

SCHOOL ACHIEVEMENT INDICATORS PROGRAM

Bridging the Gap to Jurisdictional Assessments

Gilles Fournier
SAIP National Coordinator
Council of Ministers of Education, Canada

Although large-scale assessments have taken place in Canadian provinces for many years — some in Ontario date back to the 1930s — there continues to be controversy surrounding these assessments. These controversies relate to the administration, the suitability of content, the reliability, the interpretation and the dissemination of results, and the appropriate uses of the assessment results.

While taking some of these concerns into account, this paper will examine the School Achievement Indicators Program (SAIP) as an example of the advantages that large-scale assessments can provide. Our experience with three cycles of SAIP assessments shows that these advantages include: measures allowing for the accountability of ministries/departments of education to the public for Canada's education systems; a picture of student achievement in core subjects across jurisdictions and over time; the collection of valuable inter-related achievement and contextual data that assist researchers and policy-makers to make the changes that enhance student achievement. The data from SAIP have been used in a variety of ways by provincial and territorial ministries and departments of education: to evaluate their own assessments and standard-setting processes; as an aid to monitoring the outcome of policy changes; and as an indication of which parts of the education system are strong and which require attention.

Historical Background

Because of the federal/provincial distribution of powers in Canada, responsibility for education falls to the government of each jurisdiction within the confederation, and no federal department of education exists. There were no concerted efforts to examine educational issues on a pan-Canadian basis until the Council of Ministers of Education, Canada (CMEC) was established. Founded in 1967, the Council includes all ministers responsible for education from the ten provinces and three territories, providing them with opportunities to discuss and act on

educational issues of common interest and to speak out and act on issues of pan-Canadian and international scope.

In Canada, as in many other countries, increasing attention has been paid over the past years to education systems and their performance.

Do our schools prepare students appropriately to live in an age where economic competition is global in scale and where learning is a lifelong enterprise?

Parents, the labour market, and the business world, as well as the taxpayers who finance education systems all ask that question.

In their response to the question, ministries and departments of education have participated in a variety of studies. Most provincial ministries and departments have taken measures to assess students at different stages of their schooling. On the international level, through CMEC, jurisdictions have taken part in the International Education Indicators Program of the Organisation for Economic Co-operation and Development (OECD); some have also participated individually in various achievement studies such as those of the International Assessment of Educational Progress (IAEP) organized by the International Association for the Evaluation of Educational Achievement (IEA).

Given their common concern for maximizing the effectiveness and quality of education systems, ministers felt it useful to create an additional collective mechanism to assess system performance in each jurisdiction. They therefore decided to develop a pan-Canadian program to evaluate student performance.

SAIP Assessments and Questionnaires

In 1989, CMEC initiated the School Achievement Indicators Program (SAIP), the first-ever attempt to arrive at a consensus on the elements of a national assessment. The ministers decided to assess the achievement of 13-year-olds and 16-year-olds in reading and writing and mathematics, adding science later on. The first pan-Canadian assessment was administered in 1993 — in mathematics — and the reading and writing assessment in 1994 followed it; the science assessment was administered in 1996. The second cycle began again with mathematics in 1997, using the same assessment instruments in order to provide a good basis for longitudinal comparison of results from one cycle to the next.

CMEC began the third cycle in 2001 with the administration of the mathematics assessment. For this cycle, developers have been allowed to change some of the test items so that innovations in the pedagogy of the subject areas could be incorporated into the assessments. Of course, a core set of items serving as anchor points was retained to ensure comparability with the previously administered assessments. Ministers agreed to administer the same assessment instruments to both age groups, in English and in French, in order to study the change in student knowledge and skills following additional years of instruction.

SAIP assessments are the result of a collaborative effort by all provinces and territories and ultimately require each jurisdiction to approve all facets of the process. In addition, excellent financial and technical support is provided by Human Resources Development Canada.

SAIP results do not identify the performance of individual students, schools, or school boards. SAIP is essentially a measure of how well each jurisdiction's education system is doing. It does not replace individual student assessment, which is the responsibility of teachers, schools, boards, and ministries and departments of education.

