

LEARNING OUTCOMES

Henry Schulz, Glenn Clark, Robert Crocker

**Discussion Paper Prepared for the
Council of Ministers of Education, Canada
in support of the
Pan-Canadian Education Research Agenda**

December, 1998

Background

This discussion paper was prepared in response to a call from the Council of Ministers of Education, Canada for proposals for a series of discussion papers on various issues leading to the development of a Pan-Canadian Education Research Agenda. The context for the project is the increasing importance of education for individual and societal well-being, the need for accountability for the large investment of public resources in education, and the efforts over the past decade to institute large scale educational reform with a view to increasing efficiency and improving the quality of education. More specifically, the Pan-Canadian Education Research Agenda project seeks to direct educational research in Canada towards policy-relevant issues and especially to bring some coherence to a number of large scale surveys and indicators projects which have emerged over the past decade.

The topic being addressed in this paper is learning outcomes. Specifically, the review is guided by three broad questions: (1) how do we identify the full range of desired learning outcomes? (2) does a basis exist for assessing outcomes beyond the core school subjects? (3) how can we identify the key factors that affect learning and develop explanatory models for learning? In addition, a number of issues in the development, administration and use of assessments are discussed. In this paper, an attempt is made to identify both gaps in knowledge which might be filled by research specifically in the pan-Canadian context and some specific research issues which need to be pursued. The main focus of the paper is on the third question, since knowledge of what affects learning is key to making informed decisions on matters of policy and practice. The paper does not attempt to identify specific assessment instruments, nor does it examine provincial/territorial or school-based assessments.

Measurement of the Full Scope of Learning

Goal Statements

A brief review of provincial/territorial goals documents indicates that it is possible to classify goals into five broad categories, which may be labelled intellectual, social, utilitarian, attitudinal and personal enrichment. Intellectual goals include what is usually thought of as the basic literacy and numeracy skills as well as higher level goals such as problem-solving or critical thinking. Social goals have to do with citizenship, and transmission of culture and, in some cases, with spiritual and moral values. Utilitarian goals are those associated with preparation for work. Examples include technological and subject area competence, with overlap into basic skills. Attitudinal goals have to do with promoting learning, cooperation and social cohesion, and personal satisfaction. Finally personal enrichment goals include those related to physical fitness and fine arts and involve both performance and appreciation components. Obviously these goals are far broader than any of the goals captured by conventional assessments. The important question, of course, is whether some of these goals can be formulated in ways which are amenable to measurement.

Since the mid-1980s, most provinces and territories have conducted royal commission studies or other large scale reviews of education. Many of these reviews have resulted in restatements of provincial/territorial goals of education. One of the most striking results of these reviews has been a noticeable shift in goals away from general philosophical statements to statements expressed in outcome form. Typically, goal statements are now preceded by phrases such as “students should be able to . . .” This shift has important implications for measurement, because such outcome statements represent a first step in defining what needs to be measured and what it is possible to measure.

One of the clearest examples of the expression of goals as outcomes is found in the “Essential Graduation Learnings” developed by the Atlantic Provinces Education Foundation (APEF) (1996). Similar statements can be found in other jurisdictions (e.g., Ontario Ministry of Education and Training, 1993).

APEF has identified six broad goal areas as follows: (1) Aesthetic Expression (2) Citizenship (3) Communication (4) Personal Development (5) Problem Solving (6) Technological Competence.

From the point of view of measuring outcomes, what is most interesting is the development, under each of these broad areas, of outcome statements which identify what high school graduates are expected to know or be able to do. For example, under citizenship is found the statement “graduates will be able to demonstrate understanding of sustainable development and its implications for the environment.” Under technological competence is the statement “graduates will be able to demonstrate understanding of the impact of technology on society.” An examination of curriculum documents reveals even more explicit statements within specific subject areas at particular levels of schooling. For example under communication in the English language arts curriculum is found the statement “by the end of grade 6 students will, independently and collaboratively, use a range of strategies and techniques to revise selected drafts for cohesion, clarity and impact.”

This formulation possesses a number of features relevant to the issue of measuring outcomes. First, stating goals as outcomes immediately conveys the message that there is a requirement to determine whether the stated outcomes are being achieved. Second, there is clearly an explicit effort to identify the full scope of outcomes expected of K-12 education. Finally, the formulation extends beyond broad goal statements to a series of increasingly explicit outcome statement which, at their most specific, approach the operational level required to act as a basis for developing measurement instruments.

