figure 2, the influence on Management Decisions and Allocation of Teaching Time is listed as .6 to .9 because surveys usually show considerable variation in the number of teaching days per year and even more variation in the number of hours per course.

Choice of Teaching Method/Approach is shown as about 40 per cent influenced by society's institutions because of provincial and district influence on selection of textbooks. Several studies, including the Second International Mathematics Study mentioned earlier, have found that much teaching is done straight from the textbook. These influences are mentioned because visits to districts and schools during the CEA study revealed that provincial departments and district administrations make little effort to influence student evaluation in the schools and classrooms. Teachers are left very much to their own devices to evaluate student achievement.

One possible exception to the previous observation is the influence provincial diploma examinations have on teachers' evaluation practices. The diploma examinations had only just been resumed when the author visited Alberta and British Columbia, but already teachers in courses with exams were talking about how they had prepared their classes for the examinations and how they would in the future. They consistently estimated that two weeks were taken for review rather than instruction and that they would in future use questions in their own tests like the questions on the provincial examinations.

External examinations, where they exist, are part of the .6 to .9 influence on the Allocation of Teaching Time. Provincial curriculum guidelines are commonly quite general and leave much to local initiative. Test blueprints, on the other hand, are necessarily much more specific and teachers often said that the test blueprint had become the operational guideline. External tests have both advantages and disadvantages.[25]

---

[24]Jo Mortimore and Peter Mortimore, *Secondary School Examinations*, Bedford Way Papers No 18 (London: University of London Institute of Education, 1984), p. 76.

[25]One principal pointed out that since test blueprints were more faithful to the guideline than textbooks, teachers who taught straight from the textbook would be putting their classes at a disadvantage. Jo and Peter Mortimore discuss advantages and disadvantages and conclude that the latter outweigh the former in England.

CONCEPTIONS OF TEACHING — AND STUDENT EVALUATION

IN AN EFFORT to understand teaching, various writers have found it useful to compare teachers to craftspersons, professionals, bureaucrats, managers, labourers and artists.[26] Almost 30 years ago, Broudy argued that teaching was more like a craft than a profession,[27] and in his often-cited book, Lortie wrote, "In thinking about teachers it is useful to conceive of members of the occupation as engaged in a craft; we can then compare conditions affecting the practice of this craft with those in other crafts."[28] For Lortie, "a craft is work in which experience improves performance — the job cannot, like many unskilled or semi-skilled types of work, be fully learned in weeks or even months."

Teaching as *craft* requires a repertoire of specialized techniques as well as generalized rules for their application. Anyone who has seen the work of someone skilled at the craft of pottery or woodworking will appreciate that to portray teaching (or evaluation) as craft is not to devalue teaching.

---

[26]Linda Darling-Hammond, Arthur E. Wise, and Sara R. Pease, "Teacher Evaluation in the Organizational Context A Review of the Literature," *Review of Educational Research* 53 (No. 3, 1983), pp. 285-328.

[27]H S Broudy, "Teaching — Craft or Profession?" *The Educational Forum*, January 1956, pp 175-184

[28]Dan C Lortie, *Schoolteacher — A Sociological Study* (Chicago: University of Chicago Press, 1975), p 135

The distinction between craft and profession is not always sharp, but it is a good point for discussion. An essential difference is that the professional is expected to master a body of theoretical knowledge as well as a range of techniques and to make independent judgements about when the techniques should be applied. Under this conception, teachers are clearly expected to become more and more professional as they gain experience and pursue further education, but as two reviewers of Lortie's book noted, "Teachers are neither required to be conversant with the theoretical constructs which seek to explain the teaching and learning processes nor are they expected to contribute to the development of the craft."[29] Before turning to student evaluation, we shall consider the metaphors of art and science.

Teaching as *art* may be novel, unconventional or unpredictable. Specialized techniques are used, but the rules for their application are loose guidelines and a premium is placed on individual expression and creativity. According to Gage, teaching *uses* science but cannot be a science because the teaching environment is not predictable.[30] Those who would have teaching become more scientific devise ways to reduce variability and unpredictability.

Using these conceptions of teaching, the evaluation of student achievement is most like a craft. Teachers receive little formal training in evaluation (sometimes none at all). They are seldom presented with systematic theoretical knowledge about evaluation, much less expected to master it. They make judgements about when evaluation techniques are to be applied, but the range of options is quite restricted. As we shall see below, the prevailing view is that evaluation must be predictable, and variability is usually considered undesirable.

Teachers learn about evaluation as potters learn about working with clay — from other skilled practitioners. Most know little about the underlying theories — why one technique works in a given situation and another does not. According to their accounts, they learn by experience with little or no supervision, and in-service training opportunities are becoming fewer and fewer every year. Unlike teaching in general, evaluation could often be scientific, but as it is generally practised, we are far from a science of evaluation.

Most teachers become skilled at evaluation — some are less skilled and some are professionals (in the sense defined above), but most apply a very few specialized techniques according to general rules that are rarely stated explicitly. They are made uncomfortable when they have to explain how they do it, and more uncomfortable still when asked how they justify doing it that way.

---

29 K George Pedersen, and Thomas Fleming, *Canadian Journal of Education* 4 (No 4, 1979), pp. 103-110

30N L Gage, *The Scientific Basis of the Art of Teaching* (New York: Teachers College Press, 1978), p 15

One reason for this state of affairs is that the assessment of student learning in the classroom has a weak scholarly (theoretical) base. There are good theories and techniques for differentiating among individuals on wide, abstract variables (*science, mathematics, vocabulary, intelligence*, and the like) but no consensus on theory or techniques for defining and measuring the large range of achievement linked to teaching in classrooms. There do exist concepts and techniques that would improve most teachers' practices,[31] but classroom evaluation practice is not so far from professional practice as, for example, amateur athletics or music is from professional performance.

Such a conception of teaching and of student evaluation is certainly consistent with the observed preference of teachers for experience-based professional development. They are seldom prepared, by training or experience, to learn from general examples, and still less by deductions from theory. Those who design effective in-service training experiences know this and build their presentations around experienced and admired practitioners. The improvement of the practice of student evaluation will likely have to proceed in the same way. In the next section, we will examine the criteria teachers and officials now have for high quality evaluation programs.

---

[31] See, e g , Mark Holmes, *What Every Teacher and Parent Should Know About Student Evaluation*, Informal Series/46 (Toronto: OISE Press, 1982). See also a critical review by Traub in OISE's *Field Development Newsletter*, September 1983, and Holmes' reply in the same issue.

CONTEMPORARY QUALITY CRITERIA AND FUTURE TRENDS

IN CONTRAST to the tangible crafts of pottery or woodworking, teaching is intangible. Potters know very soon, in a day or two at the most, whether their work has been well done, but the results of teaching are usually remote, sometimes only known many years in the future. Moreover, there is never a single criterion of excellence. The daunting challenge that results has been described as:

> The teacher's craft, then, is marked by the absence of concrete models for emulation, unclear lines of influence, multiple and controversial criteria, ambiguity about assessment timing, and instability in the product.[32]

This same impression was formed during the interviews, where officials and teachers were asked to list their criteria for an excellent evaluation program. Answers did not come readily; this was not a familiar question. When they did come, they were grounded in experience rather than in theory.