Since curricula differ from one part of the country to another, comparing data on how students perform after being taught this diverse content is a complex and delicate task. Students across Canada, however, do learn many similar skills in reading, writing, mathematics, and science. SAIP assessments can therefore help determine whether students attain similar levels of performance at about the same age.

The original intent of SAIP assessments was to describe what students knew and what they were capable of in each of the three designated subject areas. This information would then be used to compare student performance from one jurisdiction to another, to report to the public about the efficiency and equity of each system on a pan-Canadian basis, and to provide information that would improve instruction in the participating jurisdictions. The program was designed to answer two questions for the various ministries and departments of education.

What is the level of achievement of students in Canada in the three subject areas being assessed?

Has the achievement of Canadian students in these three subject areas changed over time?

To attempt to answer these questions, the three assessment packages were developed using a common framework. They were designed to be criterion-referenced assessments, allowing ministries/departments to report results according to five levels of performance: Level 1 being the performance expected of students that are at the onset of specialized studies in a subject area and Level 5 being the performance expected of students once they have completed their secondary education with specialized studies in the subject area.

In order to obtain a representative sample for Canada and for each participating jurisdiction, students are selected randomly from the entire population of 13-year-olds and 16-year-olds across the country. For the smaller jurisdictions, this means that a significant proportion of their students participate in the assessment in order to assure a large enough sample to generate small confidence intervals when interpreting the data.

Course of study or grade level is not a consideration when students are selected to participate. This allows for an assessment of student performance in each sampled population as well as for comparisons between assessments. It should be noted that the same instruments are administered in both official languages (English and French) to both 13-year-old and 16-year-old students in each assessment.

The assessment developers designed the instruments with differing expectations of performance for each age group, with Level 2 being the expected level of performance for 13-year-olds, and Level 3 being the expected level of performance for 16-year-olds. The assessment criteria are not drawn up to reflect the curriculum of any one particular jurisdiction. Rather, these criteria were intended to define what experts in each of the subject areas considered to be what students of each age group should know and be able to do. SAIP is the only Canadian large-scale assessment that provides outcome data simultaneously for two age groups, demonstrating learning progress over time.

Each assessment consists of **two** components, each administered simultaneously to separate samples of students: a knowledge-based component and a performance-based or an integrated-skills component. Administration of the assessment takes approximately two and one-half hours, with extra time allowed for the completion of an accompanying

student questionnaire. Since the assessment is to be as inclusive as possible, some accommodations for special-needs students are permitted.

Each assessment in the first two cycles of SAIP was accompanied by a student questionnaire designed to provide some contextual information for the output data. As a result of a review of SAIP commissioned by CMEC in 1997, assessments administered since 1999 have included an expanded student questionnaire and two new components, a teacher questionnaire and a school questionnaire, all of which are linked directly to the performance data.

The student questionnaire provides information on students' grade level, language spoken at home, parental education level, gender, aspirations beyond high school, extracurricular interests, and other facts and opinions about a variety of matters pertaining to the subject matter being assessed. The teacher questionnaire provides information about teachers' professional background, instructional practices, kinds of students, and attitudes toward teaching of the subject areas. The school questionnaire provides information about the characteristics of the school, its staff, its facilities and services as well as information about the community in which the school is located.

Once again, SAIP does not provide reports on individual student performances, or describe performance by school, school district, or region, therefore making it a "low stakes" assessment.

Dissemination of Information

Two types of report follow each SAIP administration. The public report (usually released one year after the administration) on student performance is compiled for Canada as a whole; it illustrates through charts, graphs and tables the results for each jurisdiction based on age, gender, and language of instruction for those jurisdictions with distinct linguistic populations. It also provides the results of the Pan-Canadian Expectations-Setting Session, where a group of Canadians apart from the education sector consider what they believe the performance of students should be. These expectations are presented in the report for comparison with actual performance results.

For the assessments administered in the second cycle, the reports also provide comparisons of performance from one assessment to another. As of 1999, the public report includes an addendum that summarizes

information gathered through the expanded student questionnaire and the new teacher and school questionnaires. As noted in the addendum, the comprehensive Wang, Haertel and Walberg 1990 study of factors affecting student performance provided the framework for the additional questions through which information is gathered on seven categories of contextual factors affecting student performance: Provincial/District Context; Out-Of-School Context; School; Student; Program Design; Teacher; Classroom Instruction and Climate. The subcategories that were part of the design of the questionnaires are listed in the 1999 CMEC publication titled *Science Learning: The Canadian Context*.

