Assessment and measurement

Beyond testing

Summary

- Assessment and measurement play a key role in informing educators' practice in classrooms, students' understanding of their own learning, parents' capacity to support their children's success in school, and policy-makers' decision-making.

- Competencies in creativity, social-emotional learning, citizenship, and health should be assessed for the same reasons that reading, writing, and math are assessed – to provide relevant, specific information about student learning in these vital areas.

- Assessment of these competencies is complex, and we cannot rely on the tools and strategies typically used to assess other skills or knowledge.

- At the classroom level, a range of assessment strategies can be used to understand students' growth in these areas.

- It is possible to assess these competencies at a jurisdictional level, however standardized assessments or surveys can only give information of limited quality about complex competencies.

as·sess·ment (n):

Assessment is a process of gathering information that reflects how well a student, classroom, school, or school system is doing against a set of purposes, learning criteria, or curricula (Ontario, 2010).

meas·ure·ment (n):

Measurement is a specific process through which a learning experience, phenomena, or context is translated into a representative set of numerical variables.

Assessment and measurement are closely linked concepts in education. Both can be used for a variety of purposes, including:

- reporting to students, parents and the public;

- as a method for supporting students' understanding of educational goals and their own progress;

- as a means for teachers to understand the effectiveness of their own practice;

- strengthening school-wide collaboration;

- to provide the data necessary to discern whether the system is working equitably for all students.

The complexity of assessment

Across the world, educators, policy-makers, and experts agree that student success in both school and life includes more than literacy and numeracy skills and academic content knowledge (e.g. Christensen & Lane, 2016; OECD, 2016). Many education systems are endeavouring to embed broader competencies – referred to as everything from 21st century skills to global competencies – into curricula, outcome expectations, and assessment strategies.

But expanding assessment strategies beyond outcomes in "the basics" is complex. It is difficult (and, some would argue, inappropriate) to set benchmarks for students' competencies in areas such as social-emotional learning, creativity, health, and citizenship. These competencies are not "knowledge" that can be captured on traditional tests. Some of these competencies are related to a process rather than a product, and some are more likely to be observed through social interactions.

Assessment strategies start in the classroom

Since 2015, educators from across Ontario have been field testing the use of a set of concrete, observable competencies in health, creativity, social-emotional learning, and citizenship. These educators are exploring how they might teach these competencies in their classrooms, monitor student progress, and provide feedback to move students forward.

Their assessment methods include checklists, observations, student journals, and collaborative work developing scales to track how frequently students apply specific competencies.

A number of themes are emerging from our work with educators on the ground and experts in the field.

Focusing on growth: By using assessment processes that rely on feedback loops – from teacher to student, student to teacher, and students to students – the educators are able to favour an orientation toward growth, rather than solely focusing on achieving benchmarks. They are designing learning environments where students have opportunities to practice the competencies and focus on their progress.

This type of assessment, with a focus on feedback rather than grades, promotes intrinsic motivation in students (Deci, Koestner, & Ryan, 1999).

In one school, teachers are using a numberless "sliding scale", where the student and the teacher rate the student independently, and then talk about how their ratings compare. They find that using the sliding scale leads naturally into a discussion of how the student might move up the scale.

Collaborative assessment to support student learning: There is a large body of evidence that points to collaboration among educators as an important contributor to instructional improvement in schools, and to the value of collaboration within assessment processes (e.g. Cochran-Smith & Lytle, 2009; Hargreaves, Earl, Moore, & Manning, 2002).

Through collaboration, educators are able to explore and assess their own practices, develop assessment tools, and continually evaluate how their classroom conditions support students in developing the competencies.

Using assessment to adjust learning conditions: The information educators collect is used "in the moment" to adjust learning opportunities (assessment *for* learning), so that classroom activities are tailored to the students' levels of readiness. This type of assessment has been shown to raise overall student achievement, and larger gains are seen among students with lower achievement, meaning it can reduce gaps (p. 141, Black & Wiliam, 1998).

Specific language supports effective assessment: Because the competencies are defined in specific, observable terms, they can be used as learning goals. This allows the educators to use evidence gathered in a variety of ways to assess student progress toward those goals.

