

Assessment and Accountability in Public Education:

a discussion starter for the use by the
Education Advisory Council in British Columbia

prepared and submitted
by the
BC School Superintendents Association

Questions may be directed to
Keven Elder, President, BCSSA

This paper has been prepared on behalf of the BC School Superintendents' Association as a contribution to policy discussions of the provincial Education Advisory Council (EAC). It does not represent the official view of the Superintendents' Association, but rather the perspectives and opinions of the authors supported by input from the Executive. The paper considers broad policy issues appropriate to the EAC forum and does not discuss the details of practice. It is intentionally brief, providing only the minimal scaffold necessary for a constructive dialogue by the Council. Some useful references are provided in the footnotes.

Since this paper is intended to stimulate and support a dialogue, it begins with some thoughts on what that means. Next, it introduces some key terms to be explicit about how those terms are being used in the paper and to minimize the danger of differing interpretations. Frameworks for considering the multiple purposes and audiences for assessment and accountability are then introduced, and some observations are made with reference to them.

The paper concludes by proposing some questions about assessment and accountability with the intention of stimulating a dialogue that will lead to common understandings of issues, identification of areas of agreement and clarification around areas of disagreement or concern.

Overview of the Paper

This paper is organized under the following nine headings:

- Dialogue Defined
- Functional Concepts and Terms
- Multiple Purposes and Audiences
- Responsibility and Accountability
- Classroom Assessment, Evaluation and Reporting
- Common Standards and Common Assessments
- Foundation Skills Assessment
- Conclusion
- Questions for Discussion

Dialogue Defined

Common understandings about assessment and accountability are essential to constructive engagement with the many public and professional issues that surround these vital topics, and constructive engagement is essential to strong, effective public education. Dialogue is the best way to develop common understandings.

A dialogue is different from a mere discussion and radically different from a debate.¹ Isaacs, for example, describes a dialogue as follows.

Generally, we think of dialogue as “better conversation.” But there is much more to it. *Dialogue*, as I define it, is a *conversation with a center, not sides*. It is a way of taking the energy of our differences and channeling it toward something that has never been created before. It lifts us out of polarization and into a greater common sense, and is thereby a means for accessing the intelligence and coordinated power of people and groups ... The aim of negotiation is to reach agreement among parties who differ. The intention of dialogue is to reach new understanding and, in doing so, to form a totally new basis from which to think and act. (Isaacs, p. 19)

The differences between a dialogue and a debate are summarized in the following table.

Dialogue is:

Assuming that there are many valid, if partial, views and that together a group can develop a better, although still imperfect, view than anyone alone.

Attempting to broaden your own understanding by listening to appreciate others' perspectives.

Exploring common ground.

Revealing your assumptions.

Seeking consensus while keeping questions open for further consideration in the future.

Debate is:

Assuming there is one “true” and complete way of looking at things that can be discovered through the rational contest of ideas.

Attempting to find “the truth” by listening for flaws in others' arguments.

Winning your point.

Defending your assumptions.

Seeking closure through an external adjudicator or vote.

¹ See, for example, *The Magic of Dialogue* by Yankelovich, *Dialogue and the Art of Thinking Together* by Isaacs, and the original work on this topic, which is *On Dialogue* by Bohm.

Dialogue is rare, difficult and time consuming. It seems to run against past experience and habits, and it is probably not fully achievable, particularly when first attempted. However, with a genuine common commitment to try, it is often possible to move out of debate mode and into an environment that is more hospitable to differing views. This leads to common understandings and that is what this paper is intended to encourage as a means toward resolving differences and thus strengthening public education. It is understood that any movement on the provincial policy front is a larger task which will require other processes and supports that go beyond the Council's mandate.

Foundational Concepts and Terms

Dialogue benefits from careful use of terms and common understanding of those terms. Therefore, we begin with a discussion of some key ones that are used in this paper. Others may use these terms in slightly different ways. The purpose here is simply to be transparent.²

Assessment is an ongoing process of gathering information about another's or one's own learning. It is gathered intentionally and strategically by teachers, sometimes through tests or assignments, but often also incidentally through interaction with students in the normal course of instruction. Assessment information is intended only to describe learning.

Evaluation refers to judgments made on the basis of assessment evidence. This evidence often includes marks for assignments and tests, but it should not be reduced to a calculation on the basis of those marks. There is an essential element of professional judgment. Evaluation information may be provided anecdotally and/or in the form of a letter grade.