A pattern did emerge, however, from the lists of quality criteria, and this will be reported first. Attention will then be turned to some future trends that are becoming visible and a commentary will be offered on the state of the art.

---

[32]Don Lortie, *op. cit.*, p. 136.

School Criteria for Excellence in Evaluation

There was consensus on a few criteria and diverse opinion on others. Everywhere there was a general concern for fairness and equality. In Alberta, the principal and the superintendent have to certify to the department that their evaluation methods are fair and just. In Quebec, provincial policy also stresses this point:

> Le Ministère croit utile d'identifier les valeurs que l'évaluation pédagogique doit respecter. Puisque l'évaluation fait partie intégrante de l'apprentissage, on peut dire que ce sont les mêmes valeurs qui président à l'une et à l'autre. Il nous semble toutefois que la justice et l'égalité se trouvent, de façon particulière, à la base de l'évaluation pédagogique.[33]

Principals and teachers were more down to earth, but the criteria most often mentioned could be summarized under this same general banner.

Fairness and equality. To achieve fairness, many stressed the importance of communicating the school's expectations to students and parents, preferably at the beginning of the year. A principal summed up the objective as "no surprises." The most frequent complaint (from parents and students to principals and teachers) was that they had not known what to expect. Under this same heading the importance of setting reasonable standards was mentioned, although this was acknowledged as difficult in practice.

The only explicit mention of equality was in the matter of consistency among teachers in the same school teaching the same grade or subject. Equal treatment of students was not raised as a criterion, perhaps being taken for granted (or perhaps being a touchy subject).

Overlap with the curriculum. The second most frequent criterion volunteered by teachers and officials was some version of validity — that evaluation must be linked to the teaching objectives. This is, of course, part of fairness. After the course objectives are communicated to students and parents, the evaluations must reflect them. Setting the objectives and devising an evaluation are functionally separated in schools, however, so it is not surprising that each received explicit mention.

Standards — criterion- or norm-referenced? Here is where consensus ended. Some argued strongly for the establishment of criteria and awarding of marks accordingly — no restriction on the number of As and Bs (or Ds and Fs). Others felt just as strongly that

---

[33]Ministère de l'Education du Québec, *Politique générale d'évaluation pédagogique, section préscolaire, primaire et secondaire* (Québec: Ministère de l'Education, 1981), p. 4

no criteria were valid without reference to what students had learned (or not learned) in prior years. Since the students' performance was the only viable source of information about this, the marks distribution had to take the group norms into account. One sensed that in practice there was a blending of the two, either consciously or unconsciously. One official said that norms, preferably provincial or national norms, were needed for political reasons.

One place where norm-referencing would be expected and indeed was found was in the reporting of results on commercial standardized tests. These tests, for example, the Canadian Tests of Basic Skills (CTBS), were very commonly administered by the district; scoring and reporting services were purchased from the test publisher. The Newfoundland Department of Education arranges (and pays for) the CTBS to be given to all students in grades 4, 6 and 8 in successive years (i.e., one grade per year). Testing was started in high school in 1982. The test publisher produces reports at the class, school and district level and these are given to the schools.

Careful, persistent questioning of officials and teachers at every opportunity revealed very few uses of this test information, of any kind. According to all informants, the results are never used directly in calculating students' marks, and in a strong majority of schools are never carefully studied. Teachers do not regard them as relevant to their curriculum, and research would support them in that perception.[34] A CEA study published in 1978[35] reported slightly different conclusions from a mail survey of chief executive officers of districts (superintendents, directors). In that survey, 5-10 per cent of the CEOs said that standardized tests were used "to assist in determining final grades." These replies came more often from non-urban than from urban districts, so the present study (with only a few contacts in non-urban districts) may not have turned up these uses. One use of test information that was discovered among the hundreds of interviews in six provinces suggested what was required to make the standardized tests useful to teachers. A district program co-ordinator became curious about low mathematics scores from two elementary schools and pulled out the detailed reports from those schools. After several hours of study, the co-ordinator was able to see that the low scores were due almost entirely to a number of questions using S.I. (metric) units of measure. In those two schools the curriculum was in the process of being converted to metric units,

---

[34]See, e.g., M.W. Wahlstrom, R.R. Danley, and D. Raphael, *Measuring Achievement at the Primary and Junior Levels* (Toronto: Ontario Ministry of Education, 1977). See also the companion volumes for intermediate and high school divisions.

[35]Verner R. Nyberg, and Brigitte Lee, *Evaluating Academic Achievement in the Last Three Years of Secondary School in Canada* (Toronto: Canadian Education Association, 1978).

and students had not been taught to use those units before the standardized test was given.

The above example is informative in several ways. First, one sees why teachers are reluctant to give the test scores (or the norms derived from them) very serious consideration. The scores can be affected to an important extent by several items not in the curriculum or not taught in time for the test or just not taught. To discover such potentially useful information from test scores, however, requires learning how to read the computer printouts and then spending considerable time comparing results with the test booklets themselves. Teachers argue persuasively that they can make better use of their time analyzing their own tests and homework exercises. Officials such as the program co-ordinator rarely have the experience, inclination and time to make such detailed analyses.

It was said several times that teachers use test scores to check their perceptions when they think they detect students who are having problems and use scores in grouping for instruction. This was not verified, nor could a careful survey be done. Such evidence as was obtained, however, indicates that these uses cannot be very common. It is reasonable that the most accessible goals are those that only require the test score (or a derivation from it, such as grade equivalent). These uses also account for the preference teachers and officials have for norms. The majority of uses made of test scores can be classified under the heading "finding out where we stand," not always the easiest thing to do in an uncertain craft.

Continuous and comprehensive. Viewed as an opinion poll, the study found a slight edge in favour of continous evaluation, but there were teachers who regarded end-of-unit or end-of-term marks as the only valid indicators.

Provincial policy in Quebec specifies that both continuous (formative) and comprehensive (summative) evaluation are important. In other provinces, the overall policy rarely stated a preference, leaving this to the local jurisdictions. The issue was most frequently settled at the school level (in departments in high schools), where it is often decided that a term mark must be based on at least *n* pieces of information, with *n* usually three or more.

Miscellaneous. One teacher mentioned explicitly that evaluation decisions should be arrived at democratically, giving the impression that this was not always so. Group decision-making would be the exception, since in the great majority of schools teachers rarely discuss evaluation. If one were considering terms such as *democracy* to describe decisions about evalaution, the term *anarchy* would be more accurate.[36]

---

[36]In the spirt of Proudhon, "As man seeks justice in equality, so society seeks order in anarchy."

Another teacher mentioned the need for a variety of techniques, and we will return to this point. In view of the small amount of instruction pre-service teachers receive and the virtual absence of discussion or in-service training, it is not surprising that a few techniques dominate practice.

Commentary on the Criteria

Throughout the study, it proved difficult to get details. Evaluation is not so much a deliberate activity as a familiar skill. Teachers do not constantly apply quality criteria any more than the potter thinks about the wetness of the clay, the necessity to get the lump centred on the wheel or the pressure required to shape the pot. The difference is that the consequences of wet clay, an off-centre lump or wrong pressure are immediately obvious. Errors in evaluation appear later, if at all.

