

Education Policy Analysis Archives

Volume 10 Number 46

October 20, 2002

ISSN 1068-2341

A peer-reviewed scholarly journal

Editor: Gene V Glass

College of Education

Arizona State University

Copyright 2002, the **EDUCATION POLICY ANALYSIS ARCHIVES**.

Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the [Education Policy Studies Laboratory](#).

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the [ERIC Clearinghouse on Assessment and Evaluation](#) and are permanently archived in *Resources in Education*.

Senior School Board Officials' Perceptions of a National Achievement Assessment Program

Marielle Simon

Renée Forgette-Giroux

University of Ottawa

Citation: Simon, M. & Forgette-Giroux, R. (2002, October 20). Senior school board officials' perceptions of a national achievement assessment program, *Education Policy Analysis Archives*, 10(46). Retrieved [date] from <http://epaa.asu.edu/epaa/v10n46.html>.

Abstract

The *School Achievement Indicators Program (SAIP)* has been collecting data across Canada on 13- and 16-year-old student achievement in mathematics, in science, and in reading and writing since 1993. In 1999, it completed its second assessment cycle and was reviewed in Spring 2000. The review design included a survey of officials from all the school boards/districts that participated in the science assessment program held in 1999. The results of this study show that this stakeholder views as the most pressing issue for SAIP to succeed in its mandate, the need for development in four areas: a) Increased teacher

and student motivation to participate wholeheartedly in the program; b) Effective dissemination options; c) Leadership through innovation in teaching and in assessment practices despite high accountability orientation; and d) Cost-effective, yet rigorous means of providing both snapshot information and longitudinal means of comparisons. Although universally appealing, such approaches have yet to be supported by sound educational theory and methodology.

Since 1993, the *School Achievement Indicators Program (SAIP)*, under the responsibility of the Council of Ministers of Education, Canada (CMEC), has been collecting data across Canada regarding 13- and 16-year-old student achievement in Mathematics, in Science, and in Reading and Writing. In 1999, it completed its second assessment cycle and was reviewed in Spring 2000. The review had three objectives: a) To determine the degree to which the CMEC had succeeded in implementing the recommendations adopted from the review of the first round of assessments (Crocker, 1997); b) To measure the extent to which SAIP's objectives, set at the beginning of the second cycle, had been attained; and c) To formulate specific recommendations about the SAIP's aims, operations, and uses. To achieve these objectives, the review design included various data collection approaches, one of which was an investigation of the perceptions of all the school boards across Canada that participated in the 1999 science assessment, toward the national assessment program. The purpose of this paper is to report the results of this survey in order to provide researchers and policymakers with a better insight into the general interests, particular views, and specific needs of one of the most important stakeholders in large-scale educational assessment programs.

As a large-scale assessment program, SAIP is similar to the National Educational Assessment Program (NAEP), conducted in the United States. Both are national, cyclical programs, administered across regions (states or provinces/territories). In both countries, these regions have sole jurisdictional rights over education. Like NAEP, SAIP is designed to complement existing assessments in each province and territory. It is essentially a standards-based program focussing on assessment of content and ability via a mixture of multiple-choice, constructed response, or short hand-on performance tasks. The program also consists of student, teacher, and school administrator questionnaires intended to provide contextual data. Testing usually occurs in May and the final report is made available a year later.

As mentioned above, SAIP is governed by the Council of ministers of education, Canada. The CMEC's role in the SAIP essentially combines that of the United States National Centre for Education Statistics (NCES), responsible for NAEP's operation and technical aspects, and the National Assessment Governing Board (NAGB), mandated to select the subject areas to be assessed and their content framework. In Canada, the CMEC also coordinates national participation to other large-scale international testing programs such as the Program for International Student Achievement, run by the Organisation for Economic Co-operation and Development (OECD) and the Third International Mathematics and Science Study (TIMSS) governed by the International Association for the Evaluation of Educational Achievement (IEA). Not all the Canadian jurisdictions participate in the various international programs. For a more detailed look at the general CMEC operations, SAIP's sampling techniques, and the nature of the results, see for example the most recent SAIP Report (CMEC, 2000) or the CMEC WEB site. (Note 1)

SAIP was initially established to meet the following three objectives: a) To set educational priorities and plan improvements to curricula; b) To provide the best education to all young Canadians; and c) To report on certain indicators to the Canadian public. Despite these noble goals, SAIP's practices and orientations are geared principally towards accountability and overall instructional enhancement rather than improving local teaching, learning, and assessment practices. In 1999, all thirteen Canadian jurisdictions, namely the ten provinces and the three territories, participated in the Science assessment program. A random sample of students was drawn from each of the participating jurisdictions and, within some of these, students were also sampled by linguistic groups (English and French). The next four sections respectively present the framework, the methodology, the results, and a discussion of the investigation component of the program review.