The technical reports, intended for researchers and jurisdictions, provide a detailed description of the assessment, including a variety of different analyses of the student results, the reliability of the scoring process, a series of item analyses to determine difficulty of items and consistency of forms, and multiple regression analyses to examine the links between the questionnaire and the student performance data. Analyses using Item Response Theory and differential item functioning were used to examine results for the mathematics and science assessments allowing for detailed analyses of results by domain and skills as described in their respective framework documents. Tabulations of student responses correlated to the questionnaire response are also present in these technical reports.

Controversies around Assessment

As noted at the beginning of this paper, there continue to be controversies surrounding large-scale assessments, mostly related to issues of administration, suitable content, reliability, interpretation and dissemination of results and uses of results.

Some people label assessments as instruments of state control over educational matters; others hail assessments as instruments of improvement and quality assurance. The irony is that both views are legitimate in their assumptions about large-scale assessments. Large-scale assessments serve a variety of purposes, which include those mentioned above. Issues of accountability, compliance, effectiveness of pedagogical practices, and improvement of learning conditions are all part and parcel of the purposes of assessments, whether they be large-scale exercises mandated by the state or classroom assessments used by individual teachers assessing the progress of their students.

The resistance to the renewed interest in large-scale assessments in the 1980s was based on the concern that assessments would determine —

and therefore limit — curricular content to what was being tested, which would push policy decisions toward solutions that might not be pedagogically sound.

Uses of SAIP

However, the results of SAIP assessments do provide information that jurisdictions can use in conjunction with the data from their own assessment activities to set or to alter policy. Consequently, it is often difficult to determine the precise impact of SAIP results on individual jurisdictions.

Jurisdictions with their own assessment programs require complementary information that will either reinforce or contradict the information provided by their own endeavours. For those jurisdictions without their own assessment programs, they receive basic information about their students' performance and, in addition, the contextual information that can influence policy measures to a much greater degree than in other jurisdictions.

Generally, SAIP results can be used in four categories: Accountability; Program Improvement; Benchmarking; Exemplary Processes

1. Accountability

Accountability drives assessment. The public's desire to know how well students are performing is not a new phenomenon. Schools without report cards would make little sense. What is new is the public's growing preoccupation with the competitiveness of students, their preparedness for the work environment, the suitability of their acquired knowledge and skills for meeting the demands of an economy in constant flux, which drives the need for assessments of a quite different design from those administered by teachers in the past.

Assessments of the overall education system allow comparisons of student performance at the school, district, provincial-territorial, pan-Canadian, and international levels. The public's desire to know drives the development of new and comprehensive assessments in varying forms. Leithwood et al. (1999) state quite clearly that the intended purpose of the administration of a particular assessment dictates the method of reporting. SAIP's stated purpose has been to provide ministers of education with comparative information about their individual education

systems. Leithwood goes on to state “With jurisdiction-wide reporting of results, the national, province, state-wide education authorities that they sponsor are being held accountable to the public at large” (p.59). The SAIP reports provide the public with an account of how students performed in each jurisdiction, by age, by language of instruction, and by gender. SAIP is an accountability tool, measuring the performance of students over time in various subject areas. But it also gathers a large amount of contextual information related to academic performance.

Does it meet the definition of an indicator system? Leithwood et al. make a distinction between **an indicator system**, which was the original purpose for the development SAIP, and **a monitoring system**, which requires that “collected information be translated into courses of action” (p.60). Such courses of action are not part of CMEC’s mandate — any activity stemming from SAIP results is the responsibility of each jurisdiction. Therefore, SAIP as it is delivered by CMEC, is not a monitoring device, but it does become one if individual jurisdictions act on the results. There is a hefty body of anecdotal evidence indicating that SAIP results and reports do inspire action in the provinces and territories that participate.

To improve student performance, educators must examine the present state of the systems of education at the policy level, at the curricular level, or at the delivery level to determine what actions would bring about the desired change. By examining SAIP’s comparative performance data, each jurisdiction can determine which other jurisdictions could help them to improve student performance. Alliances among jurisdictions foster cooperative ventures that provide educational support for those with limited resources, but seeking to improve student learning outcomes. SAIP results are instrumental in this process. Since SAIP also reports on student progress over time, jurisdictions that have implemented curricular change can observe and report on the effectiveness of their innovations.

A case in point occurred at the very onset of SAIP. When the SAIP Mathematics assessment was administered in 1993, the New Brunswick Francophone Education Sector had just begun the implementation of a new mathematics curriculum. Students participating in SAIP that year had not been exposed to the new curriculum, but those participating in the 1997 SAIP Mathematics assessment had. SAIP allowed the provincial ministry of education to demonstrate to their public that differences in the assessment results were a clear indication of the positive impact of the new curriculum on student outcomes. It is

anticipated that SAIP will demonstrate the impact of curricular changes for other jurisdictions proceeding with innovations in the delivery of educational services.

For smaller jurisdictions with no assessment program of their own, SAIP is often the only tool at their disposal to report to the public on the state of their system. SAIP's sampling methodology is such that, for smaller jurisdictions, the number of students participating in SAIP is significant, allowing for accurate determination of the performance of student populations, almost on a school-by-school basis. In some jurisdictions, essentially all 13- and 16-year-olds participate in SAIP, which allows for a comprehensive examination of student outcomes. This provides jurisdictions with valuable information about school-level determinants of pedagogical effectiveness on which to report. It is most evident in these jurisdictions that using the SAIP assessments as evidence of accountability leads directly to program improvement.

2. Program Improvement

CMEC does not recommend policy initiatives to improve program delivery or curricular changes since it is the responsibility of each jurisdiction to examine student performance results for this purpose. Jurisdictions with their own assessment programs examine SAIP results as a counter measure or a quality control measure linked to their own systematic review of student performance. This allows them to gauge the effectiveness of actions taken on their part or the relevance of their policies vis-à-vis the pan-Canadian context.

Some jurisdictions in the process of implementing curricular change examine longitudinal data of student performance to determine the extent to which these curricular changes have a positive impact on performance results. Smaller jurisdictions without local assessment programs can use SAIP results as a direct measure of the efficacy of their curricular content and pedagogical structures because of the inclusiveness of the sampling design. For these jurisdictions, participation in the assessment program is near census in scope because their population size necessitates participation by a large proportion of their students. Based on SAIP results and on the performance descriptors for the criteria, they can then plan effective measures to improve curricular content and put into place policy initiatives serving that purpose. By comparing their results with those of larger populations, they can examine the contextual information and determine which policy initiatives based on observed best practices from other jurisdictions would meet their needs.

In addition, comparisons of gender performances, and the outcomes for minority-language populations have drawn a lot of interest from jurisdictions. CMEC, with Ontario acting as lead on this project, has commissioned a study of contextual factors affecting minority francophone student performances on SAIP assessments. It is hoped that the recommendations from this study would be forwarded to jurisdictions and would lead to improved learning outcomes for these populations.

Jurisdictions are also examining performance differences between boys and girls in the reading and writing assessments and are considering a plan of action that would entail the modification of some teaching strategies in an attempt to minimize this gender gap. Measures to address this issue have been implemented by some jurisdictions as a result of the review of the SAIP Mathematics II data.

Many jurisdictions also perform in-depth analysis of the information provided in the technical report. Saskatchewan, in its comprehensive standard-setting sessions, uses information derived from the SAIP student and teacher questionnaires, along with other provincial sources of information, to examine students' opportunities to learn. These opportunity-to-learn measures are published alongside provincial standards and the province's outcome results, and are distributed to school divisions and schools. Administrators and teachers can then examine their practices and plan toward increasing students' opportunities to learn and improving students' learning outcomes in the desired skill and application areas.

Item Response Theory analyses allow for the re-scaling of items within one assessment in order to provide a continuum of performance results on one standard scale. This allows jurisdictions to examine their student results in light of their curricular content, enabling them to ascertain if the domains within one subject area are being equitably treated in comparison to what other jurisdictions have deemed important to prioritize. Since the SAIP design is based on a consensus of what subject experts deem important in curricular content, jurisdictions are performing curriculum match studies in order to examine whether some adjustments are required in their own programs.

3. *Validation of Benchmarks*

Benchmarks, established by ministries/departments of education, serve as goals toward which students should progress. They indicate the knowledge and abilities that students are expected to have acquired once they reach certain grade levels. Benchmarks serve as points of reference from which measurements of student performance may be made.

A less often publicized use of SAIP results is the validation of benchmarks using SAIP as a measure of the validity of standards set by the jurisdictions. It may be less publicized but it is one gaining in importance, especially among those jurisdictions with comprehensive assessment programs. While jurisdictions set their own benchmarks for student achievement and outcomes, they look to external sources for assessment packages to validate the locally set standards. By examining their students' performance in comparison to those of the rest of the country, they can ascertain how demanding their standards or benchmarks are. In other words, SAIP becomes an indicator of the level of difficulty of their individual benchmarks. This allows jurisdictions to determine whether or not the benchmarks set locally are realistic or appropriate. For those jurisdictions that wish to focus on standards of performance as part of their accountability process, this allows them to use SAIP criteria and other resources, such as the *Common Framework of Science Learning Outcomes* (1997) as the basis of locally designed performance standards for a curricular topic or subject. By using student outcome information from SAIP, they can determine whether their program content should be realigned to conform to curricular expectations.

This brings us to the question of pan-Canadian standards of student achievement for the different subject areas. Even though SAIP was not the sole determinant in the decision to set pan-Canadian standards of achievement in science, it did play a role in convincing jurisdictions of the advantages of participating in such a project. Many jurisdictions adopted these learning outcomes, and the next SAIP Science Assessment will provide feedback on the effectiveness of their implementation.

Controversy surrounding the adoption of content standards and achievement standards in the other subject areas has prevented any further development along these lines, but it is important to note that descriptors of performance for all SAIP assessments have been examined by curricular experts in the jurisdictions who have reached a consensus

that they represent adequately what students at two particular ages should know in the three subject areas. In fact, criticism stemming from opponents of SAIP state that the acceptance of these performance criteria serves to undermine the autonomy of jurisdictions in setting their own curricular priorities, an opinion not shared by the jurisdictions themselves. It is, however, doubtful that pan-Canadian content or achievement standards for other subject areas can be adopted at the present time.

4. Exemplary Processes

SAIP results do not just provide performance results. SAIP is a total program involving informed framework principles, using the most recent innovative instrument designs, complex procedures for the administration and scoring of the assessments, and the most current analytical techniques. SAIP assessments are developed by experts in each subject area, from across Canada. Most of these experts have extensive experience in the field of student assessment design within their jurisdiction and have demonstrated proficiency in designing innovative assessment instruments. Some jurisdictions are looking not only at SAIP results, but are examining the processes encompassed by SAIP with the intent to adopt many of its innovative techniques for their own testing programs. The development of every aspect of SAIP is based on obtaining the cooperation of all jurisdictions in the examination and approval of its components and relies on obtaining consensus from all jurisdictions as to the acceptability of the criteria, instruments and publications that devolve from the results. Many jurisdictions send personnel to participate in SAIP-related activities to gain expertise that can be shared with their staff upon their return. Technical information about our scoring sessions, especially those associated with mechanisms assuring reliability and validity, have been scrutinized and utilized by jurisdictions in their own scoring sessions. Observers have attended scoring sessions and expectations-setting sessions in order to prepare guidelines based on SAIP practices to facilitate similar local exercises. SAIP processes and procedures are considered state-of-the-art and continue to attract the best in jurisdictional personnel because these are considered valuable professional development exercises.

Future Orientation

Aboriginal education issues are fast becoming a major concern to jurisdictional authorities. In some jurisdictions, the Aboriginal population attending local schools is increasing at a faster rate than that

of non-Aboriginals. Yet relatively little information is available to the jurisdictions about the challenges these students and their teachers face in preparing them for the work environment and in knowing how to meet those challenges. Minority groups, such as francophone students outside Quebec or Aboriginal students, face unique challenges as indicated in their overall performance in SAIP. SAIP could be used to gather more information on the performance of these students providing some comparative data about the factors that might improve educational opportunities and learning environments for them. The contextual information from the SAIP questionnaires could also provide a valuable resource for research in this area.