Expectations and Standards

Before we can move towards the measurement of a broad range outcomes, it is necessary to formulate the desired outcomes in terms of explicit behaviours which students should be expected to exhibit. In many cases, the required statements can be found newly-developed curriculum documents such as the Common Framework of Science Learning Outcomes (CMEC, 1997). Most current large scale assessment programs are based on a conceptual framework for the area of interest and on statements of performance expectations, similar to what can be found in the Science Learning Outcomes document.

One example of a comprehensive attempt to identify outcomes over a wide range of school subjects and life skill areas is a set of standards compiled by the Mid-continent Regional Educational Laboratory (McREL) in the United States (Kendall and Marzano, 1997). These standards have been validated by being synthesized to standards produced by a wide range of professional organisations. These are, of course, American standards and do not necessarily match desired pan-Canadian outcomes in some areas. However, the fact that such standards have been developed and professionally validated illustrates that it is possible to go beyond the core subject areas in determining measurable outcomes.

Two examples will illustrate the potential of the McREL document as a source of outcome statements that might be used as a validated base for the creation of measures of

non-conventional goals. Outcome statements are organized here at three levels. Physical education, for example, is broken down into five units, two of which emphasize use of physical skills while the other three emphasize understanding. These are each further subdivided and arranged so that from five to ten near behavioral outcomes are stated for each of the second level outcomes for key stages of schooling. The level of these outcomes is illustrated by the following examples: “(the student) engages in more advanced activities that develop and maintain muscular fitness and endurance.” “(the student) understands the role of exercise and other factors in weight control and body composition.”

Life skills is any even more interesting example of how it is possible to identify outcomes in the personal development area. This category is divided into four areas at the first level; thinking and reasoning, working with others, self-regulation, and life work. At the second level, for example, self-regulation is subdivided into six categories; sets and maintains goals, performs self-appraisal, considers risks, demonstrates perseverance, maintains healthy self concept, and restrains impulsivity. At the third level self regulation has thirteen separate statements. The level and scope of of these outcome statements is shown by the following examples: “(the student) has basic belief in ability to succeed”, “(the student) examines “shoulds” to determine their positive and negative effects”, “the student) uses high self-esteem body language”, and “(the student) avoids overreacting to criticism.”

Once outcomes are stated at this level of specificity, it is possible to see how assessment instruments can be developed in areas which go well beyond the core school subjects. Indeed, the example outcome statements are at a level comparable to what is found in the conceptual frameworks for large scale assessments such as the School Achievement Indicators Program (SAIP). This is not to underestimate the complexity of the developmental task, nor does it assume that SAIP-like paper and pencil instruments can be used in all situations. Indeed, the greatest developmental challenge is likely to lie in devising performance tasks for many of these outcomes and in finding ways to efficiently administer these tasks.

Public Validation of Outcomes

The development of goal and outcome statements is essentially a political process. In large scale reviews the process typically involves a series of public hearings and solicitation of briefs from concerned citizens and groups. In Manitoba, for example, consultation began with a series of meeting with teachers, school trustees, and superintendents, and continued with thousands of briefs being submitted at hearings and the creation of a Parent’s Forum on Education (Manitoba Education and training, 1994). Quebec’s Task Force on Curriculum Reform began by making use of documents produced by the Estates General on Education and followed this with wide consultation with education groups (Government du Quebec, Ministre de l’Education, 1997) In Alberta the process began with the organisation of two roundtables on basic education (Alberta Education,1993a).

All of these cases illustrate attempts to obtain some form of public for educational reform. However, the nature of the process and the fact that the form of consultation is controlled by “insiders” suggests the possibility that public views will not be fully represented. In many cases, the strongest voices are those of interest groups (both internal and external to the system) rather than of the public at large. This is reflected in the reluctance to place priorities on goals and perhaps even in the inclusiveness of goals. There is obviously a risk that outcomes formulated through such processes would not reflect more global public priorities A case can thus be made that the political process should be followed by a more scientifically valid process of determining whether the formulations produced do indeed enjoy public

support.

Given the complexity of the measurement task, and the time and resources needed to conduct the developmental work, it would also be desirable to ask what priorities the public would assign to the various categories of goals and outcomes before determining where initial efforts at broadening the scope of assessment should be focussed. Indeed, the public is likely to have some view on whether it is even appropriate to measure outcomes in some areas. For example, although enhancing self-concept or developing physical fitness might be accepted as reasonable outcomes of schooling, it does not follow that the public would support the testing of such outcomes. Indeed, it is possible that the public would view participation in certain activities to be sufficient evidence that the goal is being achieved, without demanding that particular levels of proficiency be demonstrated. Our first research issue is therefore:

Research Issue 1:

There is a need to assess the level of public support for contemporary provincial/territorial or regional formulations of outcomes and to determine the priorities that might be assigned by the public to various outcomes.