Framework

The theoretical framework for the investigation evolved first from the examination of three documents: The call for proposals initiated by the Council of Ministers of Education, Canada, the report on the review of SAIP's first cycle (Crocker, 1997), and the official SAIP memorandum of understanding between the Human Resources and Development, Canada and CMEC, 1999). Second, various documents on general program reviews were consulted, such as the *Standards for Evaluations of Educational Programs, Projects, and Materials*, (Joint Committee on Standards for Educational Evaluations, 2nd Ed., 1999). This also included works on: a) The purpose of a program evaluation (Wilde & Sockey, 1995), b) The role of program reviews (Chelimsky & Shadish, 1997), c) The use of indicators (Posavac & Carey, 1997; Shavelson, McDonnell & Oakes, 1991a, 1991b), and d) The methodological approaches to program evaluations (Boulmetis & Dutwin, 1999; Popham, 1999). Third, the literature review for this study considered other actual systematic evaluations of large-scale assessment programs, related models, or proposals (Crooks, 1996; Madaus & Pullin, 2000; Ryan, 2002; Shepard, 1977). Finally, the previous experiences of this study's authors in the field of program evaluation also contributed to shaping the framework of the review of SAIP's second cycle (Cousins & Simon, 1993; Macdonell, A., Forgette-Giroux, R., Schmidt, S., Mougeot, Y., & Levesque, J., 1999). This led to the development of a framework that essentially consisted of five general areas of questioning on SAIP: a) Nature and position among other assessment initiatives, b) Goals and objectives, c) Operations, d) Design, and e) Impact. The nature and position focussed on the role, function, relationship, differences, similarities, and linkages between SAIP and other large-scale assessment programs. The appropriateness of SAIP's objectives and the identification of possible barriers to attaining these objectives concerned the second theme. SAIP's operations dealt with sampling, standard-setting, grading, and reporting procedures currently in place to meet the stated objectives, whereas validity, reliability, and questions that focussed specifically on content and format, nature of data collected, and motivational issues, all served to study SAIP's design qualities. Finally, its impact was examined through questions on significance, perceptions, values, and overall influence of SAIP's results on various educational and public settings.

Methods

The investigation described here represents only one component of the methodological design for the review and targets only one group of stakeholders: School boards (districts or councils) officials. Interviews with jurisdictional coordinators or directors were also conducted for the review along with in-depth content analyses of various relevant documents. These various components enabled the triangulation of the data that were ultimately reported in an aggregated format. The following sections present the subjects, instruments and research design of the survey component of the SAIP review.

Subjects

All school boards or districts across Canada with one or more schools that participated in the 1999 Science assessment program were contacted and invited to participate in the investigation. This meant reaching a total of 412 school boards from 19 jurisdictions: Ten provinces and three territories, with six of these also broken down into two linguistic sub-populations at the time of the assessment. Only one member of each board, a senior official, was asked to participate. This official could be a Director of Education, a Superintendent of Education, a designated SAIP school Official, a local Coordinator or Liaison person, a Board Consultant, or the Principal of a school that was involved in the 1999 SAIP assessment. In some areas, the Director of Education is the highest ranking official while in others it is the Superintendent of Education. When both are found within a school board, the Director of Education usually has priority.

Instrument

A written questionnaire was developed based on the framework's five themes. A first draft was submitted for validation to two experts familiar with SAIP, one French-speaking and one English-speaking. Their task was to determine the relevancy, comprehensiveness, and linguistic equivalence of the questionnaire. The final version included seven general information questions and six attitudinal questions with 73 sub-questions. Twelve questions offered multiple options while two were open-ended although all questions provided space for additional comments. The open-ended question # 13, for example, asked "What suggestions would you offer regarding SAIP and its various components in order that your school board or district gives it high priority among all assessment initiatives? Of particular interest however were questions 9 and 10. Question 9 asked : "To what extent do you agree with the following *actual* SAIP parameters?", while Question 10 read as: "To what extent do you agree with the following *proposed* SAIP parameters?" The proposed parameters emerged from comments made by the SAIP's designated jurisdictional coordinators who had previously participated in the semi-structured interviews, from the recommendations that resulted from the review of SAIP's first cycle, and from their relevance with respect to general school boards' interest and needs.