In at least one subject area measured by SAIP, reading and writing, there continues to be a gender gap in student performance — in favour of girls. Many jurisdictions have put measures into place to reduce this performance discrepancy, and SAIP is well suited to provide cross-sectional data to help determine ways to reduce this gap.

Over the last few years, a philosophical shift in the way to respond to the requirements of students with special needs has taken place. SAIP does provide information about some services provided to this group of students, but more work is needed to provide valuable information about their performance and how to better assess their learning successes.

For some jurisdictions, SAIP continues to act as a monitoring device for curriculum implementation. Future developers of SAIP instruments must ensure that modifications to the instruments, be they at the criterion level, the descriptor level, or the test item level, remain pertinent to the needs of those jurisdictions that value their use.

Cross-curricular skills and general competencies, beyond the mastery of a particular subject matter or skills in a particular area, have been included in some of the recent international assessments. In part, this is a response to the needs of the knowledge economy, which requires flexibility and the ability to integrate information quickly and to use it in a variety of contexts. SAIP must examine the possibility of integrating the assessment of cross-curricular competencies, in order to provide data for this emerging area. The question is whether an assessment based on general competencies would give information that could be tied to specific programs.

SAIP Research

Until recently, relationships between performance and contextual elements have not been the focus of SAIP research. With the introduction of the expanded questionnaires, however, exploration of the possibilities for research using SAIP data has become an important element of our program. SAIP will continue to gather data to confirm standards and results from provincial assessments and to provide good comparative information about student performance for each jurisdiction and over time. The possibility of creating linkages between SAIP and international assessments, and between SAIP and provincial assessments remains to be explored. These linkages, too, would assist in the development of a body of research examining the factors that affect student achievement.

At the moment, SAIP data allow for the tracking of the performance of Aboriginal students, French immersion and Francophone students, English as a Second Language students, and students in special education programs. It is also possible to track the performance of students who receive their education by different means of delivery (virtual schooling, home education, distance learning) relative to national and international results.

Conclusion

Some jurisdictions are strengthening their own assessment programs. Others are increasing their involvement in international assessment programs. Some jurisdictions are doing both. This heightened interest in large-scale assessment may, in part, be a measure of the influence of SAIP as a model of assessment. In light of these jurisdictional developments, SAIP must continue to provide information that complements jurisdictional efforts or that fills a gap with data they cannot obtain on their own.

The world of assessment has changed since SAIP's inception. SAIP, like all programs, must evolve and adapt if it is to remain valued by its supporters and users. As assessment programs have been initiated or expanded in many jurisdictions, SAIP may be seen by some decision-makers and field-based or school-level staff as an added burden or responsibility. One of the tasks for our SAIP team is to ensure that the value of the program justifies that burden.

The value of SAIP derives from: a combination of its ability to provide comparable pan-Canadian data on student achievement, over a period of eight years; its rich bank of contextual data and the opportunities that it presents for researchers; and the information that SAIP results offer to individual jurisdictions — whether as the jurisdiction’s main assessment program, as an external measure against which to consider the jurisdiction’s own benchmarks, or as a source of information about the performance of minority students within the jurisdiction.

In order to provide the greatest usefulness to jurisdictions, SAIP must complement and add value to jurisdictions' own assessment programs. As described earlier, the provision of more extensive equity-related data is one area that has been identified for the development of SAIP. Other possible future steps could include assessing knowledge, skills, and competencies that jurisdictions are currently not assessing, such as cross-curricular competencies; non-content-specific critical thinking and problem solving; second-language acquisition, especially through immersion programs; fine arts; information technology; citizenship and social responsibility. The possibility of creating linkages between SAIP and other assessments needs to be explored. There will also be an increased emphasis on the development of a body of research using SAIP data, in order to provide information to policy-makers on the links between achievement and context variables. These may prove to be the most valuable contributions SAIP can make to education in Canada.