Technique and technology of evaluation. One would not have expected teachers to be preoccupied with technical measurement concerns. Item discriminations, internal consistency of tests and the standard error of measurement are not on the tips of their tongues. It was sobering to this researcher, however, to find these indices entirely absent — never mentioned and not recognized except very vaguely.[37] The elemental stuff of the measurement courses and textbooks is as foreign to classroom teachers as spectral analysis of glaze mixtures is to potters -- and apparently as irrelevant.

One element of technique many teachers have learned is the multiple-choice question, where students choose from four or five possibilities supplied by the teacher. The select sample of people interviewed in this study showed themselves to be aware of the limitations as well as the advantages of such questions. Many schools have informal guidelines that limit the proportion of multiple-choice questions that can be used in term or final tests.[38] There was a general understanding that attainment of many higher-order objectives cannot be evaluated well or at all with such questions. Extensive use of them by untrained teachers who do not have access to the tools of item analysis, however, may well result in poor measurement.

There was considerable enthusiasm everywhere for the construction of "banks" of high-quality questions linked to the curriculum, and several provinces and districts have made a start.

---

[37]There are measurement specialists in district offices, of course, who are knowledgeable professionals. The technical terms are on the tips of their tongues and the techniques at the tips of their fingers, but these people are a very small minority. They seldom have time for in-service work on evaluation with teachers.

[38]The proportion reaches 100 per cent, however, when schools schedule examinations so that final marks must be submitted a day or two after testing. Such a practice narrows the range of available evaluation methods more than most teachers believe to be desirable.

More often, the questions are already packaged into tests, which makes it difficult or impossible for teachers to tailor measure them to their curriculum. Tests can suggest areas of difficulty but give few hints on what to do to make the situation better; they are best left for the summative evaluation tasks to which they are suited.

Such technical pluralism gives teachers great f..edom but offers no security. It is difficult to defend a continuous, respo' ve marking scheme tailored to the local curriculum against critics, many of whom passed provincial examinations set in an era of greater consensus. Everywhere, teachers and officials are searching for ways to demonstrate that their flexible programs are working. By default, common examinations may appear to be the only method available. The teacher's situation was described sympathetically this way.

> The freedom to assess one's own work is no occasion for joy; the conscience remains unsatisfied as ambiguity, uncertainty, and little apparent change impede the flow of reassurance. Teaching demands, it seems, the capacity to work for protracted periods without sure knowledge that one is having any positive effect on students. Some find it difficult to maintain their self-esteem.[39]

Official concern about quality. In only one province was a very critical official view encountered. One discussion paper (that attracted considerable criticism in return) said:

> Student evaluation appears to be a weak link in the instructional process. It tends to deteriorate with each higher grade in the system. The single test to produce infallible grades for report cards is common practice in many classrooms. In others, each project, essay, lab, or assignment may be graded and the resulting scores aggregated to make the term mark for the subject. The policy of "counting" everything a student does differs little from the single test in practice.
>
> . . . .
>
> Despite the rhetoric concerning the asking of higher level cognitive questions, most examinations which have been reviewed reveal a premium on recall. Furthermore, it is only rarely that teachers demonstrate to their students how to deal with questions which require application, analysis and synthesis — probably because most teachers have not been taught how to provide such demonstration.
>
> . . . .
>
> In summary, there is a general recognition of the need to monitor student achievement in the province, but in ways that

[39]Dan Lortie, *op cit.*, p. 144.

are defensible, equitable and just. There is evidence that although there are exceptions, testing practices and test construction are basically inadequate.

. . . .

The need to impose some structure on the system seems obvious.[40]

In the same discussion paper, five components were proposed for the aforementioned structure:

- Clearer provincial policies,
- a central registry of marks (as a means of monitoring grade inflation and student achievement generally),
- greater emphasis on evaluation in the new curriculum guidelines,
- a common high school diploma, and
- compulsory examinations in mathematics (penultimate year) and English (in the final year).

The criticisms are much more explicit than in other provinces, but the remedies are very familiar. Only the proposals for examinations and the common diploma are sure to be acted on, and, since there was more resistance to the common high school diploma than to the examinations, the examinations will likely be implemented.

Defences offered by teachers and officials. Confident officials in smooth-running schools defend decentralization. One high school principal who had already affirmed the importance of communicating objectives to students at the beginning of the year and who reported excellent participation on parents' nights volunteered that changes in the system have a stimulating effect on students. "Greater decentralization leads to greater creativity," he said. In all his years of teaching he has met few bad teachers. In his opinion, tight control doesn't help anyone.

An elementary school principal cited his district's policy as ideal. It boiled down to the following two requirements — The principal must assure that:

- the results of student evaluation are transmitted to parents three times a year,
- contact is maintained with parents and parents informed if any problems arise.

The chief district official had written an interpretation of the policy that left no doubt where responsibility must rest:

> The relationship among the goals, objectives, standards, and planning is one that can only be fully realized in the classroom. Although outside influences from the province, the public, the school board, and the school will help establish standards, set

[40]Discussion paper from Atlantic Region, January 1984.

goals, and provide patterns of planning, it is teachers and students who need to find harmonious and appropriate ways of putting them into practice.

Evidence and opinion. The province delegates responsibility to the trustees who delegate responsibility to the officials who delegate responsibility to the teachers. Teachers are accountable to the students in their classes and to the parents of those students, as well as to the officials. When everything works smoothly (meaning there are few complaints from parents and the public), the teacher is usually left alone. When there are complaints, however, such a system has few defences. Good and poor schools alike are vulnerable to charges that students learn nothing, charges usually supported by one or two examples of student work. In the absence of evidence that can be communicated to the public, opinion polls come into play. Both the Alberta and British Columbia Ministers cited polls in support of their decision to reinstate diploma examinations. One reason Ontario has been slower to reach for the examination button may be that a poll showed little or no enthusiasm for provincial examinations.[41]

The critical provincial discussion paper cited above ends with this sentence: "In the province, the examinations should serve the purpose of acknowledging the importance of academic rigour and standards in the public school system and, hence, of restoring confidence in education generally." Someone had concluded that confidence had been lost.

A recent national poll[42] asked, "How much confidence do you, yourself, have in the following institutions to serve the public's needs?" As regards the public schools, 75 per cent overall had either a "great deal" of confidence (25 per cent) or a "fair amount" (50 per cent), and the percentage in the Atlantic region was nearly 90 per cent! It would not appear that elaborate and expensive means were needed for the job of "restoring confidence in education generally" in the Atlantic region. In the prairies, 78 per cent had a great deal or a fair amount of confidence, and in British Columbia the figure was 73 per cent.

Some hint to sources of uneasiness was provided by the question, "In general, how would you compare elementary and secondary schools of today to schools of your day, whether in Canada or elsewhere? Standards have. . .". In Quebec, Ontario and British Columbia, 40 per cent or more felt that standards had worsened, but in the Atlantic region only 25 per cent thought so, and only 30 per cent felt that way in the prairies. One qualification was that the higher the status of a respondent's occupation, the higher the

[41] D.W. Livingstone, D.J. Hart, and L.D. McLean, *Public Attitudes towards Education in Ontario 1982. Fourth OISE Survey* (Toronto: OISE, 1983).

[42] *Speaking Out — The 1984 CEA Poll of Canadian Opinion on Education* (Toronto: CEA, 1984).

percentage who felt standards had worsened. In Canada, few longitudinal studies have been done that could help decide the issue, but the "Interface" studies mentioned earlier found no evidence for lower standards in Ontario and a test given again at the time of the Second International Mathematics Study found remarkably stable end-of-high-school performance over a 15-year period.[43] *Remarkable stability* would not be a bad description of student evaluation either. It hasn't changed much in the past two decades. Before we come to recommendations, a brief look at future trends would seem to be in order.

The Road Ahead

The present climate of uncertainty will yield more traditional examinations in the near future. Most of these will be confined to the end of secondary school, however, leaving room for new initiatives at both elementary and secondary levels. There are already some experiments worthy of attention.

Creative use of computer technology. We have seen isolated examples of direct instruction and testing by computer, what is usually called computer-assisted instruction (CAI), but the costs and difficulty of creating good lessonware have prevented this innovation from spreading. Powerful microcomputers and microcomputer clusters are now appearing in schools, however, with enough storage capacity for the development of local item banks. By local, the school and even the high school department is meant. Chemistry teachers now have access to the Chemistry OAIP on microcomputer cassette, and an enterprising high school teacher is marketing his own program for creating tests from any collection. There are powerful data base management programs on the market (albeit fairly expensive as yet) that can take this effort out of the cottage industry class.

The Calgary Board of Education is off to a fast start with a mathematics item pool based on a minicomputer in the board office. Though still in the experimental stage (three schools in early 1984), the system gives teachers a look at what is possible and creates a nucleus of teachers with hands-on experience in using a computer-based system. Such experiments, valuable in their own right, set up a base for a more rapid spread of new technology when other systems become available.

Another development now being widely discussed in England will get a strong boost from technological developments — what they

---

[43] M W. Wahlstrom. D. Raphael, & L.D. McLean, *Comparative Analysis of Ontario Mathematics Achievement 1968-1982: Results from the Second International Mathematics Study*. paper presented at the annual meeting of the Ontario Educational Research Council, December 1983, Toronto. (Available from OERC, 1260 Bay Street, Toronto)

call there "graded tests."[44] "They are assessments based, not on set proportions of candidates gaining particular grades, but on the achievement of specific levels of skill, regardless of age." These are, of course, familiar masterly learning tools brought forward to the status of a general assessment system. The Secretary of State for Education, Sir Keith Joseph, is particularly keen on the idea, especially as a way of providing monitoring for the 40 per cent of students who do not now take any public examinations. The new technology is important because item banks can be expanded to include other than multiple-choice questions and the flexible production, marking and recording of results from good quality tests becomes a practical possibility.

Profiles and their variants. Strong and vocal opposition to graded tests came from English teachers who felt that the development of mainstream language competence (as opposed to "foreign" or second-language competence) did not lend itself to sequential, piecemeal assessment. A group of teachers who perceived that they were badly misunderstood got together and produced a delightful book, *English in Schools — What Teachers REALLY Try to Do.*[45] They advocate the accumulation of a comprehensive record of student attainment, often called a "writing folder." As its name implies, a writing folder is designed to preserve examples of students' written work, but the concept of such a "writing folder" is the same as that of profiles — students choose and judge some of the work and teachers decide on other entries. The essential point is that students build and carry with them a meaningful, concrete, directly interpretable record of their achievements.

Profiles are extensions of this concept to include formal tests, assignments and, very important, students' own personal record of achievement. Proponents of profiles see positive contributions to student self-esteem and their desire to work and succeed at school tasks. Some offer profiles as an alternative to public examinations.[46] Extensive work was carried out at the Scottish Council for Research in Education, where only teacher-controlled and teacher-assessed profiles were employed. The study found very favourable reactions to the scheme but noted difficulties with the complexity of the assessment pattern, the high cost of materials and the need for considerable in-service training.[47] New technology promises to help at least with the complexity.

---

[44] A fuller discussion can be found in the Mortimores' book, *op. cit.*, pp. 64-68.

[45] Available from the English Department, Institute of Education, Bedford Way, London W C 1

[46] P. Broadfoot, "Alternatives to Public Examinations," *Educational Analysis* 4 (No. 2, 1982), pp 33-45

[47] Scottish Council for Research in Education, *Pupils in Profile* (Edinburgh: Hodder & Stoughton, 1977)

Orientation. A recent French educational reform has replaced a series of examinations (*not* the celebrated and feared *Baccalauréat*) with an elaboration of the profile, writing folder system.[48] Teachers keep a cumulative *dossier* that is reviewed at regular meetings among the teachers, a guidance counsellor, school doctor and psychologist, and representatives of the parents. Two major meetings are held at the end of the second and last years at *collège* when future educational directions are being decided. The decision on type of secondary school, a very important one in France, is made by the Guidance Council, but parents may appeal if they do not like the decision. Only if the impasse is not resolved is an examination set, marked by a committee external to the *collège*.

Schools in Canada long since lost the staff resources to implement such a system, if they ever had them, but the alternative is worth noting as an example of how far educational systems can change when they decide to. Teachers in Canadian schools perform a version of *orientation* at the end of each year in "promotion" meetings, but the involvement of parents and official outsiders in France is unique.[49]

Theory of assessment and testing. Just about the only scholarly support for educational assessment has been classical and modern test theory, now dormant for a decade. The theory helps hardly at all with profiles and large-scale monitoring. There is clearly a need for scholars to work with teachers and officials to provide a better understanding of the process of evaluation of student achievement — to develop better indices of quality, for example.

These comments about test theory will likely be disputed by scholars who feel that item response theory (sometimes called latent trait theory) is the development needed to modernize assessment and testing. (The present author is on record in opposition to this view.[50]) Reasonable people disagree and the issue is certainly not settled.

A group at the University of London Institute of Education is proposing a research program to work toward a new theory of assessment [illegible]ng, and everyone, including proponents of item response th[illegible]ikely to applaud and join in such an effort. The monitoring [illegible]evaluation of student achievement deserves the best intellectual effort to go along with the common sense and hard work now carrying it forward. It is a long way, however, from a craft to a science.

---

[48] *Rentrée scolaire. 1977 ce qui change* (Paris Service d'Information et de Diffusion, 1976)

[49] One critic thinks the system is a cosmetic reform with as many or more drawbacks as the previous one See Broadfoot, *op cit.*

[50] L McLean & R Ragsdale, "The Rasch Model for Achievement Tests — Inappropriate Before, Inappropriate Today, Inappropriate Tomorrow," *Canadian Journal of Education* 8 (No. 1, 1983), pp. 71-76.

Remembering the context. Before we leave future trends and move on to recommendations, it would be prudent to look around us at what is happening in society in general. The possible contributions of new computer technology to evaluation have been noted, but the impact of computers on the types of jobs we do and the way we do them has been mentioned only indirectly. Reference was made to microcomputers that will do all the algebraic factoring and equation-solving that now take up most of the mathematics curriculum, and it should not be too difficult to imagine how these sorts of machines are changing engineering and other technical occupations. Before most people have tried a word processor for themselves, the technology has moved on to document preparation systems, one of which was used to prepare this report. On completion of each draft, yet another computer program was used to check for spelling errors, and it was somewhat comforting to find out that the computer program did not find all of them. The program has to be taught how to spell a number of words, just as its user does. The difference is that the computer program doesn't forget.

Can we really go on setting traditional examinations over yesterday's basics when today's newest technology is already old hat? To paraphrase the mathematics professor quoted earlier, "We won't be able to keep these secrets from the parents forever." One day soon they are going to change their most common tune (more emphasis on "the basics") and angrily demand to know why the schools keep spending so much time on routine tasks that computers can do and too little time setting real problems and teaching students to use computers to solve them. It will take the wisdom of Solomon to move fast enough but not too fast. This report has stressed examinations as instruments of control. Their use can retard the schools' responses to technological change or strengthen and improve them. Before I list some recommendations, a summary of the main findings will be presented.

SUMMARY OF MAIN POINTS

1. District and provincial standardized achievement testing is common, but few uses of the results were found.

2. Evaluation of student achievement for marks and promotion is carried out by individual teachers or small groups; there is little communication with others. As a skill, student evaluation is neglected and has a weak scholarly base.

3. Provinces and districts are turning to examinations as a substitute for the program consultants and inspectors that used to assist with quality control in schools.

4. Employers are more concerned with attitudes and behaviour than with marks, although more academically related businesses, such as banking and insurance, pay attention to marks.

5. Post-secondary institutions are concerned about the comparability of school marks, especially in the absence of common examinations. They are satisfied if supplied with scores from common examinations, although there is only very weak evidence that they make better decisions thereby.

6. Increasingly rapid technological change is putting pressure on schools to be flexible and to make the best possible use of the current strengths of faculty and community. Such pressures appear to be in conflict with the idea of a common core curriculum and common examinations.

RECOMMENDATIONS

1. **Districts should raise the status of evaluation by giving it more attention in professional development activity and supervision.**

Rationale. By far the most evaluation is done by individual teachers in classrooms. The work suffers from neglect, and the district level would seem to be the natural place to start giving it more attention. Staff closest to the teachers are likely to be the most effective, especially as provincial departments have too few resources to mount a wide effort. Such an objective seems worthwhile because it suggests that each district do what it can. Many officials could and would be pleased to prepare a paper on the importance of fairness and equity and the necessity to link evaluation to objectives (with examples) and to discuss the paper with principals. Principals (and department heads, where appropriate) could work out more of the operational details with teachers. Teachers will often respond with requests for and suggestions about evaluation in-service opportunities.

What was missing nearly everywhere during the study was a sense that it is important to do evaluation well, that there are criteria everyone can use now. It should be feasible to build on the consensus about fairness and equality as desirable objectives.

2. **Districts should develop promotional materials that explain to employers and the general public how students are evaluated and what marks mean.**

Rationale. From this study and from several on-line radio "phone-ins" lately, it is clear that the public either knows nothing about how schools evaluate students or has some quite distorted

ideas. This works against the students, the teachers and the cause of education in general.

Some districts may find that they are not ready to publish the description of their evaluation system as it now exists — in which case the exercise would be a very useful one. The public wants to be reassured that students get high marks for sustained effort resulting in solid achievement — and not for anything else. Such is precisely the case in the vast majority of schools and a little effort will enable officials to communicate these facts of educational life to the people who pay the bills. If the district policies need a little work before exposing them to public view, then now is as good a time as any to get them in shape.

3. Each provincial department should establish a task force on evaluation and technological change.

Rationale. The task force would have two mandates: (a) to consider how schools can make the best use of technology in evaluation itself (item banks, teacher and school record systems for profiles, testing for guidance) and (b) to consider how content about technology will get into the curriculum and be evaluated. Many claims will be made for computers and their offshoots, and provincial departments will do their districts a favour by convening an expert group, selecting some options and making recommendations. It is not implied that no scope is left to districts for experimentation, but only that the province can gather special resources and provide inspiration that most districts outside the largest cities cannot. Such a task force would be better advised to work for two months every two years than to take four months in the first year. Getting started modestly now is preferable to launching a major effort six months from now.

4. The CEA should organize a series of regional conferences for officials, teachers and trustees to discuss student evaluation — quality, policies, making the best use of examinations and communicating about the process to the general public.

Rationale. The district is a good place to tailor policies to local needs, but most of the problems are common ones. The cross fertilization that can happen at a regional conference could be a very valuable input to the task of modernizing student evaluation, and CEA is a natural agency to organize such gatherings. There are resource people in all parts of Canada who can help the conference along, but it would be a mistake if participants came expecting to sit back and be instructed. Participants should have at least an active interest, and preferably an involvement, in student evaluation policy or practice at the local level. Most of the time should be spent sharing fears, experiences and triumphs among themselves rather than listening to someone say how things should be done. Provincial

officials might choose to sit back and be instructed, the better to learn where the needs are the greatest.

5. The CEA should initiate a national study of teachers and the evaluation of student achievement.

Rationale. The present study provides the groundwork and the overview, but it also confirmed that the most important work goes on inside classrooms and inside teachers' heads. This study barely caught a glimpse of the real process in secondary schools and didn't really see it at the elementary level. It would serve everyone well, including the intrepid scholars who hope to advance the theory of assessment and testing, to document the beliefs, skills, fears and talents of a number of teachers about the student evaluation work they do.

Such a study ought to be responsive and proactive — responsive in that it portrays the situation faithfully from the teachers' points of view and proactive in that it obtains teachers' reactions to a number of ideas they may not yet have thought of. Some combination of these two features will be a useful way to learn what present reality is like but still see what the obstacles are to some of the changes that have to come in the near future. Crafts of various sorts greatly enrich our lives, and people skilled at crafts are truly admirable. A craft of student evaluation is inadequate to the needs of education, however, and ways must be found to move it toward a profession, if not a science.

APPENDIX A

Selected References[1] Relevant to the Model of Evaluation of Student Achievement (See Figure 1 on page 22)

Society's Institutions

Broadfoot, Patricia. *Assessment, Schools and Society.* London: Methuen, 1979.
Draws sociological analyses together to present a comprehensive account of the consequences of examinations for the life of the classroom. Explores developments in accountability, assessment of performance, and monitoring around a range of informal assessment techniques which not only affect classroom life but also adapt and modify the relationship between society and schools.

Rentrée scolaire, 1977: ce qui change: Actualités service. Paris: Service d'Information et de Diffusion, Ministère de l'Éducation, 1976.
Describes in detail the *dossier* and the process of orientation.

Russell, H.H., Wolfe, C., Evans, P., Wolfe, R., Traub, R., and King, A. *Programs and Student Achievement at the Secondary-Post-secondary Interface: Interproject Analysis.* Toronto: OISE Educational Evaluation Centre, 1976.
Synthesis of findings from an interrelated set of studies of the transition paths between high school and post-secondary education. (Note: Out of print; photocopies only.)

Science Education 11-18 in England and Wales: The Report of a Study Group. London: The Royal Society, November 1982.
The Study Group reviewed the teaching and examination of science (including mathematics), considered the needs of potential employers and how to meet these needs. They made 25 major recommendations to government, to the school system, to Examination Boards (and the projected Examination Council) and to the Council of the Royal Society.

Stager, David *Accessibility and the Demand for University Education.* Toronto: Commission on the Future of the Universities of Ontario, June 1984.
Discussion paper examining factors affecting accessibility to universities, including economic and social. Excellent bibliography.

[1]For references published before November 1983, this study relied mainly upon a computer search of the ERIC database, concentrating on student evaluation and academic achievement. The mainstream academic literature in education was searched manually from November 1983 to December 1984.

Airasian, Peter W., and Madaus, George. "Linking Testing and Instruction: Policy Issues." *Journal of Educational Measurement* 20 (No. 2, 1983): 103-118.

Establishes the context for the succeeding papers in this special issue on the state-of-the-art in linking achievement testing to the cognitive processes employed in test responses and to the instructional experiences of students. [N.B Six papers make up this special issue (for measurement specialists).]

In the editor's view, the foundation of achievement measurement rests heavily on the validity of the interpretation of a given measurement as the consequence of specific cognitive processes employed by the examinee.

Broadfoot, Patricia. "Alternatives to Public Examinations." *Educational Analysis* 4 (No. 2, 1982): 33-45.

Suggests answers to questions, "What is a public examination?" "Why do we need alternatives?" "What might these be like?"

Darling-Hammond, Linda, Wise, Arthur E., and Pease, Sara R. "Teacher Evaluation in the Organizational Context: A Review of the Literature." *Review of Educational Research* 53 (No. 3, 1983): 285-328.

Presents a conceptual framework for examining the design and implementation of teacher evaluation processes in school organizations. Research on teaching, organizational behaviour, and policy implementation suggests that different educational and organizational theories underlie various teacher evaluation models.

Fuchs, Lynn S., Deno, Stanley L., and Mirkin, Phyllis K. "The Effects of Frequent Curriculum-based Measurement and Evaluation on Pedagogy, Student Achievement and Student Awareness of Learning." *American Educational Research Journal* 21 (No. 2, 1984): 449-460.

A study in special education classes in New York demonstrated desirable teacher and student effects when teachers used the data-based modfication system.

Glasman, Naftaly S. "Student Achievement and the School Principal." *Educational Evaluation and Policy Analysis* 6 (No. 3, 1984): 283-296.

Principals were identified as most and least effective in efforts to improve student achievement. Both groups believed strongly that sharing data with teachers had a positive effect on achievement gains and that gains should be used to evaluate teachers. Fewer than half believed use of gains in evaluating teachers could affect classroom practice.

Le Mathieu, Paul G. "The Effects on Achievement and Instructional Content of a Program of Student Monitoring through Frequent Testing." *Educational Evaluation and Policy Analysis* 6 (No. 2, 1984): 175-187.

Both positive and negative effects of an intensive teaching/testing program were demonstrated.

Mortimore, Jo, and Mortimore, Peter. *Secondary School Examinations.* London: University of London Institute of Education, 1984. Bedford Way Papers No. 18.

A comprehensive examination of the British examination system — advantages, disadvantages and alternatives.

Nagy, Philip. "An Examination of Differences in High School Graduation Standards." *Canadian Journal of Education* 9 (No. 3, 1984): 276-297.
Analysis of process by which the Newfoundland Department of Education compares means of school marks with mean provincial examination marks and adjusts school marks that are too far out of line.

Teachers' Attitudes and Feelings

Lortie, Dan C. *Schoolteacher — A Sociological Study*. Chicago: University of Chicago Press, 1975.
Deals with a variety of issues in the organization of teaching work and inquires into various sentiments teachers hold toward their daily tasks. The unifying theme is a search for the nature and content of the ethos of the occupation.

Pedersen, K. George, and Fleming, Thomas. "Review of *Schoolteacher — A Sociological Study.*" *Canadian Journal of Education* 4 (No. 4, 1979): 103-110.
Lortie's emphasis is not so much on who teachers are, but on why they are who they are. The book is a handbook of researchable topics in education and the sociology of work and is as well informative to practitioners and teachers in the areas of administration and policy analysis.

Focused Learning Time

Brunelle, Jean, Tousignant, Marielle, et Godbout, Paul. "Notion de temps d'apprentissage et son évaluation en situation d'enseignement." *Canadian Journal of Education* 8 (No. 3, 1983): 232-244.
Integration of concept of learning time into research in physical education.

Gettinger, Maribeth. "Achievement as a Function of Time Spent in Learning and Time Needed for Learning." *American Educational Research Journal* 21 (No. 3, 1984): 617-628.
Model presented with quantitative estimates of causal influences derived from path analysis.

Peterson, Penelope L., Swing, Susan R., Stark, Kevin D., and Waas, Gregory A. "Students' Cognitions and Time on Tasks during Mathematics Instruction." *American Educational Research Journal* 21 (No. 3, 1984): 487-515.
Students' report on attention, understanding and cognitive processes were more valid indicators of classroom learning than observers' judgements of students' time on task. Students reported affect as well as cognitions mediated the relationship between instructional stimuli and student achievement and attitudes. In particular, students' negative evaluative self-thoughts may be potentially debilitating both in terms of student achievement and attitudes.

Airasian, Peter W., and others. "Proportion and Direction of Teacher Rating Changes of Pupils' Progress Attributable to Standardized Test Information." *Journal of Educational Psychology* 69 (No. 6, 1977): 702-709.

In 10 per cent of the cases, teachers raised their ratings after learning the test scores.

Bejar, Isaac I. "Educational Diagnostic Assessment." *Journal of Educational Measurement* 21 (No. 2, 1984): 175-189.

It is concluded that the development of powerful diagnostic instruments may require a reexamination of existing psychometric models and possibly the development of alternative ones. The psychometric and content demands of diagnostic assessment all but require test administration by computer.

Bellanca, James A. *Grading. NEA Professional Studies*. Washington, D.C.: National Education Association, 1977.

A brief overview of the social context for current grading practices forms the background for a discussion of alternatives to the assignment of letter or numerical grades to represent student performance.

Centra, John A., and Potter, David A. "School and Teacher Effects: An Interrelational Model." *Review of Educational Research* 50 (No. 2, 1980): 273-292.

Examines a model for investigating school and teacher variables which influence student achievement.

Engel, Brenda S. *Informal Evaluation*. Grand Forks: North Dakota Study Group on Evaluation, March 1977.

Intended for non-experts in evaluative techniques, this monograph presents suggestions and examples for assessing (1) the child, (2) the classroom, and (3) the program or the school

Fair, J.W. and others. *Teacher Interaction and Observation Practices in the Evaluation of Student Achievement*. Toronto: Ontario Ministry of Education, 1980.

This study investigated the importance and meaning of the role of observation in teachers' assessment of student achievement.

Holmes, Mark. *What Every Teacher and Parent Should Know about Student Evaluation*. Toronto: OISE Press, Informal Series/46, 1982.

A handbook of practical advice for teachers and parents from an ex-principal and director of education now a professor of educational administration.

Marx, Ronald W. "On 'Test Purposes and Item Type': A Comment on Mason." *Canadian Journal of Education* 4 (No. 4, 1979): 14-19.

Item type should be related more specifically to task domains, including the process components of objectives, and not simply to the referencing procedure for tests or their formative or summative role. (Mason's reply is in the same issue.)

Mason, Geoffrey, P. "Test Purpose and Item Type." *Canadian Journal of Education* 4 (No. 4, 1979): 8-13.

Constructed-response type of items will generally be required in both

formative evaluation and in criterion-referenced summative evaluation.

Mitchell, Allison C. "Using Microcomputers to Help Teachers to Develop their Assessment Procedures: A Development Project Report." *Programmed Learning and Educational Technology* 19 (No. 3, 1982).
Describes a Scottish project — "School-based assessment using item banking" — investigating the feasibility of producing computer-based [illegible] marking and reporting facilities.

Northcroft, David. "Education and Distributive Justice: Some Reflections on Grading Systems." *English in Education* 13 (No. 2, 1979): 7-18.
Focuses on the distribution of grades as symbols of educational merit. The social function of the artificially created shortage of high marks is discussed and different characteristics of grading systems are considered. The effects of co-operative and competitive distributive systems are summarized.

Quinto, Frances, and McKenna, Bernard. *Alternatives to Standardized Testing*. Washington, D.C.: National Education Association, 1977.
NEA suggests alternatives to standardized, norm-referenced tests: (1) performance contracts; (2) teacher-student and teacher-parent-student interviews; (3) teacher-developed tests; (4) criterion-referenced tests; and (5) an open admissions policy in higher education.

Richmond, John (Ed.). *English in the Schools — What Teachers Really Try to Do*. London: University of London Institute of Education, English Department, 1983.
Compilation of statements contributed by 230 teachers at the Language Teachers by Candlelight Conference of Language in Inner-City Schools.

Roid, Gale, and Haladyna, Tom. "The Emergence of an Item-Writing Technology." *Review of Educational Research* 50 (No. 2, 1980): 293-314.
A continuum of item-writing methods is proposed ranging from informal-subjective methods to algorithmic-objective methods. Each method is critically reviewed and empirical studies are described.

Traub, Ross E. "There's More to Know. And Different!" *OISE Field Development Newsletter* 14 (September 1983).
Critical review of the book by Holmes (*What Every Teacher and Parent Should Know about Student Evaluation*) by a professor of measurement and evaluation. Holmes's reply is in the same issue.

Student Characteristics Outside the Influence of the School

Belz, Helen F., and Geary, David C. "Father's Occupation and Social Background: Relations to SAT Scores." *American Educational Research Journal* 21 (No. 2, 1984): 473-478.
Father's occupation was associated with quantitative and verbal SAT scores. It is a potential interacting variable associated with scholastic achievement.

Edmonds, Ronald R., and others. "Comments on 'Should We Relabel the SAT . . . or Replace It?'" *New Directions for Testing and Measurement* (March 1982): 51-57.

The need for accuracy in testing, the unintended social consequences, and the contrast of achievement and aptitude tests are discussed in response to the views of Jencks and Crouse (see below) regarding whether to change the functions of the SAT.

Halsey, A.H., Heath, A.F., and Ridge, J.M. *Origins and Destinations: Family, Class and Education in Modern Britain*. Oxford: Clarendon Press, 1980.

Literature from developed Western countries is reviewed in discussing the link between background and education.

Jencks, Christopher, and Crouse, James. "Should We Relabel the SAT . . . or Replace It?" *New Directions for Testing and Measurement* (March 1982): 33-49.

Shifting from aptitude to achievement tests for college admissions is discussed with implications toward the positive educational effects of rewarding diligence and serious study in high school. (See "Comments on 'Should We Relabel the SAT . . . Replace It?' " above.)

Schulte, Dan. "The Relationship between IQ, Rates of Learning, Standardized Achievement Tests and Classroom Observation." Paper presented at The Council for Exceptional Children Conference on The Exceptional Black Child, New Orleans, February 1981.

There was a substantial relationship between IQ, standardized tests, and rates of learning, but not classroom observation. Observation, however, had the advantages of observing the current levels of academic responding, was not influenced by rates of learning, and had the capability of being diagnostic.

Svanum, Soren, and Bringle, Robert G. "Race, Social Class, and Predictive Bias: An Evaluation Using the WISC, WRAT, and Teacher Ratings." *Intelligence* (July-September 1982): 275-286.

A substantial relationship between standardized measures of IQ and achievement was found which was independent of race, but decreased with increasing socio-economic status.

Students' Feelings about School

Riley, Roberta, and Schaffer, Eugene. "Testing without Tears." *English Journal* 64 (No. 3, 1975): 64-68.

Various techniques for involving students in evaluation are described, all of which make evaluation a learning activity.

Steinkamp, Marjorie W., and Maehr, Martin L. "Affect, Ability, and Science Achievement: A Quantitative Synthesis of Correlational Records." *Review of Educational Research* 53 (No. 3, 1983): 369-396.

Science achievement is positively related to affect, but the relationship is weaker than was expected; science achievement correlates more strongly with cognitive abilities than with affect (interest, preferences).

Students Feelings about Themselves

Power, Marian E. "The Grading Syndrome." *Journal of Reading* 19 (April 1976): 568-572.

Describes what competitive grading procedures do to students and suggests alternatives.

Torshen, Kay Pomerance. *The Relationship of Evaluations of Students' Cognitive Performance to their Self-Concept Assessments and Mental Health Status*. Chicago: Illinois University, Department of Psychology, March 1973.

Norm-referenced grades assigned by teachers are significantly related to the students' self-concept assessments and mental health status. The author suggests that modifying evaluation methods can provide an important avenue for dealing with the extensive personality problems found in our schools.

Students' Educational Accomplishments

Biggs, John B., and Collis, Kevin F. *Evaluating the Quality of Learning — he SOLO Taxonomy*. London: Academic Press, 1982.

Concentrating on the meaningful learning of existing knowledge (reception learning), the authors developed the Structure of the Observed Learning Outcome (SOLO) taxonomy, a model of the objective and systematic assessment of the quality of learning.

Broadfoot, Patricia M. "Trends in Assessment: A Scottish Contribution to the Debate." *Trends in Education* 2 (Summer 1977): 35-38.

The effect of changes has revealed the inadequacy of current methods of assessment and certification.

Catherwood, Vince. *Assessment: The New Zealand Experience*. Wellington: New Zealand Department of Education, August 1980.

Describes an experiment with an alternative form of evaluation of New Zealand secondary school students' English proficiency (an internal measurement of skill achievement based on a student language profile).

Cornett, Joe D. "Alternatives to Paper-and-Pencil Testing." *NASSP Bulletin* (November 1982): 44-46.

Describes three alternative methods for evaluating student achievement using rating scales, checklists, and anecdotal records.

Dobrinski, Virginia, and Liechti, Carroll D. *Profiles of Performance*. Wichita, Kansas: Wichita Public Schools, Division of Research, Planning, and Development Services, November 1981.

The performance profiles indicate pupils' strengths and weaknesses and are used to help determine individual and group development programs.

Girard, Richard, Nadeau, Marc-André, et Scallon, Gérard. "Analyse d'erreurs conceptuelles dans le cadre de l'évaluation formative de l'apprentissage de concepts." *Canadian Journal of Education* 8 (No. 2, 1983): 174-187.

Demonstrates, in the context of formative evaluation, that a measuring instrument constructed according to the diagnostic model developed by Suzan Markle and Philip Tiemann is sufficiently sensitive and sufficiently reliable to detect the conceptual errors likely to be made by a student at the beginning of concept learning.

Haertel, Edward. "Detection of a Skill Dichotomy Using Standardized Achievement Test Items." *Journal of Educational Measurement* 21 (No. 1, 1984): 59-72.

Multiple-choice reading comprehension items from a conventional, norm-referenced reading comprehension test were successfully analyzed using a simple latent class model. A classification rule for assigning respondents to "mastery" or "nonmastery" states is presented, which simplifies the scoring procedure. (N.B. Articles such as these, which are clearly outside the frame of reference of teachers and officials, were not usually included. As computers are more widely used in schools these sorts of procedures can be evaluated for their utility in practice.)

McLean, Leslie D., and Ragsdale, Ronald. "The Rasch Model for Achievement Tests . . . Inappropriate before, Inappropriate Today, Inappropriate in the Future." *Canadian Journal of Education* 8 (No. 1, 1983): 71-76.

Reaction to the use of the Rasch model to construct mathematics achievement tests in British Columbia. A reply by the original authors appears in volume 8, number 2.

Morris, Joan (Ed.). "Educational Testing." *School Guidance Worker* 38 (March 1983): 5-59.

Contains 11 articles about educational testing focused on the topics of guidance and information management, student achievement, mathematics teaching/learning, test score interpretation, alternatives to standardized testing, evaluation of multicultural and exceptional children, learning/cognitive training assessment models and "blue collar" career inventories.

Parsons, James B. *Evaluating Student Achievement in Alberta Social Studies: Report to MACOSA Committee on Social Studies Assessment.* Edmonton: Alberta Department of Education, July 1977.

This bibliographic essay discusses evaluation instruments that could be used to evaluate the K-12 social studies program in Alberta. The author points out the difficulty of evaluating the Alberta social studies program because its objectives are ill defined and it relies heavily on values and the inclusion of the affective domain.

Scottish Council for Research in Education. *Pupils in Profile.* Edinburgh: Hodder & Stoughton, 1977.

Teacher-controlled and teacher-assessed profiles were systematically studied. Favourable reactions were tempered by complexity, cost and the in-service training required.

Spencer, Ernest. *Folio Assessment or External Examinations?* Edinburgh: Scottish Council for Research in Education, 1979.

Research favoured the O-grade English examination because it is a well tried method. Problems with the O-grade examination include intermarker inconsistency and lack of fine discrimination among large numbers of average students. Although folio assessment could be more directly related to particular courses, in-service education in grading practices would be necessary, marker inconsistency would be present, and some teachers would be reluctant to take on the work involved.

Stewin, L.L. "Research Notes: A Note on Pupil Evaluation in the Soviet Union." *Alberta Journal of Educational Research* (December 1980): 276-280.

Contrasts North American and Soviet views and approaches to student evaluations, especially in the area of testing differences in academic achievement.

Strathe, Marlene, and Krajewski, Robert J. "Testing in Nontraditional Curriculum Areas." *NASSP Bulletin* (November 1982): 33-38.

Achievement testing in nontraditional curriculum areas (such as industrial arts, physical education, or music) provides an ideal opportunity for developing students' self-evaluation skills. While applying testing procedures, teachers demonstrate what skills deserve evaluation and how to evaluate them.

Taylor, Hugh. "The Misuse of Grade Equivalent Scores." *School Guidance Worker* (March 1978): 11-15.

The findings suggest the use of standard scores as the most appropriate method to measure change.

Ulibarri, Daniel M., and others. "Language Proficiency and Academic Achievemen ." *NABE: The Journal for the National Association for Bilingual Education* 5 (No. 3, 1981): 47-80.

Wahlstrom, Merlin, and others. *Assessment of Student Achievement: Evaluation of Student Achievement at the Intermediate Level. Final Report*. Toronto: Ontario Institute for Studies in Education, June 1977.

Evaluation and assessment procedures of Ontario principals and teachers at the intermediate level (grades 7 and 8) were examined. All teachers indicated a desire for more standardized instruments, and for more training in the area of assessment procedures. In many ways, grades 7 and 8 present an extension of the procedures and practices used at the elementary level.

Wahlstrom, Merlin, and Weinstein, Edwin L. "Standardized Testing in Ontario Intermediate Schools." *School Guidance Worker* (March 1976): 43-47.

Includes a brief description of the tests that are commonly used at the intermediate level and the ways in which they are used.

APPENDIX B

List of Visits and Interviews by L.D. McLean and D. Welch in the Course of the CEA Study

British Columbia

- Educational Research Institute of British Columbia
- Ministry of Education, Learning Assessment Branch
- Vancouver School District No. 39
  - Templeton Secondary School
  - Maple Grove Elementary School
  - Chief Maquinna Elementary School
- Coquitlam School District No. 43
  - Port Moody Senior Secondary School
  - Dr. Charles Best Junior Secondary School
- Prince George School District No. 57
  - Meeting of officials and teachers

Alberta

- Department of Education
- Edmonton Public School Board
  - Jasper Place Composite High School
- Calgary Board of Education
  - John G. Diefenbaker High School
- County of Vulcan No. 2
- County of Wetaskiwin No. 10
  - Millet School
- Wetaskiwin School District No. 2641
  - Wetaskiwin Composite School

Ontario

- Ministry of Education
- Carleton Roman Catholic School Board — French Sector
  — English Sector
- Carleton Board of Education
  - Rideau Valley Middle School
  - W. Erskine Johnston Elementary School
  - Earl of March Secondary School
- Ottawa Roman Catholic Separate School Board — French Sector
  — English Sector
- Ottawa Board of Education — French Sector
  — English Sector
- Sudbury District Roman Catholic Separate School Board

61

- Sudbury Board of Education
  - École secondaire Franco-Jeunesse
  - Lasalle Secondary School
  - Wembley/Prince Charles Elementary School
- Lakehead Board of Education
  - Lakeview High School
- Dryden Board of Education

Québec

- Ministère de l'Éducation
- Commission des écoles catholiques de Québec
  - Pavillons Automobile et Coiffure
- Commission scolaire Ancienne Lorette
  - École Jacques-Cartier
- Commission des écoles catholiques de Montréal
  - Polyvalante Calixa-Lavallée
- Laval School Board

New Brunswick

- Moncton School District No. 15
  - Riverview Secondary School
- Shédiac district scolaire no 13
  - La polyvalante Mathieu-Martin
  - L'École intermédiaire Vanier

Newfoundland

- Department of Education
- Conception Bay South Integrated School Board
- Bonavista-Trinity-Placentia Integrated School Board
- Avalon North Integrated School Board
- Roman Catholic School Board for Ferryland District
- Avalon Consolidated School Board
- Roman Catholic School Board for St. John's

APPENDIX C

The Ontario Institute for Studies in Education
252 Bloor Street West, Toronto, Ontario M5S 1V6 Tel: 923-6641

Educational Evaluation Centre

March 2, 1984

The Canadian Education Association (CEA) has undertaken a national study of student evaluation programs. The study itself is being done by the Educational Evaluation Centre of the Ontario Institute for Studies in Education (OISE).

One aspect of the research is a study of the use of school marks by employers. For instance, we know that schools place a great deal of importance on student grading since this is thought to be one of the most important ways in which student accomplishment can be communicated to future employers. How valid is this belief? Please respond on this letter and return it in the envelope provided.

1. Does your company, in the hiring of recent secondary school graduates, consider high school grades in the choice of candidates for employment?
   No ☐
   Yes ☐ How much weight? _____

2. Do you place as much or greater emphasis on student attitudes (punctuality, attitude towards work, etc.)?
   No ☐
   Yes ☐

3. Do you feel the present means of student evaluation, as you understand them, present an accurate picture of what students learn at school?
   Yes ☐
   No ☐ What would you like to see?

4. Any other comments?

_____

_____

Any help you can give to us in this regard would be most appreciated.

Yours sincerely,

L.D. McLean
Educational Evaluation Centre

63