Procedure

The questionnaires were sent to the top administrator of each of the 412 school boards. The senior officials were instructed to fill out the questionnaire themselves or to forward it to someone from the school board that had been actively involved in the 1999 science assessment. The questionnaires were distributed during the last week of May 2000 with

specific written instructions to return the completed questionnaires in a self-addressed envelope by June 30, 2000. Two sets of follow-up telephone calls were made by a superintendent to selected school boards from each of the jurisdictions and their sub-population in order to achieve a maximum rate of return by all sampled populations. The first was conducted in the second week of June 2000 to approximately half of the boards within each jurisdiction to see whether they had received the questionnaires. The next series of calls were made in August to approximately twenty randomly selected boards from each jurisdiction that had a return rate below 30 % by the end of July.

Results

In all, 147 questionnaires were completed and returned, yielding an overall response rate of 36%. Of the 19 populations, four had all their participating schools in a single school board. Two of these four boards returned the completed questionnaire. Closer examination of the distribution of responses by jurisdictions reveals that the nine “smaller” jurisdictions and minority groups, that is those that had from two to 12 school boards, yielded response rates varying from 33% to 100 %, with a mean of 59% and median of 54%. The six “larger” jurisdictions, namely those with over 12 school boards (actually between 22 to 62), gave response rates varying from 27% to 36 %, with a mean of 32 % and a median of 33%. The rate of response for the 15 jurisdictions with two or more school boards therefore ranged from 27% to 100%, with an average of 48% and a median of 46%. It can also be reasonably stated that respondents were drawn from both small and large school boards and districts, from both remote and rural settings, from both central and urban areas, and from both the majority and minority linguistic groups.

Most questionnaires (80%) were completed by Directors of education, Superintendents or Board consultants. Eighty-six percent of respondents said they were more or less familiar with SAIP or knew it well. Respondents participated mainly in the coordination of the study, the test administration, and the communication of results. Although most questions included a four-point rating scale, dichotomized results (e.g. totally agree and agree versus more or less agree with disagree) are reported here with respect to each of the five themes.

Nature

The specific information on this topic was provided mainly in the two open-ended questions. Respondents generally recognize that SAIP provides a valuable index from a national perspective but state that it should distinguish itself from international and provincial initiatives by: a) Highlighting cross-curricular competencies if sampling remains age-based; b) Introducing innovative teaching, learning, and assessment approaches that are applicable to the classroom; and c) Adopting a diagnostic and interpretative approach to contextual and achievement data. Approximately 70 % of respondents also suggest that secondary analyses be performed on the data collected.

Goals

Although indicating being familiar with SAIP, three quarters of respondents (75 %) also admit that they are somewhat or not very aware of its stated goals and objectives. With respect to the appropriateness of the objectives, they generally suggest the need for an

objective that would implicate some form of comparison of results. The rank order of the percentage of respondents in agreement with the various types of comparisons of results is presented in Table 1.

Table 1
Rank Order of Percentages of Respondents in Agreement with
Types of Comparisons of Results

Longitudinal comparisons from one assessment cycle to next	87%
With national results (averages)	81%
With national expectations (standards)	65%
Between the two age groups (13 and 16)	64%
With local expectations	50%
Among jurisdiction results	44%

Table 1 shows support from school board officials for comparison of data from one assessment cycle to the next and with the national averages.

Operations

Results under operations are reported in terms of SAIP's general administrative, sampling, expectation/standard-setting, scoring, and reporting procedures. Eighty-nine percent (89 %) of respondents agree with the statement that SAIP meets the given time-lines and 67% believe that SAIP is relatively easy to administer. It is interesting to note that only 51 % of the respondents favour the month of May as the best time for test administration, 68 % disagree with conducting these in February or March, and 77 % disagree with a Fall administration, thus failing to achieve a consensus on the best time of year for administrating the various assessment programs. Three quarters of respondents favor sampling of 13- and 16-year-old students rather than sampling grades. Two thirds agree with comparisons of results with expectations or standards and generally approve of the five-point rating approach to scoring. Finally, nearly 70 % of respondents support the idea of disseminating some form of school board, school, or individual-based results for motivational purposes.

Design

Questions on SAIP's design asked for level of agreement with the assessment's focus on collecting disciplinary and contextual data as well as from theoretical and practical aspects of achievement data gathering techniques within Mathematics and Science. They also looked at motivational issues and perceptions around assessment cycles. Results show that 77 % of respondents agree with the collection of contextual data. Most concur with SAIP's mandate to conduct assessments in Mathematics, in Science, and in Reading and Writing and over 85 % favor both the theoretical and practical components of the science and mathematics assessments. However, 66 % of the respondents state that SAIP does not sufficiently encourage or motivate students to give their optimal

performance, thus addressing validity and reliability concerns. Finally, nearly 80 % of respondents support the three-year assessment cycle within a discipline.