The terms "assessment" and "evaluation" are often used as if they were synonyms, but they are not, and this can lead to confusion.

Standards are the expectations against which evaluative judgments are made. Both content and quality standards may be used. The Prescribed Learning Outcomes in Integrated Resource Packages provide content standards. Quality standards are generally defined by teachers on the basis of their professional judgment. Performance Standards for Reading, Writing, Numeracy, Social Responsibility and Information and Communications Technology Integration have been developed for voluntary use in BC schools. Both the content and quality standards being used by teachers should be known to students and parents. This will help them to understand the teacher's evaluation as well as assist them on their own.

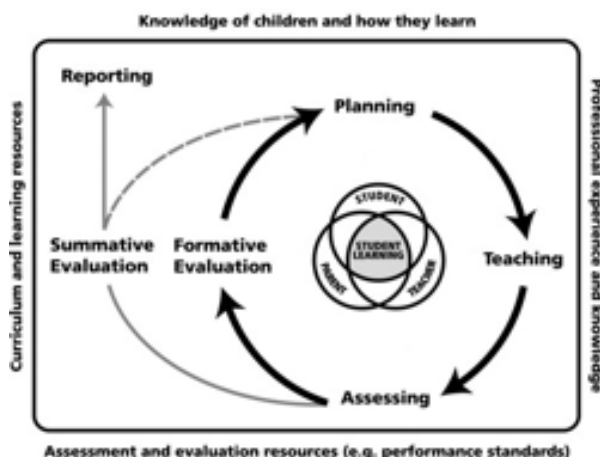
Indicators are the products and performances teachers use to enable students to demonstrate their learning. From these products and performances teachers infer what has been learned and how. There is no entirely objective or reliable way to determine what a person knows, and certainly not one that is equally valid for all kinds of learners. Therefore, a wide range of indicators should be used to give the student the best opportunity to demonstrate learning and to increase the likelihood that the teacher will make accurate inferences from them.

Formative Evaluation is usually conducted while learning is in progress to provide information about what has been learned to the student and teacher that will assist them in monitoring and adjusting their performance.

Summative Evaluation is usually conducted at the end of a unit of study in order to draw conclusions about the quality of student achievement relative to the intended learning outcome.

² Chapter 8 of the Primary Program provides an excellent discussion of assessment, evaluation and reporting. It can be found at http://www.bced.gov.bc.ca/primary_program/primary_prog.pdf.

Formative evaluation is intended to provide feedback to the student and the teacher as learning occurs in order to help both adjust what they are doing. Summative evaluation is intended for reporting on the end results of that learning in order to demonstrate the level of achievement. The relationship is illustrated in the accompanying diagram, which is adapted from *The Primary Program* (RB0109, p. 161). In some cases, a single assessment may be used to provide both formative feedback and summative evidence.



Criterion-Referenced Evaluation is a judgment about learning that is made with reference to pre-defined content and quality standards. In a criterion-referenced evaluation, one student's results are entirely independent of any other student's results. This approach focuses on the learning demonstrated by the student in comparison to what it was intended s/he would learn.

Comparative Evaluation is a judgment about learning made by comparing the achievement of students to each other. Provincial policy in BC is that student evaluation is to be "in relation to expected learning outcomes set out in the curriculum;" that is, evaluation is to be criterion-referenced, not comparative.³

Norm-Referenced Evaluation is a judgment about learning that is made with reference to norms that have been established through prior assessments of a population that is similar to the student for whom the evaluation is being conducted. This approach focuses on the learning demonstrated by the student in comparison to what is typical of similar students.

Performance Evaluation is a judgment made on the basis of complete, complex realistic tasks. It is sometimes referred to as "authentic" evaluation. Assessments and evaluations based on authentic performance tasks provide the most valid, reliable and useful information.

Standardized Assessment is the process of gathering information using the same indicators for all students. *Standardized Evaluation* is a process of applying a standard method of valuing indications of learning obtained through a standardized assessment. Summative evaluations, particularly those based on tests, are generally standardized to the classroom level, sometimes to the school level, and sometimes to the district or provincial level. Standardized evaluations may be either criterion-referenced or norm-referenced.

Multiple Purposes and Audiences

Judgments about the learning that occurs in schools are made by many people for many reasons, including providing feedback to students on their learning, providing feedback to parents regarding progress of children, reporting achievement for certification, providing insights into system performance and providing for public accountability. The kinds of evidence, standards and indicators that will lead to sound judgments vary according to the purpose.