Professional Validation of Outcomes

It is reasonable to assume that attempts at public validation of outcomes would be conducted at a relatively high level of generality, and that the public at large would not be in a position to respond to detailed behavioural statements within particular curriculum areas. Finding the appropriate balance between public and professional views is, indeed, one of the difficulties in advancing the identification and assessment of outcomes. For example, the public might be asked in broad terms about the priority to be assigned to outcomes in science, citizenship or personal development but would not be expected to determine the priority or the placement of atomic physics, immigration policy or interpersonal relationships in the curriculum. All of this requires validation through conceptual frameworks, curriculum development and assessment activities in which the primary players are subject area professionals, curriculum developers, measurement specialists and other experts.

With the development of detailed outcome statements, and assuming an acceptable level of public support for broadened assessment, it seems quite feasible to move to assessment of a much wider range of outcomes than have conventionally been included in large scale assessments. A level of effort comparable to that exerted in the development of existing SAIP assessments would certainly yield comparable assessments in a variety of subject areas such as technology, social studies, or physical education. Somewhat greater effort would be required to elaborate conceptual frameworks and develop instruments in areas of personal development or aesthetic expression. That is not to say that assessments in all of these areas could be accomplished by paper and pencil instruments suitable for large scale survey use. More will be said on this issue later in the paper. The research issue at stake here is:

Research Issue 2

There is a requirement for professional validation of underlying contemporary statements of goals and outcomes and for the development of assessment instruments in areas outside the core school subjects for which there is public

warrant for measuring outcomes.

Key Factors that Affect Learning

Educational Indicators

One of the responses to increased concern about performance of the education system has been a surge of interest in the development of educational indicators. A large number of provincial/territorial, regional, pan-Canadian and international indicators projects may now be identified (e.g., Alberta Education, 1993b; Atlantic Provinces Education Foundation, 1996; Canadian Education Statistics Council, 1996; Centre for Educational Research and Innovation, 1996). A key purpose of indicator systems is to permit the investigation of analytical relationships between student learning and variables which may be changed through policy and program decisions. Developing parsimonious models which can account for such relationships should also be part of the conceptualization of an indicators system, although this rarely is the case.

While the conceptual and practical limitations of indicators systems have been pointed out in several reviews (e.g., Bryck & Hermanson, 1993), it may equally be argued that indicator systems have not yet reached their potential for public accountability or for explaining particular observed results. Most indicators systems developed to date have been largely confined to compiling descriptive and comparative data on such matters as system structures, enrolment, finances, levels of attainment, and achievement. A good example is SAIP which, although conceived as a comprehensive indicators project, has to date confined itself mainly to producing comparative indicators of achievement. Similarly the Pan-Canadian Education Indicators Program along with a variety of other large scale surveys such as the National School-Leavers Survey and the National Graduates Survey, have been concerned mainly with descriptive and comparative data. The main research issue to be examined in this section is the need to return to the conception of indicator systems as serving primarily an analytical rather than a descriptive purpose.

Basic Indicator Model

Most approaches to analysis are built around a straightforward model, commonly referred to as a production function, in which **outcomes** such as achievement are seen as influenced by **inputs** such as resources and infrastructure and by **processes** such as instructional practices engendered by these inputs. In its simplest form, this model may be depicted as follows:

Inputs _ Processes _ Outcomes

Conceptually, the I-P-O model is unidirectional and causal, in the sense that inputs are presumed to influence processes which, in turn, influence outcomes. Nevertheless, it is generally recognized that all of this takes place in an overall **context** determined by demographic features, social and economic conditions, infrastructure, and other broad characteristics of the society in which the enterprise operates. It also must be recognized that the direction of causality is not as straightforward as the model indicates. More generally, the system may be thought of as dynamic and interactive, where feedback from outcomes, for example, are used to modify inputs and processes. The system also needs to be thought of as an open one, in which learning is also influenced by variables external to the system which are

not amenable to policy changes within education.