Impact

Eighty three percent (83 %) of the respondents believe that the program has little or no positive impact on assessment practices in the classroom and approximately three quarter of respondents say that SAIP has little positive influence on setting priorities in the various disciplines assessed, on curricula, on public perception of the quality of education, on research initiatives, and on teaching practices. Approximately 85 % of respondents, however, would recommend those school boards that have not yet participated to the various SAIP programs to do so.

Discussion and Conclusion

Despite the overall low response rate obtained in this study, the results are telling. With one exception, the higher response rates are provided by those jurisdictions or minority populations with less than 12 school boards. Although many jurisdictions take part in other large-scale assessment program such as TIMSS and PISA, most respondents indicate general support for the SAIP because it is the only national long-term assessment program in which all thirteen educational jurisdictions and respective linguistic sub-populations participate. They also appreciate the complementarity of SAIP's results to those obtained through regional and other international assessment programs. Respondents stress, however, that SAIP should be forward looking and be more than a simple indicator system serving accountability purposes. Their responses lead toward interesting suggestions for each of the five themes. These are discussed in the following sections.

Nature, role, and position

With respect to the nature, role, and position of SAIP, the respondents value the nationally representative and continuing aspects of SAIP. Given that participation is voluntary, however, it becomes important that SAIP continues to be attractive and relevant to this stakeholder, particularly to those school boards that have sufficient resources to implement their own assessment program or to participate in most international ones. In that sense, SAIP should consider three suggestions offered by the school boards. First, they propose that if SAIP remains an age-based program as opposed to grade-based, then it should go beyond the assessment of disciplinary contents and basic skills in order to focus on socially relevant general competencies such as critical thinking, information management, speaking skills, and on attitudes such as civic values, self-awareness, self-esteem, and student engagement (Jones, 2001). Major organizations, such as the IEA and the OECD, already provide leadership in the assessment of such competencies via their own large-scale programs. At the onset, this option may appear to be duplicating efforts and resources given that many jurisdictions also participate in these major international assessment programs but, so far, SAIP has had the advantage of involving all jurisdictions and of being better able to respond to national educational concerns.

Second, as mentioned above, the respondents wish to see SAIP as more than an indicator

program, one that would offer specific perspectives for adjustment and intervention with respect to setting educational priorities, targeting curricular improvements, highlighting best teaching strategies, and fostering innovative assessment practices. This push for a shift from an indicator system, i.e., one that provides information that can be used to improve education, to a monitoring one, in other words, one that further analyses and interprets key contextual and achievement data in order to propose prescriptive feedback, stems largely from those smaller jurisdictions with scarce resources. The resulting feedback could subsequently translate into the introduction of innovative teaching, learning, and assessment approaches that are applicable to the classroom. This last statement relates to the respondents' third request.

These three requests contributed to the formulation of two specific recommendations in the review: a) That the concept of achievement indicator be broadened to include assessment of general competencies as defined by the pan-Canadian public and b) That SAIP be assigned a diagnostic function with interpretation of the most obvious links between contextual data and achievement (Forgette-Giroux & Simon, 2000). These recommendations are rather demanding. As with other large-scale studies, SAIP must re-examine its priorities with respect to accountability versus instructional orientations toward educational improvement (Popham, 1999) and eventually aim at establishing a balance between the two goals. So far SAIP has been mainly oriented toward the need for greater accountability. Moreover, it must rely on sufficiently sound theory or design to efficiently explore any relationships among background variables and achievement, to claim causal inferences or to explain why the comparison of certain groups yields different results (Bechger, van den Wittenboer, Hox, & De Glopper, 1999).

Goals and objectives

Although SAIP's present objectives are ambitious, universal, and aim at continually moving targets, thus ensuring their enduring validity, respondents wish to add another objective concerning the specific comparison of results from one assessment cycle to the next and with national data. This implies that SAIP should be given the dual goal of providing a snapshot of current achievement levels across jurisdictions and of measuring progress over time. However, in addition to the lack of proper theory to explain comparative differences in achievement as mentioned above, such an objective can also create tensions such as those experienced throughout NAEP's history, because a single assessment system cannot adequately serve such diverse purposes, each with its own set of assumptions, processes, and consequences (Linn, 2000). For example, longitudinal programs must find ways to rely on stable and reusable instruments while remaining fully aligned with national standards, actual curriculum contents, and evolving theories. In order to meet these challenges, NAEP presently operates two systems, the Main NAEP for longitudinal measures, and the State NAEP for cross-state comparisons. A recent NAEP review, however, has called for streamlining and for merging some aspects of the two programs (Pellegrino, Jones & Mitchell, 1999). If the CMEC decided to stress pan-Canadian longitudinal comparisons from one cycle to the next and comparisons with national averages, then such a decision would have major consequences on the entire program's structure and development (Bechger, et al., 1999). SAIP is currently not designed to meet such goals and does not have the resources to conduct parallel systems. As with all large-scale studies of achievement, if SAIP decided to opt for the comparative route, then it would have to meet at least three conditions to ensure the comparative validity of its results: Construct equivalence, scale equivalence, and