REFERENCES

- Broadfoot, Patricia (1996), "Performance assessment in perspective: international trends and current English experience", in Harry Torrance (ed.), *Evaluating Authentic Assessment: Problems and Possibilities in New Approaches to Assessment* (Buckingham and Philadelphia: Open University Press).
- Council of Ministers of Education, Canada (1997) *Common Framework of Science Learning Outcomes, K to 12: Pan-Canadian Protocol for Collaboration on School Curriculum* (Toronto).
- Council of Ministers of Education, Canada (1994) *SAIP Reading and Writing Assessment 1994* (Toronto).
- Council of Ministers of Education, Canada (1997) *SAIP Mathematics Assessment 1994* (Toronto).
- Council of Ministers of Education, Canada (1999) *SAIP Report on Reading and Writing Assessment 1998* (Toronto).
- Council of Ministers of Education, Canada (2000) *SAIP Report on Science Assessment 1999* (Toronto).
- Eisner, Elliot W. (January 2001), "What Does It Mean to Say a School Is Doing Well?", *Phi Delta Kappan* Vol. 82 (No. 5), 367-370.
- Eisner, Elliot W. (May 1999), "The Uses and Limits of Performance Assessment", *Phi Delta Kappan* Vol. 80 (No. 9), 658-660.
- Gatto, John Taylor, Kristin Kearns Jordan, Theodore Sizer, and Thomas Stewart (September 2001) "School on a Hill: On the design and redesign of American education", *Harper's Magazine* Vol. 303 (No. 1816), 49-63.
- Haertel, Edward H. (May 1999), "Performance Assessment and Education Reform", *Phi Delta Kappan* Vol. 80 (No. 9), 662-667.
- Kohn, Alfie (January 2001), "Fighting the Tests: A Practical Guide to Rescuing Our Schools", *Phi Delta Kappan* Vol. 82 (No. 5), 348-357.
- Leithwood, Kenneth, Karen Edge, and Doris Jantzi (1999), *Educational Accountability: The State of the Art* (Gütersloh: Bertelsmann Foundation Publishers).

- Madaus, George F. and Laura M. O'Dwyer (May 1999), "A Short History of Performance Assessment: Lessons Learned", *Phi Delta Kappan* Vol. 80 (No. 9), 688-695.
- Merrow, John (May 2001), "Undermining Standards", *Phi Delta Kappan* Vol. 82 (No. 9), 652-659.
- Moss, Pamela A. and Aaron Schutz (Spring 2001). "Educational Standards, Assessment, and the Search for Consensus", *American Educational Research Journal* Vol. 38 (No.1), 37-70.
- O'Neil, John (February 1991), "Drive for National Standards Picking Up Steam", *Educational Leadership* Vol. 48 (No. 5), 4-8.
- Peshkin, Alan (December 2000), "The Nature of Interpretation in Qualitative Research", *Educational Researcher* Vol. 29 (No. 9), 5-9.
- Robertson, Heather-jane (May 1999), "Bogus Points", *Phi Delta Kappan* Vol. 80 (No. 9), 715-716.
- Rose, Lowell C. and Alec M. Gallup (September 1999), "The 31st Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools", *Phi Delta Kappan* Vol. 81 (No. 1), 41-58.
- Saskatchewan Education (January 1999), *Saskatchewan Opportunities to Learn Related to Reading and Writing: Opportunity to Learn - A Shared Responsibility* (The Ministry).
- Saskatchewan Education (March 2000), *Saskatchewan Opportunities to Learn in Science 1999* (The Ministry).
- Shepard, Lorrie A. (October 2000), "The Role of Assessment in a Learning Culture", *Educational Researcher* Vol. 29 (No. 7), 4-14.
- Spalding, Elizabeth (June 2000), "Performance Assessment and the New Standards Project: A Story of Serendipitous Success", *Phi Delta Kappan* Vol. 81 (No. 10), 758-764.
- Stake, Robert (May 1999), "The Goods on American Education", *Phi Delta Kappan* Vol. 80 (No. 9), 668-670.
- Wang, M.C., G.D. Haertel and H.J. Walberg (1993), "Toward a knowledge base for school learning", *Review of Educational Research* Vol. 63 (No. 3), 249-294.
- Weiss, Carol H. *Evaluation: Methods for Studying Programs and Policies – Second Edition* (1998), (Upper Saddle River, New Jersey 07458: Prentice Hall, Inc.).