The major limitation of the I-P-O model is that it is merely a skeleton, giving no indication of what relationships should be examined or even what variables are worth measuring. In working from such a model, there is a risk of reducing research to “correlational fishing trips,” in which there is no prior concept of which input or process variables might be related to outcomes. Ideally, what is needed is mature theory which serves to explain why certain relationships might be expected and from which hypothesized relationships, leading to new research, can be derived. While there is no dearth of theory in education, most of this theory is derived from other disciplines and is not readily applicable to school learning. In the absence of comprehensive theory, and to ensure that progress is made towards theory development, it is useful to continually extend the existing knowledge base as well as to think of ways of organizing this knowledge base into more useful models than conveyed by the basic I-P-O model.

The IEA (TIMSS) Model

Early studies conducted by the International Association for the Evaluation of Educational Achievement (IEA) were based on a restricted version of the I-P-O model. The focus of these studies was on the idea that “students learn what they are taught.” This led to an emphasis on the curriculum as the main influence on learning. Input, process and output were represented by the concepts of intended, implemented and achieved curriculum respectively. IEA studies have always placed great emphasis on analyzing curriculum documents and textbooks to obtain evidence on the intended curriculum, have used teacher ratings of topics or test items as evidence of the implemented curriculum, and of course have considered the achievement test results as the evidence of the achieved curriculum.

One of the most recent, and certainly the most comprehensive of the IEA series of studies, was the Third International Mathematics and Science Study (TIMSS). TIMSS elaborated on the core I-P-O model and earlier IEA models to develop a composite model in which four research questions were embedded in a framework which included the earlier concepts of intended, implemented and achieved curriculum, and which identified levels of the system responsible for various variables and which would then serve as the sources of data. This model is shown in Figure 1. Once specific variables within each of the components of such a model have been measured, analytical techniques now exist for examining the patterns of direct and indirect relationships among these variables and thus for empirical validation of the model.

- Insert Figure 1 Here -

(Note: it is not possible to electronically insert the graphic at this time; it will be available in hard-copy in the symposium package available at registration on February 16, 1999.)

As with many other indicators systems, the idea behind TIMSS seemed to be to gather as much data as possible and address the analytical issues later. Nevertheless, the TIMSS model is more specific than the basic I-P-O model in helping identify clusters of variables and the levels at which data might be gathered. In particular, while it appears obvious, the separation of variables into the four levels of system, school, teacher and student is useful in identifying

sources of data.

The Wang-Haertel-Walberg Synthesis

Perhaps the best example of an attempt to synthesize what is currently known about factors influencing achievement is the work of Margaret Wang and her colleagues, which has taken place over more than a decade, and which has appeared in the literature in various forms. The particular version to be discussed here appears in two major articles by Wang, Haertel & Walberg (1990, 1993) and in a further refinement of the model presented in 1994 (Wang, Haertel & Walberg, 1994). This work will be referred to collectively as the WHW synthesis.

In the 1993 version, WHW synthesized a total of 270 research reviews and meta-analyses, using a combination of content analysis, expert opinion, and meta-analysis. They identified 228 variables involving influences on learning. These in turn were organized into 30 scales, under six broad categories on an approximate proximal-distal continuum, based on how closely the variable touched on the lives of teachers and students in the classroom. For example, broad state and school district policies were considered the most distal variables, while time on task, discipline and other classroom variables were considered most proximal. These items formed the basis for the coding of more than 11,000 separate statements about the strength of the association between one or other of the 228 items and learning outcomes. The outcome of the whole exercise was a ranking of the 30 scales and the six broader classifications with respect to strength of association with learning outcomes.

In general, the results supported the hypothesis that proximal variables are more closely associated with learning than more distal variables. The order of influence of the six main categories was:

1. Program Design (e.g., curriculum and instruction)
2. Out-of-School Contextual Variables (e.g., home environment, out-of-school use of time)
3. Classroom Instruction and Climate (e.g., classroom management)
4. Student Variables (e.g., motivation, placement)
5. School-Level Variables (e.g., parent involvement policy)
6. State and District Variables (e.g., state level policy)

In a subsequent report (Wang, Haertel & Walberg (1994) the model was refined slightly to produce composite ratings for 28 of the original scales. From this analysis, it is possible to identify the specific scales which show the strongest and weakest associations with learning. The five strongest were:

1. Classroom management
2. Metacognitive processes
3. Cognitive processes
4. Home environment/parental support
5. Student/teacher social interactions

Motivation, peer group influences, quantity of instruction, classroom climate, and other proximal variables also received high rankings.

The scales receiving the lowest mean ratings were:

1. District demographics
2. School policies
3. State-level policies
4. School demographics
5. Program demographics

This WHW work is by far the largest attempt to synthesize the research on variables associated with learning. Nevertheless, there have been criticisms of this approach to developing a knowledge base. Some of these have been discussed in a series of commentaries on the review in the same issue of the Review of Educational Research. These include the issue of whether such a distillation provides any basis for policy-making (Elmore, 1993) or is of any use to teachers (Kleibard, 1993), the atheoretical nature of the exercise (Kerdeman & Phillips, 1993) and the failure to consider context (Hedges & Waddington, 1993).

Although WHW claim that their identification of a hierarchy of influences and amounts to a theory of school learning, and they certainly draw on other theories as their starting point, at best this can be treated as a “descriptive” rather than an “explanatory” theory. Most of the criticisms of this model fall under one of two areas of concern, the absence of plausible explanatory theories and the difficulties in conveying the results to practitioners in ways which would be convincing. This points to the need for theory development and for devising more comprehensive ways of disseminating the results.

What do we Need to Know?

From a pan-Canadian perspective, an obvious limitation of the WHW synthesis is the overwhelming influence of United States-based research and American perspectives on research in reviews of this type. Much such research is motivated and supported by what is widely regarded as a crisis in education in the United States and the initiatives being taken to extricate the nation from this crisis. While there is no way of avoiding this in comprehensive reviews, because most of the original research is American, it is not obvious that the same relationships would be found if a similar corpus of work could be found in other parts of the world.

A second and related point is that little of the review literature on which the synthesis is based comes from large scale studies. Most of the work summarized in the original reports involved bivariate correlations or small-scale experimental treatment studies, rather than large more externally-valid multivariate studies. With the current interest in expanding indicators systems there is some promise that a broader range of data will eventually become available. In addition, many countries, including Canada, have initiated projects which have the capacity of generating much more comprehensive and representative data bases than have previously existed. The planned enhancement of SAIP Science for 1999 is an example. The proposed inclusion of student, teacher and school questionnaires in SAIP will, for the first time, allow the investigation of the links between science achievement and a large number of input and process variables. Both the TIMSS conceptual model and the WHW synthesis are being used in developing the questionnaires to provide focus to the questions being asked and to capitalize on existing knowledge.

It is also important to note that the achievement outcomes used in most of the research

included in the WHW synthesis are mainly in the core subject areas of language, mathematics and science. It is not at all obvious that the same pattern of relationships would be found for outcome in other areas. For example, an argument can be made that the type of classroom process most conducive to high achievement in mathematics or science might not be what is required for the development of high self-esteem. In particular, in order to achieve high outcome levels in some areas it may be necessary to reduce emphasis on other areas. The simple matter of differential time allocations can be expected to contribute to differential outcomes.

One of the most striking results of the WHW synthesis is the relatively low influence of state and district policies on learning. This result is obviously problematic for policy-makers who, presumably, make their decisions on the assumption that they will have some impact. Nevertheless, when interpreted in terms of the distal-proximal continuum this result not only seems to make sense but also suggests that a different approach is needed to analysis. It might be argued, for example, that broad policies exert only indirect influences on outcomes but exert more direct influences on processes. More comprehensive attempts at modeling school learning could be built around such a hierarchical concept.

In the meantime, there are strong policy reasons for conducting studies in Canada which will contribute to our understanding of the factors which influence learning. Although some of the basic influences, such as cognitive processes or use of time are likely to be the same everywhere, factors such as parental influences, curriculum structures, classroom management and many others may well operate differently in a different society. Moreover, Canada is unique in the degree of independence of its provinces/territories in education, with the result that provincial/territorial policies may well play a larger role in influencing learning than would be the case in the United States and other countries.

Finally, advances in statistical methodology and in data processing capability in recent years now make it possible to conduct much more comprehensive analyses than previously. This makes it much more feasible to conduct more comprehensive analyses of large scale surveys. In particular, the ability to examine the effects of school and classroom practices and the influence of particular variables in the presence of other variables has been substantially enhanced. It is also feasible to analyze data sets which are structured in a hierarchical manner and to investigate the "fit to data" of various hypothetical models. The research community has barely scratched the surface of the potential inherent in new analytical tools, perhaps because there continues to be a preoccupation with comparative achievement and other relatively simple presentations of results, and relatively few resources are devoted to more comprehensive analysis. It is clear, however, that the time is now ripe to move from descriptive-comparative uses to more analytical uses of the large scale survey data that are becoming available.