³ See Student Progress Report Order 191/94 at <http://www.bced.gov.bc.ca/legislation/schoollaw/e/m191-94.pdf>.

The multiple purposes and audiences for assessment and evaluation information might be thought of in terms of the following grid.

	Student	Staff	Parent	System	Public
Formative					
Summative					

In a classroom, students and teachers (and increasingly paraprofessionals working with the teacher) gather evidence and make judgment about learning on a continuous basis. From time to time this may be formalized as an assignment or a test, but formal data gathering should not be confused with the much richer and more influential process of assessment and evaluation that occurs within every lesson and interaction, both consciously and subconsciously.

In order for this process to be effective, both content and quality standards should be known in advance. This enables students to understand the teacher’s assessment. It helps teachers to focus clearly on intended outcomes in their assessment so that they make valid judgments and provide useful feedback. It also permits parents to better understand what their children are learning, how they are doing and how they can help.

Classroom assessment practice is a complex and extensive topic that will not be discussed in this policy-focused paper. However, the interested reader is referred to the many excellent resources available, including those provided by the Ministry of Education.⁴ In general terms, the trend in education has been from an exclusive focus on summative evaluation to a greater emphasis on formative assessment and more immediate feedback to students.

An important consideration in classroom assessment and evaluation is professional autonomy. Although there are definable elements to *effective* practice, there is no single standard for *best* practice. Teachers in an inclusive school system must have a sophisticated understanding of effective practices in order to adapt their practices as necessary to enable the particular students in their care to succeed. The autonomy they are afforded to do so comes with a responsibility for continuous learning about effective assessment, evaluation and instructional practices.⁵

There is also an important role for assessment and evaluation beyond the classroom. The Ministry of Education has a clear responsibility to monitor student achievement and system performance, and the general public has a legitimate interest in the success of this essential public service. Although students, teachers and parents are not responsible for these system-level processes, they should have a role in their design and in determining how the results will be used because of the knowledge they can contribute, and because of the implications of the results for them. As with classroom assessment, the indicators and standards should be carefully selected and publicly explained in advance.

⁴ Classroom assessment is discussed at http://www.bced.gov.bc.ca/classroom_assessment/. See also *Rethinking Classroom Assessment With Purpose in Mind* at <http://www.wncp.ca/media/40539/rethink.pdf>.

⁵ Professional Autonomy is also constrained by an employment duty to accept direction from administrative supervisors and to participate in system-wide assessments.

In 1993, a broad-based committee of stakeholders, educators and the provincial and territorial ministries, developed a statement of *Principles for Fair Student Assessment Practices for Education in Canada*.⁶ This continues to be a useful reference for both classroom and system practice.

Responsibility and Accountability

According to the New Oxford American Dictionary, *responsibility* is “a moral obligation to behave correctly toward or in respect of [someone or some obligation] and *accountability* means being “required or expected to justify actions or decisions.” Generally, one is responsible prior to an action and accountable afterwards, but the notion of obligation applies throughout.

The *Mandate for the School System* identifies three goals, all of them shared between schools, families and communities.⁷ Intellectual Development is the primary responsibility of schools, with the support of families and communities. Human/Social Development and Career Development are shared responsibilities. The Mandate also describes various responsibilities for students, parents, staff, school boards, the community, the College of Teachers and the Ministry of Education. Clearly, public education is undertaken in partnership and involves shared obligations.

Thus, one might ask, “Who is responsible to whom and for what?” The following grid provides a framework for considering this question.

		Student	Staff	Parent	System	Public
Responsible to ...	Student					
	Staff					
	Parent					
	System					
	Public					

For example, with reference to the first column, in what ways is a student responsible to school staff, parents, the school system and the public? With reference to the last column, in what ways is the public responsible to students, school staff, parents and the school system? Not all of the sectors in this grid are equally relevant or important but all raise valid questions. Considering the grid as a whole can help to balance the discussion of responsibility and accountability in the school system.

Some obligations relate to tasks while others relate to conduct. For example, the School Act defines the legal duties of teachers while the College of Teachers defines standards of ethical conduct. Other obligations, such as those implied by “professionalism” are unwritten but significant. Thus, teachers are afforded the autonomy and respect of professionals in exchange for acceptance of the obligations of a profession, which include mastering a well-defined body of expertise, being able to make independent decisions in a complex context, holding members of the profession to a high standard of competence and conduct, and committing to ongoing learning and skill development.

⁶ See http://www.bced.gov.bc.ca/classroom_assessment/fairstudent.pdf.

⁷ See http://www.bced.gov.bc.ca/legislation/schoollaw/d/oic_1280-89.pdf.

Responsibility in a partnership is always reciprocal. Students are responsible for making the most of the opportunities afforded to them and for conducting themselves in a manner that contributes to a safe, supportive learning community. In return staff are responsible for adapting the educational program to accommodate their needs and abilities, parents are responsible for providing love and support, the system is responsible for providing appropriate facilities and resources, and the public is responsible for supporting the school system and the families.

Accountability requires a bias for transparency and a willingness to explain one's intentions, decisions and actions. Sometimes, accountability requires acceptance of consequences for inappropriate intentions, poor decisions or incorrect actions. Generally, however, accountability is not punitive and the greater good is achieved in an appreciative environment that provides the safety and support that encourages everyone involved in public education to admit error, continuously seek to improve and learn from experience.

The standards-based reform movement found in some countries is based on the assumption that one must create an external motivation for educators to improve. A community partnership approach to educational accountability assumes that educators have an internal motivation to improve based on their moral and professional commitment. This approach is reciprocal and has the potential to unleash the enthusiasm and creativity of not only educators but the entire school community and system.

Classroom Assessment, Evaluation and Reporting

Assessment, evaluation and reporting at the classroom level is a complex process that requires careful consideration of many contextual factors and development of sophisticated skills.⁸ It is not possible to consider it in any detail in a short policy-oriented paper of this sort. However, certain fundamental observations can be made and fundamental questions asked.

Assessment involves gathering data that will allow a teacher to infer what a student knows and can do. This includes a wide range of student knowledge, skills and attitudes that are relevant to the process of learning and to intended learning outcomes. Student knowledge of facts is relatively easy to ascertain. Understanding of concepts is harder to gauge. Skill development is harder yet, and it is also more time consuming to assess skills, particularly physical skills. Attitudes are by far the most difficult thing to assess and, in fact, there are few tools available to do so.

To illustrate the outcomes that are to be assessed, consider Science in the Intermediate Grades. The curriculum has four goals.

GOAL 1: Science, technology, society, and the environment (STSE)

Students will develop an understanding of the nature of science and technology, of the relationships between science and technology, and of the social and environmental contexts of science and technology.

GOAL 2: Skills

Students will develop the skills required for scientific and technological inquiry, for solving problems, for communicating scientific ideas and results, for working collaboratively, and for making informed decisions.

⁸ Classroom assessment is discussed at http://www.bced.gov.bc.ca/classroom_assessment/. See also Rethinking Classroom Assessment With Purpose in Mind at <http://www.wncp.ca/media/40539/rethink.pdf>

GOAL 3: Knowledge

Students will construct knowledge and understandings of concepts in life science, physical science, and Earth and space science, and apply these understandings to interpret, integrate, and extend their knowledge.

GOAL 4: Attitudes

Students will be encouraged to develop attitudes that support the responsible acquisition and application of scientific and technological knowledge to the mutual benefit of self, society, and the environment.

It is very difficult to assess student learning and growth in some of these areas. In all areas it requires professionally informed inference by the teacher on the basis of daily interactions with the student, assigned work and tests. Unfortunately, the most difficult learning outcomes to assess may be the most lasting and influential on the student's life.

Since assessment of what the student knows and can do is subtle and imprecise, evaluation is also imprecise. Marks on a test may give the appearance of accuracy and precision, but that is misleading for many reasons, including complexities arising from the selection of items to include on the test, the variable ability of students to represent their learning in this way, the interpretation of the student response by the teacher and so on. These factors are particularly vexing in the evaluation of thinking skills and attitudes such as those described above.

Provincial policy requires assessment and evaluation to be referenced to the expected learning outcomes in the curriculum. Although percentage grades based on assignments and tests may be used, student evaluation should not be reduced to calculations based on such data. Ultimately, evaluation is a professional judgment that should consider all available information about the student.

Ongoing assessment and evaluation serve two formative purposes. The first is to provide immediate feedback to the student about his or her learning that enables him or her to focus on what needs to be learned, seek assistance, adjust learning strategies or increase effort. The second is to inform the teacher so that the style, pace or content of instruction can be adjusted to the needs and abilities of students. Research shows that the formative use of assessment and evaluation is a powerful way to improve student achievement.

Formative reporting to parents should be ongoing and can take many forms, including parental review of homework, parental inquiries, class newsletters, conferences and open houses. Formal summative reports are required three times per year and in Grades 4 to 12 they must also include letter grades, which are defined in provincial policy. For example:

- A = The student demonstrates *excellent or outstanding* performance in relation to expected learning outcomes for the course or subject and grade.
- B = The student demonstrates *very good* performance in relation to expected learning outcomes for the course or subject and grade.

Letter grades provide a broad-stroke summary and convey very little information about what the student has learned, the student's strengths, what the student needs to improve, how the student can improve or how the parent can assist. Thus, reporting can also include anecdotal comments and provision for personal conferences with the student and/or parent. In the secondary grades, of course, this is limited by the large number of students seen by each teacher.

Formal reporting also includes written comments that describe student behaviour, including information on attitudes, work habits, effort and social responsibility. However, behaviour and effort are not to be considered in determining a student's grade, which is purely reflective of what has been learned in relation to expected outcomes. This means that absence, late assignments or cheating on tests must be dealt with in ways other than a marking penalty.

Both formative and summative reporting should serve to create a partnership between the teacher student and parent in which responsibility is shared. This is best accomplished when expectations are known in advance, when communication is ongoing and flows in both directions, and when each partner contributes what they know and observe from their unique perspective.

One important consideration within an overall analysis of assessment and evaluation is the rapid emergence of 'assessment for learning' (AFL) in classrooms and schools. While not described in depth here, the fundamental principle behind AFL is one of shifting the locus of learning from the teacher to the student. This is done by:

1. Providing learners with clarity and understanding of the learning intentions.
2. Providing and co-developing with learners the criteria for success.
3. Providing regular, thoughtful feedback.
4. Designing thoughtful classroom questions to lead discussions to evidence learning.
5. Putting learners to work as learning/teaching resources for each other.
6. Doing everything possible to ensure that learners are the owners of their learning.⁹

Common Standards and Common Assessments

One question that continues to circle around the province is how to ensure consistency of assessment and evaluation in domains of learning, including those defined within mandated curriculum, and then communicate these results to students and parents with clarity and consistency. As students, parents and educators have said for many years, "How can we be sure that 'meeting expectations' in one class means 'meeting expectations' in another?"

For that consistency we need common standards, particularly in key areas of learning such as reading, writing, numeracy, social responsibility and information technology. Those common standards do exist and are widely respected across the system - the BC Performance Standards. Although optional, those standards are in widespread use and form the basis for not only classroom practice but assessment systems in use throughout the province.

⁹ These principles come from the work of Dylan Wiliam and, in BC, Caren Cameron among others.

We also need commonly administered assessments based on those standards. In many districts, adherence to the Performance Standards has been accompanied by collective commitment to using high quality assessment tools at common times throughout the year. The use of these tools provides teachers, students and parents with a good sense of where children are in relation to standards early in a learning cycle (for example, 'cold writes') and as a means of informing instructional practice for children, thereby serving as assessment for learning. The same instrument, administered later, becomes a more summative assessment of learning based on common standards.¹⁰

An emerging mindset is that it is actually possible to make a province-wide commitment to working from common standards in key areas of learning, most obviously the BC Performance Standards, and to have commonly administered assessments used consistently within classrooms across the province. This would serve multiple purposes including: ensuring consistency in standards being used for assessment and evaluation; creating system-wide integrity in regard to the use of those standards for communication of student progress; engendering confidence in parents of the same; generating a collection of authentic classroom-based assessment data that can be used for grade-specific, school-wide and system-wide planning for improvement of practice; and, providing the public with the confidence in public education that it needs and deserves.

Foundation Skills Assessment

There are several components to the provincial assessment program, but the Foundation Skills Assessment (FSA) is currently the only system-wide measure of achievement for elementary students. It is valued and used by many, including parents who see it as providing the only means of determining progress via a consistently administered standards-based assessment, and by the system as a basis for planning. It has been questioned by others in several respects, including the inherent merits of FSA data, the appropriate uses of the data, the cost-benefit ratio of the assessment program and the use of FSA data by the Fraser Institute.

The FSA examines core knowledge and skills in reading, writing and numeracy that have been acquired over a period of several years. It includes only a subset of the intended learning outcomes in these parts of the curriculum and is not intended to examine the current learning of students at their particular grade level. It is, therefore, quite different in its design and purposes from the ongoing assessment and evaluation conducted by the teacher.

The FSA is a criterion-referenced performance assessment aligned with the Prescribed Learning Outcomes in provincial curriculum and it is field tested before use to refine the test items. Consequently, it is 'valid' and 'reliable' in a technical sense.¹¹ However, as with any single evaluation, it has a high margin of error for individual students. A typical school would have a relatively small number of students at each grade level so the margin of error remains quite high, but at the district level that margin is much smaller due to the larger sample size.

The FSA consists of 6 sessions totaling 4.5 hours spread over a period of approximately two weeks in February. Some have asked whether the opportunity cost in terms of effort and lost instructional time is worth the benefit.

¹⁰ Examples include DART (District Assessment Reading Team) and RAD 36 (Reading Assessment District 36)

¹¹ Statistical validity and reliability is explained at <http://www.socialresearchmethods.net/kb/relandval.php> and at <http://writing.colostate.edu/guides/research/relval/>.

Parents receive FSA results for their children, meaning that FSA requires a full census approach. Although the Ministry does not use the results to evaluate school or district performance, it does provide the data to schools and Edudata generates an item analysis that is used by many schools.¹² If the FSA were only intended to determine whether provincial curriculum is effective and to monitor the overall effectiveness of the school system, random sampling would be sufficient. Sampling could also be designed to assess and evaluate system performance for particular sub-populations (e.g., subjects, genders, ages, special programs, students with special needs or aboriginal students). However, many have argued that due to the small sample sizes in so many schools and districts, the only way to sample those sub-populations is to continue to assess all students via a census approach. Additionally, many parents have expressed a desire to see the results for their children as FSA is the only commonly administered standards-based assessment in use.

While there are reasons to support a census approach to FSA, the concerns about the limited scope, large margin of error and costs are worthy of consideration. Of further interest is the concern expressed by some that FSA is, or may become in the future, a high stakes assessment such as occurs in Britain and the United States. Currently, although FSA is a large-scale standardized test, it is very different from these high-stakes assessments in that it technically has “no stakes” in the sense of specific consequences for students, teachers, schools or districts. However, the annual “Report Card” produced by the Fraser Institute using FSA data is inappropriately interpreted by some as a reliable measure of school quality, which it is not. This affects parents’ choice of schools, undermines confidence in public schools in general and is even said to affect real estate values in some parts of the province.

The Fraser Institute Report Card is based on a statistic that it claims to be an indicator of school quality and which it uses to rank schools. The statistic itself is a manipulated combination of FSA data with no clear logic to its composition so that both its meaning and its value are unclear.¹³ What is clear is that it is at best a very limited indicator of student achievement. It is said to create false impressions for the public in a manner that has been decried by major educational stakeholders and, in regard to ranking of schools, the Ministry.

Many legitimate questions have been asked about the Foundation Skills Assessment, but there has been little constructive dialogue amongst stakeholders about them. The Education Advisory Council may be able to begin to rectify this through its leadership of a collaborative system-wide dialogue about assessment.

Conclusion

Effective assessment and evaluation practices are essential to both student and system success. Properly employed and communicated they can also contribute to public confidence in and support for public education. It is important, therefore, that there be well-informed and thoughtful dialogue about policy and practice within the Education Advisory Council. It is hoped that this paper has contributed to that dialogue.

¹² See http://www.edudata.ca/apps/fsa_item/

¹³ For example, no rationale has been provided for the weighting given to the various elements and no reason is given for the 20% of the statistic attributed to the difference between scores for boys and girls, which inflates the statistic for single gender schools. Moreover, the statistic is used without any information about its margin of error, which gives rankings an unwarranted appearance of precision and reliability.

Questions for Discussion

Out of this overview emerges a set of questions that could serve as discussion starters for consideration by the Education Advisory Council, out of which might come broader provincial dialogue and consideration of new or revised policy directions for the province. Those questions begin with:

1. What are our commonly held beliefs about the purposes of assessment?
2. How can teacher, student, parent and “system” understandings of the purposes and processes of assessment and evaluation be aligned? Specifically, how can we collectively address the need for commonly administered standards-based assessment?
3. How can we balance the need for system-wide practices with teachers’ desires for professional autonomy? Can those become aligned and harmonious ideals?