The specificity of the competencies also means that teachers are able to give students feedback about how they demonstrate the competency within a task, experience, or process. Feedback that is directed at the task, rather than the person, has been shown to improve achievement, and is an important part of effective social-emotional learning interventions (e.g. Hattie & Timperley, 2007; Kyllolen, 2013).

The risks and rewards of jurisdiction-level assessment

The purposes of classroom-level assessment are different from the purpose of assessment at the jurisdictional (board, province) and international level.

Jurisdictionally, measurement and assessment are more often used to provide information to policy-makers and the public about how systems are doing. Large-scale assessments can include performance tasks (including tests or essays), third party (e.g. teacher or parent) questionnaires, or self-report student surveys. They can be census-based (everyone is assessed), or sample-based (a portion of the population is assessed).

While many systems are exploring ways to assess competencies in areas such as creativity, health, social-emotional learning, and citizenship, as well as ways to gather data about learning environments across the system, there are both risks and rewards to this type of reporting.

The benefits
Large-scale assessments can:

- identify where additional system support is needed and provide accountability to the broader public in relation to system performance in these areas.

- provide information about the impact of policies and programs in relation to creativity, citizenship, social-emotional learning, and health.

- track the system's progress in providing learning conditions that support equitable outcomes in these vital areas of schooling.

- promote the visibility and importance of these areas in a climate of educational accountability that currently relies on more narrow indicators of achievement.

The challenges
Large-scale assessment can:

- provide information that can be misinterpreted. For example, it may be possible to measure one or two competencies in creativity at a jurisdictional level, but it cannot be assumed that this information represents system performance in creativity overall.

- create a false assumption that achievement on a test provides a complete picture of a student's ability in relation to the competency being measured (e.g. Johansson, 2016; Sellar, Thompson, & Rutkowski 2017).

- drive policy, funding and accountability in the system in ways that can narrow pedagogic approaches, and constrain experimentation and risk-taking (e.g. Kempf, 2014).

- increase time and resources spent in classrooms to prepare for and deliver the assessment, and resources spent at the jurisdictional level to administer assessments.

Addressing the challenges of large-scale assessment

Because it is important to have information about these areas of learning, it is worth exploring ways to navigate the challenges of jurisdictional assessment.

For example, standardized assessments given to a sampled population of students, on a sampled selection of competencies, could provide information about system performance, while avoiding some of the negative consequences of large-scale measurement (e.g. FINEEC, 2016). Or, regionally-based teams of teachers could develop standards and then conduct team-based assessment at different levels/grades in the system. This would provide jurisdictional information about student performance, but it would require more resources, infrastructure, and capacity (e.g. Mooney & Mausbach, 2008).

Assessment is a core component of education. But before exploring jurisdictional measurement of the system's performance in building students' competencies in creativity, social-emotional learning, health, and citizenship, we must understand more fully how to support educators in assessing these vital areas, and how to prevent jurisdictional measurement from skewing classroom practices.

Assessment for, as and of learning in the classroom

Classroom assessment is commonly categorized into three "purposes," based, to a large degree, on how the information is used (Black & Wiliam, 2007; Ontario, 2010).

- *Assessment for learning* – assessment used to determine what students already know and what they need to learn. This may be either a 'diagnostic assessment' used prior to teaching a new concept to determine students' readiness to learn, or an ongoing assessment of students' progress during the learning process. Feedback is a key part of assessment *for* learning.

- *Assessment as learning* – assessment process in which students assess themselves or their peers within the lesson.

  Both assessment *as* and *for* learning fall under the category of formative assessment, and the ongoing feedback loop created throughout the process is considered as much a learning process as it is an assessment process. (Ramaprasad, 1983; Ontario, 2010).

- *Assessment of learning* – assessments used to determine the accumulation of knowledge/ skills that students have gained from their learning. This is the type of assessment that is most often associated with grades on projects, tests, or student report cards. Standardized test results are another form of summative assessment. This example of this type of assessment. assessment is summative, rather than formative.

The distinction between assessment *for, as,* and *of* learning lies in how the information is used, and by whom. The same instrument - a written test, for instance - may be a tool for assessment *for, as,* or *of* learning. The test may be an opportunity for students to compare their understanding to a standard (assessment *as* learning), a way for a teacher to gauge where her students are in learning a concept (assessment *for* learning), or a moment to reflect on what has been learned and to grade performance (assessment *of* learning).