The establishment of the Canadian Education Statistics Council, the recent initiation of several large scale surveys (SAIP, National Longitudinal Survey of Children and Youth, OECD Programme for International Student Assessment, the Canadian Longitudinal Youth in Transition Survey), and the current high levels of cooperation among major agencies (particularly the provincial/territorial ministries/departments, CMEC, HRDC and Statistics Canada) has set the stage for the development of the required comprehensive studies. Nevertheless, there is a need for greater integration of the various projects to ensure that the data bases can be merged and to ensure that resources for research are used to greatest effect.

In particular, the means will have to be found to integrate achievement data with the substantial amount of input and process data now being gathered. Indeed, an argument can be made that higher priority should be placed on integration than on developing additional surveys. In addition, given what appears to be a strong relationship between classroom processes and learning, there is a requirement to investigate more closely what goes on in classrooms. The most appropriate approach to this is through observational studies rather than through the more typical questionnaire surveys. Finally, there is a need to enhance the analytical phase of all of these studies and to develop the capacity to use the analytical tools available.

A research agenda designed to address the key factors affecting learning outcomes should thus attend to the following issues:

Research Issue 3

There is a need to develop a Pan-Canadian database from which long-term secondary analyses and longitudinal results can be derived.

While existing surveys cover an enormous range of context and input variables existing surveys are of limited use beyond description because there is as yet little process data and most of the existing data cannot yet be linked to achievement. It would be better to consolidate all of the existing efforts into one omnibus study than to continue with a series of unconnected studies.

Research Issue 4

There is a need to examine classroom events more closely than is possible in large scale surveys.

It is difficult to gather adequate data on classroom process through teacher and student surveys. The only way to get close enough to classroom events is through observational studies. Although such studies are costly and labour-intensive, they are no more so than large scale surveys which rely on individual interviews or which, like SAIP, require an immense test scoring effort. Such studies would be particularly useful as a means of corroborating promising relationships found in self-report data.

Research Issue 5

There is a need to examine more closely the factors that affect outcomes beyond the core subject areas. In particular, the possibility of conflicts in the conditions required to achieve different outcomes needs to be investigated.

Research Issue 6

There is a need to further develop the capacity for data analysis and model building, to build a comprehensive analysis phase into major projects and to devote a larger proportion of resources to this task .

Administration and Use of Assessments

This section examines two issues which follow from the previous discussion. The first concerns the direct involvement of teachers in the assessment process. This follows directly from the need to assess a broad range of outcomes, particularly those which require direct observation or performance components. The second is the use by schools and teachers of the results of large scale assessments to enhance learning. This is linked to the issues of factors influencing learning in the sense that, aside from other factors, feedback of assessment results is one direct means of influencing instruction.

Teacher Involvement in Assessment

Some outcomes obviously do not lend themselves to measurement by conventional paper and pencil tests. Indeed, in recent years, there has been much debate over finding more “authentic” ways of assessing outcomes, even in conventional subject areas, but especially in areas such as aesthetic expression or personal development. Good examples of this are found in the practical component of SAIP Science and in the SAIP writing assessment. The latter, in particular, represents a strong move to create an environment which resembles the conditions under which writing is normally done (e.g., development of the written document over time, successive drafting and editing). What is obvious from SAIP and other similar assessments is that this substantially increases the intensity and cost of the assessment effort, while still presenting validity problems because of the limited number of tasks used.

At the same time, it is obvious that assessment is an integral part of the daily work of teachers and students, that almost all teachers amass a substantial amount of information on the knowledge, skills, behaviours and attitudes of their students, and that such information is used regularly in making important decisions about students. This kind of data can greatly increase the scope of assessment information that is collected, but it leads to difficulties of systematic data collection procedures and comparability. However, there are ways of enhancing comparability across administrative conditions and tasks (e.g., Haertel & Linn, 1996). Also, large-scale assessment projects are beginning to develop procedures for using teacher-collected portfolios (e.g., Koretz, Stecher, Klein, & McCaffrey, 1994). Even in well-developed assessment areas such as direct assessment of writing, the criteria that lead to systematic assessment may not be in alignment with the process nature of written composition (Halden-Sullivan, 1996). These problems may be even greater in assessing aesthetic or personal development outcomes.

It follows that teacher assessments may have greater validity than the typical procedures imposed in large scale assessments, and that research needs to be directed towards capitalizing on teacher assessments. Moreover, performance assessments are notoriously difficult and inefficient to administer and score under large scale assessment conditions. The implication is that it would be extremely difficult to measure a broad range of outcomes unless we involve teachers directly in the process. There are two research challenges here. One is to develop ways to use what teachers already have available, while preserving the comparability requirements for appropriate interpretation of the results. The second is to develop large scale testing instruments in such a way that they can be administered by teachers, under authentic performance conditions and are sufficiently useful to teachers themselves to justify the effort required.

Research Issue 7

In order to move to the assessment of the full range of outcomes, there is a need for research on how to use, for system-wide purposes, the large amount of assessment information that teachers collect in the course of instruction, and to devise assessment procedures which may be administered and used directly by teachers.

Use of Assessment Results

The paper would not be complete without some reference to the use of large scale assessment results. In considering the factors that affect learning, assessment is an integral part of instruction, and is hence one of the most proximal influences on achievement. It thus seems plausible to argue that appropriate use of assessment results, including those from large scale assessments, should have a positive effect on learning. In moving to assess a broader range of outcomes, the effect might be to clarify and reinforce the point that outcomes other than those in the core areas are important and can be defined in ways that permit assessment.

Schools and teachers generally make little or no use of externally developed assessments, including those that are based on state or provincial/territorial curriculum (e.g., McLean, 1985). Teachers have little incentive to use the results of large-scale assessments to change instructional practice, and they believe the tests do not provide them with the timely, teaching-sensitive information they need for in-class use. For school or classroom use of assessment results, the data need to be available at these levels, and not just at system-wide levels of aggregation. The fact that some provinces and territories are now moving in the direction of testing all students and providing the results at the individual student level as well as various levels of aggregation supports this point.

LeMahieu and Leinhardt (1985) reported clear and substantial achievement gains in classrooms that were attributable to a comprehensive assessment program. Cohen and Hyman (1991) concluded the more precise the curriculum-test alignment, the better the instruction and the higher the resulting test scores. The quality of the learning, as well as the ethics of teaching directly to the test, have been questioned as being too narrow a focus for instruction and as representing inappropriate preparation of students for the test (Mehrens & Kaminski, 1989). This criticism appears less warranted with authentic forms of assessment, since these should be directly representative of the important learnings.

Keys to enhancing classroom level use of externally developed assessments are curriculum-assessment alignment, teachers' direct involvement in the process of the assessment and support for schools in the interpretation and use of the results. The alignment could be obtained through teacher involvement in specifying the assessment objectives and in developing the assessments themselves (e.g., preparing items, tasks, and prompts; scoring rubrics and procedures). Schools and teachers must also be involved in the administration of the assessments so that they are timely, and so that they can be administered in settings that are closer to what happens in the classroom and more authentic. Finally, teachers must take part in the scoring so they understand exactly what various scores mean, and in interpretation of results so that they have input into standards of performance.

All of this fits well with the idea that teacher involvement will be required as part of any major effort to expand the range of outcomes to be assessed. Thus, one process may serve two

different kinds of goals. Beyond this, it is also possible that the historic wariness of teachers and teacher organizations towards large scale assessment can be reversed through carefully thought-out procedures for bringing about greater teacher involvement in assessment.

Research Issue 8

Research is required on ways of making more effective use of assessment results to enhance instruction.

Summary and Conclusions

This paper has addressed a number of issues of identifying and assessing a broad range of learning outcomes, modelling the factors which affect outcomes and the administration and use of assessments.

An examination of provincial/territorial goal statements revealed a strong movement towards the statement of goals as outcomes. This includes outcomes in such areas as aesthetic expression and personal development, which go well beyond the core school subjects that have typically been the focus of large scale assessments. It was inferred from this that provincial/territorial policies are moving in a direction which points to a need to assess the achievement of non-conventional as well as conventional outcomes. Nevertheless, there remains some question of whether a public warrant exists for expanding assessment. An argument was thus made that research is required on the extent of public support for a much broader approach to assessment. The question of professional validation of outcomes and of the possibility of developing assessments in non-conventional areas was also raised. While there is evidence that outcome frameworks exist in a wide variety of areas at a level appropriate to form a basis for assessment, it is also clear that substantial development work would be required to validate such frameworks in a pan-Canadian context and to prepare the necessary assessment tools.