measurement equivalence (Bechger, et al., 1999). So far, these equivalencies are unattainable mainly because the theoretical frameworks underlying the most universally accepted competencies such as reading, problem solving, and scientific reasoning are constantly evolving, performance scores across time and groups are interpreted using arbitrary scales, and scale equivalence is greatly hampered by issues such as test translation. Perhaps the only way out at the present time is to document and make all related information as comprehensive and transparent as possible in order to arrive at the most valid interpretations when comparing results.

Operations

Respondents generally agree with the existing parameters around SAIP's administrative, sampling, expectation/standard setting, scoring, and reporting procedures. Some of their comments, however, stress greater participation of various school board personnel in all stages of SAIP's assessment programs. Similar statements are expressed in scholarly papers such as that of Hunter and Gambell (2000), and with respect to other large-scale studies such as New Zealand's National Assessment Monitoring Project (NEMP) (Flockton & Crooks, 1997). They claim greater empowerment and satisfaction by teachers and by other school board members who participate in substantial and independent training sessions and in the actual implementation of various aspects of the assessment program. Moreover, respondents point toward operational changes that consider greater teacher involvement and teacher input. As a result, the SAIP review offered a recommendation stating that CMEC implement, for example, a two-step standard-setting process in which the first series of expectations be formulated principally by teacher representatives.

Another operational issue that has not been overtly stated by the respondents but that was raised indirectly through their responses is the debate over grade- or age-based sampling. Although age-based sampling is seen as a more probabilistic in theory than a grade-based approach, in practice, it is not always respected. Local SAIP coordinators experience difficulty in implementing the recommended student sampling process at the school level (Forgette-Giroux & Simon, 2000). This is probably due to the fact that practitioners tend to view a classroom as a unit and thus see age-based sampling as causing significant disruptions to the class. As a result, student sampling is not done uniformly because of a conscious or unconscious need to minimize those perceived disadvantages. Although it is current knowledge that, as a result, most schools therefore select students based on their own criteria, unfortunately these practices are not always documented. School sampling is also a problem in jurisdictions with smaller populations because of over-sampling, which means that the same schools participate repeatedly in large-scale programs, again seen by many practitioners as causing significant disturbance.

Design

This section dealt with the format, content, and measurement qualities of the assessments. Sound decisions about policies and the allocation of resources depend on quality data. Although a significant number of respondents support most existing design parameters, they acknowledge the need for greater willingness by teachers and students to fully engage in the program in order to increase the validity of the data. Despite the respondents' understanding that the program is not suitable to offer school-based or

individual achievement results, their comments point toward the need to publish some type of school-based information, such as thematic reports, that would highlight current exemplary assessment practices. Increased teacher participation and support without added stress, time, and effort are also seen as a means for attracting teachers to further embrace the process and to encourage their students to provide optimal performances.

To that effect, the SAIP's review thus recommended the exploration of incorporating local assessment practices within the current SAIP design. This give-and-take approach is expected to further empower and meet the needs and satisfaction of participating teachers. However, this is easier said than done and the issue of motivation has been raised in other large-scale studies (Hattie, Jaeger, & Bond, 1999; Lane & Stone, 2002; Wilson, 1999). Within the SAIP's context, external motivation, i.e. students are motivated if teachers are motivated, plays an important role. Offering pizza to participating students, as several jurisdictions do, is not enough to entice students to invest wholeheartedly in the assessment. The debate around greater teacher and student involvement leads to several meaningful research questions such as: Can such a low-stakes program provide sufficient relevant information to schools to attract teachers and their students to such assessment programs? To what extent would the incorporation of innovative classroom-friendly assessment practices of socially relevant competencies entice teachers and their students to fully collaborate? What are some of the most successful practices to motivate students and their teachers to provide optimal performances within age-based sampling programs? What type of feedback would promote greater participation by the teachers and their students?

