

THE ADVISORY GROUP ON PROVINCIAL ASSESSMENT

FINAL REPORT

KRIS MAGNUSSON **SIMON FRASER UNIVERSITY**
BLYE FRANK **UNIVERSITY OF BRITISH COLUMBIA**

JUNE 2014

EXECUTIVE SUMMARY

The Ministry of Education, as part of its new Education Plan, is working toward improved student assessment and reporting to help government and educators build on the strengths of the existing education system. As part of this initiative, the Ministry convened the Advisory Group on Provincial Assessment (AGPA) in March, 2013. The group consisted of representatives from stakeholder groups in education in British Columbia, together with two experts in student assessment and two education deans as co-chairs. The AGPA met four times from April to June, 2013, and again three times from February to March, 2014, to develop advice and recommendations to the Ministry of Education on BC's future provincial assessment system. Each stakeholder group had the opportunity to share its perspective on BC's current program of assessment as well as its hopes and wishes for future assessment programs. The discussions were frank and respectful.

Over the course of the three meetings held in 2014, AGPA members worked collaboratively in the development of a list of the purposes of assessment. Additionally, the group was able to develop an agreed-upon set of principles that should inform a new provincial assessment program. At the final meeting, the group discussed a proposal for a framework model of computerized adaptive testing that can be used in provincial assessments. All AGPA members agree that this framework model holds great promise for answering the assessment needs of all stakeholder groups, while conforming closely to the agreed-upon principles. AGPA members respectfully request that the Advisory Group continue to meet in order contribute to discussions around development and implementation of the framework.

The Advisory Group on Provincial Assessment thanks the Ministry of Education for providing this opportunity to meet and discuss this very important topic.

PREAMBLE

Large-scale assessment in general and the Foundation Skills Assessment program in particular often serve a diverse array of purposes. That is certainly the case in BC, where different stakeholder groups hold different perspectives on the purposes of a provincial assessment program. They have also made use of assessment data for a variety of different goals. While these goals vary considerably, they are all legitimate and important issues within the K-12 BC educational system.

However, it is also true that many of these goals are not consistent with the original purpose or capability of the testing programs in place. One of the significant failings of the current Foundation Skills Assessment (FSA) program is that the results are used to make judgments that go beyond its mandate. The Advisory Group on Provincial Assessment acknowledges that any assessment tool will have limitations.

The Advisory Group on Provincial Assessment engaged in seven meetings, during which members discussed the purposes, uses, and needs for assessment data. The highly complex nature of a provincial assessment program was discussed in detail in the context of the current shift toward the new Education Plan. The AGPA understands that the guiding framework of the Plan involves cross-curricular competencies (such as thinking, communication and personal/social competencies), learning standards (curricular competencies and content), and enduring understandings (social, cultural and traditional values). The challenge of devising a provincial assessment program that is based on competencies, standards, and understandings rather than on content spurred many thoughtful conversations.

The group noted that its efforts were somewhat hampered by a lack of information about the design and content of the new curriculum. Overriding all these discussions was an awareness of the critical function of large-scale assessments in helping us understand how the education system is fulfilling its core purpose. Equally important was the awareness that data from past provincial assessments have been misinterpreted. For these reasons, any new provincial assessment system should promote the appropriate and ethical use of the data it generates.

By the end of the seventh meeting, the respectful and cooperative dialogue among Advisory Group members allowed the group to move toward consensus on a proposed framework of provincial assessments, which is outlined in this document. It is based on an agreed-upon set of principles, and has a clearly defined set of purposes. While no single assessment tool can answer all purposes, the framework that is proposed is flexible enough and detailed enough to provide important data that will inform and guide decision makers at all levels.

The Advisory Group on Provincial Assessment respectfully requests that the Ministry empower this group of committed stakeholders to continue meeting on a regular basis. This Advisory Group provides informed guidance and advice to support and improve public education in British Columbia. Because the Advisory Group represents all the major stakeholders in public education in BC, the Ministry can utilize this group to generate ideas, to provide feedback on assessment and evaluation-related Ministry initiatives, and to serve as a conduit for information to stakeholders. . This group should not function on an ad hoc basis, but should constitute a regular part of the evaluation process.

PROCESS SUMMARY

The Ministry of Education convened the Advisory Group on Provincial Assessment (AGPA) in April 2013. The mandate given to AGPA was to provide a broad set of recommendations on provincial assessments, based on the perspectives and issues faced by the stakeholder groups. The AGPA met for a series of four meetings during April, May, and June 2013. The group submitted a preliminary report to the Ministry in September 2013. The Ministry reconvened the AGPA for a series of three meetings in February and March 2014, with a deadline of April 30th for the submission of a report and recommendations. The purpose of the second series of meeting was to address the following questions:

Given the shifting emphasis in provincial curriculum and the overall transformation directions:

- *What should the purpose(s) of a provincial assessment program be in the elementary/middle grades?*
- *What areas should be assessed? At what grade levels? (note: the areas measured could be different at different grade levels.)*

The first meeting of the second series took place on February 25th, 2014. AGPA members discussed changes within their respective organizations since the previous meeting. The conversation then turned to the tensions that arise when considering how to conduct an assessment of cross-curricular competencies. Other tensions emerge when considering classroom-based versus large-scale assessments. Finally, the group discussed the purposes of provincial assessments and what the potential impact of such assessments might be.

AGPA members met for a second meeting on March 5th, 2014. The meeting took the form of a group writing session that addressed the purposes of provincial assessments, with a focus on the varied purposes of assessment at the provincial, district, school, and individual levels. All AGPA members agreed that the primary purpose for a provincial assessment program should be to determine educational priorities and needs and to

subsequently allocate resources to meet those priorities. In this way, committee members believed that assessment data could lead to system improvement, not just system description.

At the third and final meeting, held March 20th, 2014, AGPA members discussed the transformation of curriculum and assessment in BC. An important point of agreement for all AGPA members was the notion that assessment needs to be meaningful for learners, and it needs to be linked to the curriculum. At the same time, assessment needs to allow for multidisciplinary and personalized learning. During the second half of the meeting, members discussed a new model for provincial assessment that would align with AGPA's stated principles and that would provide system-wide information. At the close of the meeting, the co-chairs agreed to develop the model more fully in a written document, and circulate this to AGPA members for consideration. The final document, incorporating all stakeholder feedback, was prepared for delivery to the Ministry on April 30th, 2014.

PURPOSES OF PROVINCIAL ASSESSMENT

There are two broad and often competing purposes for large-scale assessments. First, there is a need to better understand how the British Columbia educational system is meeting the general learning needs of its students, so that the system is accountable to society in general and to parents and other stakeholders in the public education system. This system view, in essence, answers the broad question, "How well are we doing at educating our children?" Second, there is also a need to track and monitor individual students. Many students are already disadvantaged in terms of their home environment, and ensuring that this information is available to the school is a worthwhile goal. Tracking information over time about students in care, for example, might help educators and caregivers identify strategies and supports that boost graduation rates. The Advisory Group recognizes that the current provincial assessment system cannot support all of these purposes in a meaningful way. The new framework outlined later in this Report, answers all of the purposes outlined here.

The Advisory Group on Provincial Assessment proposes the following desired purposes of provincial assessment:

1. To inform the provincial education system by:
 - 1.1. Providing a broad description of system functioning;
 - 1.2. Identifying system strengths, weaknesses and gaps in students learning;
 - 1.3. Describing the performance of subsets of the school population, including Aboriginal students, students with special needs, children in care, and other groups as identified by stakeholders;
 - 1.4. Highlighting the impact of curricular, structural or other policy changes; and
 - 1.5. Providing a framework for longitudinal data.
2. To support informed decision-making in relation to:
 - 2.1. Public policy and priorities about K-12 education;
 - 2.2. Planning, curriculum development, and interventions;
 - 2.3. The allocation of resources to meet specific needs that are identified or confirmed through the data; and
 - 2.4. Priorities and support for educational research.
3. To support districts and schools by:
 - 3.1. Situating district results in relation to provincial results; and
 - 3.2. Enabling appropriate comparisons to be made over time within jurisdictions.

STAKEHOLDER PERSPECTIVES

Over the course of the seven AGPA meetings, several themes emerged and re-emerged as we discussed a new model for provincial assessments. These included discussions of what constitutes quality public education, challenges around measuring system strength and accountability, and how to ensure that the data collected are being used to target resources where they are urgently needed.

All stakeholder groups are concerned that assessment within the educational system be performed in an ethical, equitable, and consistent manner. The misuse of data needs to be avoided. A core value for all stakeholders is to ensure that assessment, and resourcing that derive from the data gathered, are used to foster equity and social justice within the public education system. A literacy and numeracy secretariat, for example, could be tasked with making recommendations to the Ministry based on the data collected through a provincial assessment.

A fundamental concern is how to ensure that the education system is held accountable and adequately resourced for groups or subsets of the population who are either at risk, or who have not flourished within the system in the past. For example, Aboriginal students' achievement scores continue to lag behind the norm, and their school completion rates remain much lower than the norm. If we do not have a way to document such achievement gaps, then we have no way of ensuring that the system addresses issues that are leading to the gaps.

A matter of concern for many stakeholders is how to link assessment to a competency-based curriculum: how do we separate content from competencies? As discussions focused on province-wide assessments, it was noted that since content will vary from classroom to classroom, devising a valid assessment becomes even more challenging. Whatever direction assessment goes in, it will have to reflect personalized learning: whatever appears on a provincial assessment has to relate to what has been taught in the classroom, or the assessment runs the risk of harming learners. As the educational system moves away from "knowing things" toward "understanding things," then it becomes more difficult to determine what should appear on an examination. The Advisory Group recognized that assessing student understanding of concepts within a discipline requires students to have a common appreciation of essential content within that discipline. Of course, this matter is more relevant to assessment in higher grades, and less pertinent to the current assessment practices in grades 4 and 7.

The public education system is charged with, among other important purposes, preparing young people for successful participation in society. Graduates are expected to be literate, numerate, and socially responsible. One of the purposes of a provincial assessment program is to provide guarantees to society that the public education system is indeed graduating literate, numerate, and socially responsible citizens in a democratic society. Equally, there is a need for system-level data that provides information on how well the education system is meeting the needs of specific subgroups, such as Aboriginal children, children in care, and other marginalized groups. The ability to share information from provincial assessments with various agencies and groups will help ensure that all children are receiving a quality education regardless of their life circumstances. The information can be used to provide the supports necessary to meet the child's learning needs, especially for those in at-risk population groups. It is important to note that information collected at the provincial level has the power to influence important policy shifts. The best possible information, therefore, is needed in order to ensure the best possible policy decisions.

Many of the tensions and frustrations around the current provincial assessment stem from a lack of classroom- or school-level control over the timing of assessments. Provincial assessments are conducted on specified dates and in prescribed ways. Little or no flexibility is offered to the classroom teacher in when or how the provincial assessment is administered. Results from the provincial assessment are delivered too late in the school year to be useful. Groups such as the Fraser Institute misinterpret and publicize results in ways that are damaging to classrooms and schools, and therefore damaging to learners. AGPA stakeholders are recommending modes of assessment that allow early intervention; that allow teachers to determine student readiness within a specified window; and that ensures that results are used appropriately. While it seems desirable that the maximum number of students possible participate in a provincial assessment program, it is nevertheless the case that some students may need access to flexible arrangements for participation. Among these might be students who have significant cognitive or emotional impairment, or who are physically dependent. In addition to provincial, district, and school level data, several stakeholders outlined a need for the assessment and tracking of individual students. An Information and tracking system that could identify, monitor, and contrast individual performance against aggregated norms (e.g., class, school, district, or provincial) would be one way of addressing this need. It would also help in the identification of individual planning, development and learning needs, and could inform the allocation of specific resources to meet individual needs.

The assessment needs to be able to show improvement over time. If we are moving toward personalized learning, then it follows that there will be many pathways toward meeting learning outcomes. The assessment system needs to show the starting point for each student as well as tracking improvement over time.

PRINCIPLES OF A PROVINCIAL ASSESSMENT PROGRAM

Regardless of the mode of administration (paper-and-pencil test or web-based test), the guiding principles ought to remain the same. In other words, the values that are expressed in the following principles need to be embedded in whatever assessment tool is used. The AGPA proposes the following principles to guide the development of a new provincial assessment program.

1. The system of education and the accompanying modes of provincial assessment must be based on an ethical foundation, in which social justice shares equal importance with system quality and excellence as the primary outcomes. While there are many definitions and conceptions of “social justice”, we use the term to mean access and opportunity for all members of society to “equity, security and safety, sustainability of communities, participation of citizens in social change, and access to opportunities for personal growth and development” (BCTF, <http://www.bctf.ca/socialjustice.aspx>).
2. We must recognize the diverse needs and purposes of stakeholder groups within the educational system. These diverse needs imply a need for diverse strategies for meeting needs.
3. Assessment comes in different forms and is used for a variety of purposes. There is a need for relevant and reliable data that can be used to inform how well the system is doing. Equally, there is a need for information on how well individuals are doing. Professional educators are the primary source of information on the performance of individual learners.
4. We must attend to all within the system, and not just the majority or the advantaged. Safeguards need to be in place to ensure that at-risk populations or individuals have access to quality education and sufficient

resources. Given the diversity of needs of learners within the system, we may need to consider additional or alternative:

- 4.1. Reporting mechanisms, particularly for learners at the elementary level;
 - 4.2. Tracking mechanisms, to better follow individuals and groups of individuals through and within the system; and
 - 4.3. Means of professional interactions (e.g., linking educators with those in social and justice systems).
5. We must consciously and intentionally link resource allocation to identified gaps or system discrepancies. It is not enough to simply track at-risk groups; the data must be used to inform priorities and to allocate budgets and other resources.