On the question of the key factors affecting learning outcomes, the strengths and limitations of educational indicators models and indicators projects were examined and one comprehensive synthesis of the knowledge base for school learning was reviewed. The research issues arising from this review were; (1) the need to consolidate existing large scale surveys into a data base which links achievement with the many context, input and context variables on which data are being gathered, (2) the need to examine classroom processes more closely using observational studies, (3) the need to extend the modelling of school learning to non-conventional outcomes and particularly to examine the possibility of conflict between conditions conducive to achieving different types of outcomes.

Teacher involvement in assessment is essential to ensure the full range of outcomes can be assessed, particularly since many outcomes involve performance components which cannot be addressed by the kinds of instruments conventionally used in large scale assessments. Research is required both to find ways of using the large amount of assessment information normally collected by teachers and also to involve teachers in assessing a broader range of outcomes and in developing new ways to conduct assessments. Finally, because of the importance of assessment for instruction, the paper has briefly addressed the classroom use of large scale assessment results and has suggested that research is required to develop more effective uses.

References

- Alberta Education (1993a). *Meeting the Challenge: What We Heard*. Edmonton: Author.
- Alberta Education (1993b). *Achieving Quality: Final Report of the Educational Quality Indicators Initiative*. Edmonton: Author.
- Atlantic Provinces Education Foundation (1996). *The Atlantic Canada Framework for Essential Graduation Learnings in Schools*. Halifax: Author.
- Atlantic Provinces Education Foundation (1996). *Education Indicators for Atlantic Canada*. Halifax: Author.
- Bryck, A.S. and Hermanson, K.L. (1993). Educational Indicator Systems: Observations on their Structure, Interpretation and Use. In L. Darling-Hammond (Ed.), *Review of Research in Education, Volume 19*. Washington, D.C.: American Educational Research Association.
- Canadian Education Statistics Council (1996). *Education Indicators in Canada*. Toronto: Author.
- Centre for Educational Research and Innovation (1996). *Education at a Glance: OECD Indicators*. Paris: Organisation for Economic Co-operation and Development.
- Cohen, S. A., & Hyman, J. S. (1991). Can fantasies become facts? *Educational Measurement: Issues and Practice*, 10(1), 20-23.
- Council of Ministers of Education (1997). *Common Framework of Science Learning Outcomes*. Toronto: Author.
- Elmore, R.F. (1993). What Knowledge Base? *Review of Educational Research*. 63(3), 314-318.
- Haertel, E. H., & Linn, R. L. (1996). Comparability. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 59-78). Washington, DC: National Center for Educational Statistics.
- Halden-Sullivan, J. (1996). Reconsidering assessment: From checklist to dialectic. *Assessing Writing*, 3(2), 173-196.
- Government du Quebec, Ministere de l'Education 1997. *Reaffirming the Mission of Our Schools*. Quebec: Author.
- Hedges, L.V. and Waddington, T. (1993). From Evidence to Knowledge to Policy: Research Synthesis for Policy Formation. *Review of Educational Research*. 63(3), 345-352.
- Kendall, J.S. and Marzano, R.J. (1997) *Content Knowledge: A Compendium of Standards and Benchmarks for K-12 Education (Second Edition)*. St. Louis, MO: Mid-continent

Regional Educational Laboratory.

- Kerdeman D. and Phillips, D.C. (1993). Empiricism and the Knowledge Base of Educational Practice. *Review of Educational Research*. 63(3), 305-313.
- Kleibard, H.M. (1993). What is a Knowledge Base and Who Would Use it if we Had One? *Review of Educational Research*. 63(3), 295-304.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13(3), 5-16.
- LeMahieu, P. G., & Leinhardt, G. (1985). *Higher test scores: Educational improvement or content manipulation?* Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Manitoba Education and Training, 1994, *Renewing Education: New Directions*. Winnipeg: Author.
- Martin, M.O. & Kelly, D.L. (1996) *Third International Mathematics and Science Study: Technical Report*. Chestnut Hill, MA. Boston College
- McLean, L. D. (1985). *The craft of student evaluation in Canada*. Toronto: Canadian Education Association.
- Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice*, 8 (1), 14-22.
- Ontario Ministry of Education and Training, 1993, *The common Curriculum Grades 1-9*. Toronto: Author.
- Wang, M.C. Haertel, G.D, & Walberg, H.J. (1994). What Helps Students Learn? *Educational Leadership*, December, 1993-January, 1994, pp 74-79.
- Wang, M.C. Haertel, G.D, & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*. 63(3), 249-294.
- Wang, M.C. Haertel, G.D, & Walberg, H.J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*.. 84(1), 30-43.