From our field trial teachers...

"What we found was that no matter where we did it, the conversation with the sliding scale had the student already giving their next steps... And that was our biggest 'Aha!'...When the student was putting their dot on the sliding scale, they were saying, 'Well, if I did this, then I would be here [at a higher point] on the scale.'"

"If I don't know what I'm looking for, I can't notice it. I can't name it. I can't respond to it. I can't plan next steps... I think [the Measuring What Matters competencies] help us unpack what meaningful learning looks like and sounds like, and then we have the autonomy and creativity to act - that's the art and science of teaching."

"For me, the domains and competencies provide a way in which we can talk about learning and not just achievement. When we only look at the expectations in the curriculum, we tend to only talk about achievement. A student can do that or they can't. We don't really think about growth. We don't think about the conditions for learning that we are responsible for creating, either as a teacher or an administrator in a building."
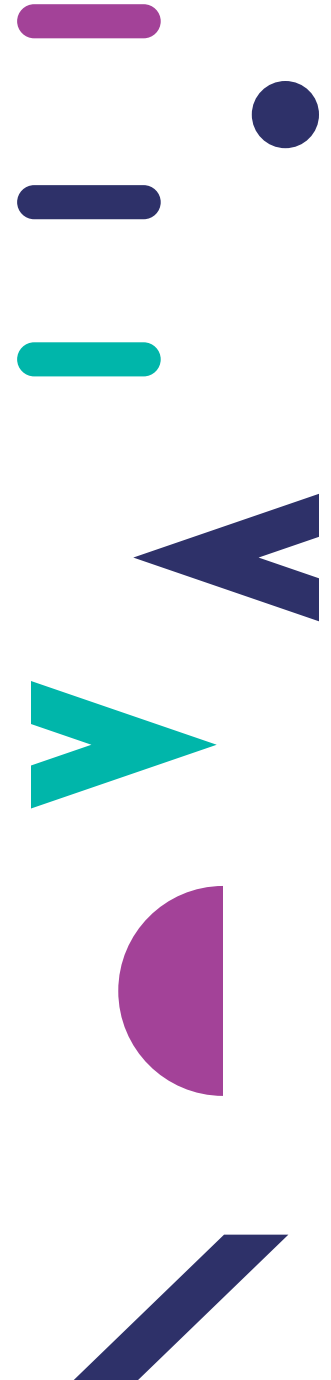
References

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P. & Wiliam, D. (2007). Large-scale assessment systems design principles drawn from international comparison. *Measurement, 5*(1), 1-53.

Cochran-Smith, M., & Lytle, S. (2009). Inquiry as Stance: *Practitioner Research in the Next Generation.* New York, NY: Teachers College Press.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *The Psychological Bulletin,* 125(6), 627-668.

Finnish Education Evaluation Centre (FINEEC). (2016). *National Plan for Education Evaluation 2016-2019.* Helsinki, Finland: FINEEC.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hargreaves, A., Earl, L., Moore, S. & Manning, S. (2002) *Learning to Change: Teaching Beyond Subjects and Standards.* San Francisco, CA: Jossey-Bass Education, Wiley.

Johansson, S. (2016). International Large-Scale Assessments: What Uses, What Consequences? *Educational Research 58*(2), 139–48. doi:10.1080/00131881.2016.1165559.

Kempf, A. (2014). "Revising the pedagogic form: Test-oriented teaching and learning" In *The Pedagogy of Standardized Testing.* New York, NY: Palgrave MacMillan, 71-97.

Kyllonen, P. C. (2013). *Measurement of 21st Century Skills Within the Common Core State Standards.* US: K-12 Center, Education Testing Service. Retrieved from https://cerpp.usc.edu/files/2013/11/Kyllonen_21st_Cent_Skills_and_CCSS.pdf

Mooney, N. & Mausbach, A. (2008). *Align the Design: A Blueprint for School Improvement.* Alexandria, USA: ASCD.

Ontario Ministry of Education. (2010) *Growing Success: Assessment, Evaluation, and Reporting in Ontario Schools.* Queen's Printer for Ontario.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4-13.

Sellar, S., Thompson, G., & Rutkowski, D. (2017). *The Global Education Race: Taking the Measure of PISA and International Testing.* Canada: Brush Education.