Significance

In terms of impact, the data indicate that a large proportion of respondents believe that SAIP has little or no influence on various educational contexts at a local level. In that respect, CMEC has not succeeded in implementing some of the recommendations from the first SAIP review, particularly those addressing a) better dissemination of results at local level, b) increased SAIP visibility, and c) greater ownership of SAIP's objectives by school-board administrators. This implies a definite need to develop an awareness-building plan that aims specifically at school-based stakeholders. Such an issue can be addressed through various approaches, namely: a) Elaboration of reports for different audiences, b) Publication of clear and pertinent frameworks that reflect the most recent theories, research, and practice (e.g., Campbell, Kelley, Mullis, Martin & Sainsbury, 2001; College Board, 1999; CCSSO, 1999a, 1999b; Flockton & Crooks, 1997), c) Release of actual and practice items and tasks, along with their respective scoring guides to serve a innovative practices (e.g., Robitaille, Beaton & Plomp, 2000), d) Development of a web site that is continually upgraded and maintained, e) Support for secondary data analysis studies, and f) Linkages of SAIP results with those from provincial and international assessment programs. NAEP, TIMSS, and PIRLS have been known for their sustained effort to inform their various stakeholders and audiences in a timely fashion and through a variety of reports. NAEP, for example, publishes the following: *Report cards*, *Highlights reports*, *Instructional reports*, *State reports*, *Technical reports*, *Cross-state reports*, *Trend reports*, *Focussed reports*, and *Service reports*, and its dissemination process is continually examined to improve the usefulness of these various reporting formats (Horkay, 1999). The review of SAIP's second cycle has led to the formulation of several recommendations to that effect, particularly with respect to the release of sound and clearly articulated frameworks, exemplary items and

scoring guides, and secondary analyses of relationships among contextual and achievement data.

Many of the jurisdictions that participate in SAIP also join other international large-scale assessment programs and most large populations, defined in terms of number of school boards or districts, implement their own. Eleven Canadian jurisdictions took part in the last TIMSS-R assessment and were involved in the PISA, 2000 program. The survey did not question the practitioners on SAIP's merits in relation to alternative programs because of the jurisdictions' varying degree of involvement in these. However, such a question was formally asked to each jurisdictional coordinator or representative. Their answers generally indicated that these programs provide valuable complementary data but they do not have a confirmed continuous cyclical plan and tend to provide a perspective that is more or less detached from the national educational concerns. With its planned long-term assessment cycles, SAIP can carve out an enviable and enduring place among large-scale evaluations by considering the following three mandates: a) Focussing on cross-curricular competencies that are valued by the Canadian public, b) Adopting a diagnostic function aimed at offering innovative perspectives for adjustment and intervention in educational priorities, in curricula, and in teaching, assessment and learning strategies, and c) Increasing its usefulness toward the smaller size jurisdictions or minority populations.

In conclusion, this study's objective was to report the voice of one major stakeholder: The school boards. For a variety of reasons, such as the timing of the investigation, a major overhaul of the educational system, a considerable employment turnover, and responsibility overload, fewer respondents than anticipated completed the survey questionnaire. Nevertheless, this voice is a fundamental one to which any large-scale program should pay particular attention given its frontline position. In other words, it is the one that provides the raw data. In future reviews of large-scale assessment programs, this stakeholder should therefore be consulted through methodologies offering further in-depth and relevant prompting of some of the major concerns expressed in this study. This is especially true if the instructional enhancement is as much a priority as is educational accountability (Popham, 1999). Such methodological approaches could perhaps include focus group sessions in which directors, superintendents, principals, local coordinators, board consultants, and teachers would share and confront their views regarding those aspects of the assessment programs that have the most impact on them (Haertel, 2002). As this study results show, the most pressing issues to be debated would likely include the following: a) Increase teacher and student motivation to participate wholeheartedly in the process, b) Develop effective dissemination options, c) Identify ways to ensure that the assessment program can continue to provide leadership through innovation in teaching and in assessment practices, and d) Finding cost-effective, yet rigorous means of simultaneously providing snapshot information and longitudinal means of comparisons. Although universally appealing, such approaches have yet to be supported by sound educational theory and methodology.

Finally, despite the fact that a large proportion of respondents in this study viewed SAIP as having little impact on various educational contexts and were more or less aware of its objectives, most indicated that they would recommend other school boards to participate in SAIP. It appears that for many, SAIP has effectively carved out an important place for itself among large-scale assessment programs because of its pan-Canadian nature, its capacity to involve all jurisdictions, and its ability to respect

many of the technical requirements of such initiatives (Forgette-Giroux & Simon, 2000). Given the voluntary aspect of this participation and the importance of gathering valid data, however, SAIP should invest in maintaining its leadership role by raising its instructional priority, by increasing its visibility, and by inspiring local policymakers, administrators, teachers, and students through attractive and meaningful ways.