FRAMEWORK MODEL

The Advisory Group recommends a model of provincial assessment that connects to each of the guiding principles described above. The proposed model is ethical, and system quality and excellence will be the primary goals. The model can be adapted to meet the needs of the various stakeholder groups. It will provide data at all levels (system, district, school, and individual) and will provide means of reporting, tracking, and information sharing to ensure that at-risk students and populations are well served. The final principle, that of linking resources with data, is beyond the scope of an assessment tool but nevertheless remains a principle that needs to be attended to.

The Advisory Group proposes that the Ministry invest in a computerized adaptive testing (CAT) system for the administration of provincial assessments that will address both system and individual needs. Such a system could capture data that reflects achievement related to the core functional skills including reading comprehension, writing, and numeracy. The system could also capture individual data including achievement, learning strategies, learning attitudes, learning engagement, and learning gains. Once personal identifiers are removed, data could be aggregated at several levels.

Implementing CAT would entail the development of a large bank of items that reflect the curricula at various grade levels, providing grade-equivalent items from Kindergarten to at least Grade 10 in order to address the range of abilities in individual classrooms. This provincial assessment would be administered twice in the elementary years. Children would sit at a computer to take the test. As the child works his or her way through the items, the program adapts to the child, selecting easier or more difficult items for the child based on the responses given. Each child participating in the test will feel successful (“I answered all those questions”) rather than frustrated (“I didn’t get most of those questions”). Such targeted data would provide a powerful additional source of evidence for teachers to identify areas in which students need additional support.

With CAT, teachers are able to determine the timing of the test within a specified window of several months’ duration. Teachers may use their discretion to make the decision that is best for the child. Often, there are emotional or contextual factors that inhibit student performance on large-scale assessments. While the assessment may still provide a reasonable picture of the system as a whole, it may cause damage for individual students who are not ready for such a test. Students who are deemed ready may take the test early on in the window, while other students may receive several weeks of additional preparation before they take the test. This model reduces test-day stress in students, as the examination becomes just another of the many assessments children are administered in the classroom. Because tests are administered and data collected over a period of months, there is no “event” around the test, and therefore less opportunity for misuse of the data.

Several of the stakeholders groups in AGPA spoke of the need for tracking of individual students and at-risk groups. While the Advisory Group did not come to a formal agreement on whether this should be done, it nevertheless suggests that this is an area that warrants further investigation and discussion with stakeholder groups. For example, the Ministry might consider using the computerized system that is created for provincial assessments as a large-scale student information system. Ideally, the system would create a student history, including tags that indicate special considerations (such as an IEP) and that would follow a student through his or her time in the public education system. In this way, student records for transient students and students in care would immediately be available to new schools. Provision could be made to allow the sharing of appropriate information with other agencies, such as social services. Data from the system could be aggregated at any time, across grade levels. Comparisons by tag to population data could be made to allow tracking of particular groups, such as Aboriginal students or students in care. Finally, such a model also would also provide a mechanism for tracking individual change across time, and would help to ensure that at-risk individuals are carefully attended to. The creation of a student information system would need to be carefully and ethically planned. The Advisory Group would welcome the opportunity to help shape this system, as more information about the new curriculum, the competencies, and the graduation requirements becomes available.

THE STRUCTURE AND TIMING OF ASSESSMENTS

There was general agreement on the necessity for collecting system data on the core skills of reading comprehension, writing, and numeracy, and the AGPA sees no need to change these areas of focus. However, there will also be a need to prepare teachers to assess student progress on the cross-curricular competencies (i.e., thinking, communication and personal/social competencies), learning standards (curricular competencies and content), and enduring understandings (social, cultural and traditional values) that are integral to the new BC Education curriculum model. The “provincial assessment system” must include attention to both the core skills of reading, writing, and numeracy and the assessment and reporting of the new curricular competencies. In fact, an expanded model of system assessment as proposed here would serve as a useful evaluation of the impact of the new curriculum; gains or losses in core skills as a result of the new curriculum could be tracked. However, implementing one without the other will be problematic. The attention of the Advisory Group has primarily been on the large-scale system assessment, but AGPA members also do not want the local assessment issues to be lost.

There was less agreement about the timing and structure of the new provincial assessment model. The current system, in which data is collected twice in the early and middle school years (i.e., Grade 4 and Grade 7), provides useful marker points to evaluate the development of core skills across a system. Testing too early (e.g., within the primary years) does not provide sufficient time for developmental interventions to take root. Testing too late reduces opportunities to respond to individual and system needs. We recommend maintaining a program of provincial testing within Grades 4 and 7. However, we also recommend that teachers have discretion to test students within a window of at least several weeks to allow for individual differences and circumstances. Virtually all students would be tested, thus providing valuable data on system effectiveness.

If the CAT model is adopted, we may be able to consider both formal and informal test intervals. The size of the item pool within the test bank would determine the number of discrete “samples” that could be reliably drawn from the pool. For example, in addition to the required participation in the window for each of the formal (Grade 4 and Grade 7) assessments, teachers could administer additional testing sessions to gauge student competence. Given a sufficient size test bank, teachers could conduct an informal assessment of any individual, identify specific learning gaps, and introduce interventions to address those gaps. Rather than “teaching to the test”,

these strategies would enable teachers to calibrate their instruction to meet learning needs. The net impact would be to improve individual performance, and as a result, improve overall system performance.

IMPLICATIONS OF THE FRAMEWORK MODEL

Planning, Implementation, and Communication

1. Planning and implementing the new model will require funding and ongoing collaboration from assessment experts, educational researchers, educators, and other stakeholders.
2. The current system involves an annual outlay for test preparation, administration, scoring, and dissemination. The proposed system would see much higher front-end costs, and much reduced costs for annual administration, especially if web-based testing and data management systems were employed.
3. There would be a need for a data base administrator to manage the volume of data and reporting structures.
4. It is important that the Ministry sponsor the creation of a provincial assessment framework that ties to learning standards. If we do not have a framework of learning standards or outcomes, then we have nothing by which to measure the assessment tool.
5. The creation of the item banks will require expertise, planning, and input from all levels. Assessment experts will be needed to create an item bank that is extensive, reliable and free of cultural and other biases. Teachers will need unlimited access to the items in order to gauge individual student growth and mastery. At the same time some items or clusters of items will need to be specified as mandatory for system purposes. Maintaining the vitality of the item bank would require regular input.
6. There will need to be a mechanism for the ethical reporting and aggregation of the data.
7. The Ministry of Education will need to provide targeted funding and release time for in-service programs that will help professional educators understand and work with the new assessments.
8. Information about achievement in the core functional skills and competencies can be collected in a provincial assessment. However, the domains of learning strategies, attitudes, level of engagement and learning gains are highly complex. Collecting data in these domains will require expertise, planning, and input to ensure that the domains are not reduced to narrow, easily surveyed types of queries.
9. Communication about the provincial assessment program with parents and other stakeholder groups must be planned for.

CONCLUSION

The Advisory Group sees the value of system-wide assessment, but only if the abuses of the data can be curtailed, and if the data is actually used to:

- inform decision-making around policy and priorities in the public education system;
- influence planning, curriculum development, and interventions;
- influence the allocation of resources to meet the needs that are identified through the data; and
- develop priorities and support for educational research.

Regardless of the decisions that are ultimately made about a provincial assessment system, the Advisory Group also sees great value in maintaining stakeholder discussions about the linkage between assessment and system

improvement. We strongly encourage the continuation of this dialogue as an important strategy for the K-12 system in BC.

APPENDIX A: ASSESSMENT SYSTEM

a. Summative Assessment

- Aligned to provincial learning standards
- Administered during the last 8 weeks of the school year
- Accurately describes student achievement;
- Provides valid, reliable, and fair measures of students' progress toward, and attainment of the knowledge and skills required for post-secondary/career transitions
- Computer based
- Adaptive
- Include performance tasks and selected responses;
- Secure

b. Optional Interim/District Assessments

- Aligned to provincial learning standards
- Administered at locally determined intervals.
- Available to all teachers
- Provides educators with actionable information about student progress throughout the year.
- Computer adaptive
- Includes performance tasks
- Helps teachers, students, and parents understand whether students are on track
- Not secure; Available through My Education BC
- Identify strengths and limitations in relation to the Provincial Learning Standards;
- Teacher scored;
- Used when students are ready

c. Formative Assessment

- Repository of professional development materials, resources, and tools aligned to the Learning Standards
- Learning objects (Videos, lessons, webinars etc.)
- Research-based instructional tools available on-demand to help teachers address learning challenges and differentiate instruction. (i.e. scoring rubrics)
- Professional development materials related to all components of the summative assessment system.

APPENDIX B: PROPOSED MODEL FOR PROVINCIAL ASSESSMENT IN BC

Proposed Model for Provincial Assessment in BC