Acknowledgement

The findings reported in this paper are part of a review funded by the Council of Ministers of Education, Canada (CMEC) and are published with the written permission of the CMEC. The opinions expressed in this paper however do not necessarily reflect the position or policies of the CMEC.

Notes

1. <http://www.CMEC.ca>

References

- Bechger, T.M., van den Wittenboer, G., Hox, J. J., & De Glopper, C. (1999). The validity of comparative educational studies. *Educational Measurement: Issues and Practice*, 12(3), 18-26.
- Boulmetis, J., & Dutwin, P. (1999). *The ABCs of evaluation: Timeless techniques for program and project managers*. Windsor Ontario: Jossey-Bass Publishers.
- Campbell, J. R., Kelly, D. L., Mullis, J. V. S., Martin, M. O., & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001*. PIRLS International Study Center, Boston College: Chestnut Hill, MA, USA.
- Chelimsky, E., & Shadish, W. R (Eds.) (1997). *Evaluation for the 21st Century. A handbook*. Thousand Oaks, CA: SAGE Publications.
- CMEC (2000). *SAIP Science Report 1999*. Toronto.
- College Board (1999). *Mathematics framework for the 1996 and 2000 national assessment of educational progress*. Washington, DC: National Center for Educational Statistics.
- Council of Chief State School Officers (1999a). *Reading framework for the national assessment of educational progress: 1992-2000*. Washington, DC: National Center for Educational Statistics. (ERIC Document Reproduction Services ED430209)
- Council of Chief State School Officers (1999b). *Science framework for the 1996 and 2000 national assessment of educational progress*. Washington, DC: National Center for Educational Statistics. (ERIC Document Reproduction Services ED431618)
- Cousins, B., & Simon, M. (1993). *A review and analysis of the thematic program, "Education and work in a changing society" of the SSHRC strategic grants program*. Final and Technical reports submitted to Social Sciences and Humanities Research

Council.

Crocker, R. K. (1997). *Study of the school achievement indicator program*. Toronto: Council of Ministers of Education, Canada.

Crooks, T. (1996). *Validity issues in state or national monitoring educational outcomes*. (ERIC Document Reproduction Services ED398285)

Crooks, T., & Flockton, L. (1999). *The design of New Zealand's national education monitoring project*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Flockton, L., & Crooks, T. (1997). *Reading & speaking assessment results 1996*. Ministry of Education, New Zealand. Dunedin: Educational Assessment Research Unit, University of Otago.

Forgette-Giroux, R., & Simon, M. (2000). *School Achievement Indicators Program - Second cycle evaluation*. Report submitted to the Council of Ministers of Education, Canada. Toronto.

Hattie, J., Jaeger, R.M., & Bond, L. (1999). Persistent Methodological Questions in Education Testing. In Asghar Iran-Nejad, & P.D. Pearson (Eds.), *Review of Research in Education*, 24, 393-446.

Haertel, E. H. (2002). Standard setting as a participatory process: Implications for validation of standards-based accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 16-22.

Horkay, N. (Ed.) (1999). *The NAEP guide*, NCES 2000-456. Washington, DC: US Department of Education. National Center for Educational Statistics.

Human Resources and Development, Canada & Council of Ministers of Education, Canada. (1999). *Memorandum of understanding*. Unpublished document.

Hunter, D., & Gambell, T. (2000). *Professionalism, professional development, and teacher participation in scoring of large-scale assessment*. Paper presented at the annual meeting of the Canadian Society for the Study of Education. Edmonton.

Joint Committee on Standards for Educational Evaluation (1999) *Standards for evaluations of educational programs, projects, and materials*, 2nd Ed. Toronto: McGraw-Hill-Ryerson.

Jones (2001). Assessing achievement versus high-stakes testing: A crucial contrast. *Educational Assessment*, 7(1), 21-28.

Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 23-30.

Linn, R. L., (2001). *The influence of external evaluations on the National Assessment of Educational Progress*. CSE Technical Report 548. Los Angeles: Center of the Study of Evaluation, National Center for the Research on Evaluation, Standards, and Student

Testing.

Macdonell, A., Forgette-Giroux, R., Schmidt, S., Mougeot, Y., & Levesque, J. (1999). *Rapport sur l'évaluation d'étape du réseau stratégique en éducation, formation et emploi*. Conseil de Recherches en Sciences Humaines du Canada, Ottawa.

Madaus, G. F., & Pullin, D. (2000). *Questions to ask when evaluating a high-stakes testing program*. Consortium of Equity in Standards and Testing. Available online: <http://www.stecp.bc.edu/CTESTWEB/documents/CTEST/NCASPress.pdf> .

Pellegrino, J. W., Jones, L., & Mitchell, K. J. (Eds.) (1999). *Grading the Nation's Report Card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy Press.

Popham, W. J. (1999). Where Large Scale Educational Assessment Is Heading and Why It Shouldn't. *Educational Measurement: Issues and Practice*, 18(3), 18-26.

Posavac , E. J., & Carey, R. G. (1997). *Program evaluation: Methods and case studies*. (5th Ed.). Upper Saddle River, New Jersey: Prentice-Hall.

Ryan, K. (2002). Assessment validation in the context of high-stakes assessment. *Educational Measurement: Issues and Practice*, 21(1), 7-15.

Robitaille, D. F., Beaton, A. E., & Plomp, T. (Eds.) (2000). *The impact of TIMSS on the teaching & learning of mathematics & science*. Vancouver: Pacific Educational Press.

Shavelson, R. J., McDonnell, L. M., & Oakes, J. (1991a). What Are Educational Indicators and Indicator Systems? *Practical Assessment, Research and Evaluation*, 2(11). Available online: <http://ericae.net/pare/getvn.asp?v=2&n=11> .

Shavelson, R. J., McDonnell, L. M., & Oakes, J. (1991b). Steps in Designing an Indicator System. *Practical Assessment, Research and Evaluation*, 2(12). Available online: <http://ericae.net/pare/getvn.asp?v=2&n=12>

Shepard, L. (1977). *A checklist for evaluating large-scale assessment programs*. Available online: <http://www.wmich.edu/evalctr/pubs/ops/ops09.html> .

Wilde, J., & Sockey, S. (1995). *Evaluation handbook*. New Mexico: EAC West New Mexico Highlands University.

Wilson, R. J. (1999). Aspects of validity in large-scale programs of student assessment. *The Alberta Journal of Educational Research*, XLV(4), 333-343.

About the Authors

Marielle Simon is currently Associate Professor at the Faculty of Education, University of Ottawa where she teaches courses in research methods, assessment, measurement and evaluation. She specializes in classroom and large-scale assessment, with particular focus on portfolio assessment and reporting.

Email: msimon@uottawa.ca

Renée Forgette-Giroux is Professor with the Faculty of Education, University of Ottawa. Her studies also focus on classroom and large-scale assessment. She has published on portfolio assessment and grading. She teaches courses in educational research, statistics and assessment.

Copyright 2002 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, [Gene V Glass](mailto:gene.v.glass@asu.edu), glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

[Michael W. Apple](#)
University of Wisconsin

[John Covalesskie](#)
Northern Michigan University

[Sherman Dorn](#)
University of South Florida

[Richard Garlikov](#)
hmwkhelp@scott.net

[Alison I. Griffith](#)
York University

[Ernest R. House](#)
University of Colorado

[Craig B. Howley](#)
Appalachia Educational Laboratory

[Daniel Kallós](#)
Umeå University

[Thomas Mauhs-Pugh](#)
Green Mountain College

[William McInerney](#)
Purdue University

[Les McLean](#)
University of Toronto

[Anne L. Pemberton](#)
apembert@pen.k12.va.us

[Richard C. Richardson](#)
New York University

[Dennis Sayers](#)
California State University—Stanislaus

[Greg Camilli](#)
Rutgers University

[Alan Davis](#)
University of Colorado, Denver

[Mark E. Fetler](#)
California Commission on Teacher Credentialing

[Thomas F. Green](#)
Syracuse University

[Arlen Gullickson](#)
Western Michigan University

[Aimee Howley](#)
Ohio University

[William Hunter](#)
University of Ontario Institute of Technology

[Benjamin Levin](#)
University of Manitoba

[Dewayne Matthews](#)
Education Commission of the States

[Mary McKeown-Moak](#)
MGT of America (Austin, TX)

[Susan Bobbitt Nolen](#)
University of Washington

[Hugh G. Petrie](#)
SUNY Buffalo

[Anthony G. Rud Jr.](#)
Purdue University

[Jay D. Scribner](#)
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia
Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación
Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de
México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de
México
humberto@servidor.unam.mx

Daniel Schugurensky
(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de
México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do
Sul-UFRGS
luceb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch—Brazil
(AIRBrasil)
simon@airbrasil.org.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu