Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Learning from the Queen's University Assessment Experience: Considerations for selecting an appropriate skills measurement tool

Jackie Pichette and Elyse K. Watkins

Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Cite this publication in the following format:

Pitchette, J. & Watkins, E. K. (2018). *Learning from the Queen's University Assessment Experience: Considerations for selecting an appropriate skills measurement tool.* Toronto: Higher Education Quality Council of Ontario.

Table of Contents

Background

Are students developing the sorts of skills we expect of a high quality education system? As researchers at the Higher Education Quality Council of Ontario (HEQCO), this question is at the forefront of our work. We appreciate that students enrolling in a postsecondary program seek more than discipline specific knowledge; they seek skills and competencies needed to succeed in their personal and professional lives.

Since 2012, HEQCO has collaborated with Ontario colleges and universities to investigate student skills development and inform our understanding of where and how Ontario can improve its education system. A few examples include:

- The **Learning Outcomes Assessment Consortium (LOAC)** began as a partnership between seven postsecondary institutions interested in testing or developing reliable and valid instruments for assessing student outcomes. As some of those projects wrap up, we are bringing five new institutions to the table and scaling up the use of effective skills assessment tools in diverse learning environments and across entire institutions.

- Our **Postsecondary and Workplace Skills (PAWS) Project** examines the relationship between critical-thinking skills in postsecondary education and labour market outcomes by linking the results of a skills assessment with income tax data.

- The **Essential Adult Skills Initiative (EASI)** piloted an international test of core skills such as literacy, numeracy and problem solving among first- and final-year students at 19 colleges and universities to understand differences in skills across entering and exiting cohorts.

Our partners at Queen's University have been involved with each of the aforementioned projects, and in the process, have gained valuable experience administering a range of large-scale skills assessment tools. Recognizing that the quality of our assessment projects hinge on the appropriateness of the tools we incorporate, we saw an opportunity to learn from the experience of Queen's University with five unique tools to capture student skills: the Collegiate Learning Assessment Plus (CLA+), the Critical Thinking Assessment Test (CAT), HEIghten Test, Education and Skills Online (ESO), and the American Association of Colleges and Universities (AAC&U) VALUE Rubrics.

As described in Table 1, these tools provide different ways of assessing skills such as critical thinking, problem solving and communication. This paper provides a look at how and why these assessments can be used to measure skills development at Ontario postsecondary institutions.

Table 1: Skills Assessment Tools Employed as Part of HEQCO-Queen's Partnership

Tool	Administrative Details	Skills Measured	Tasks	Results
Collegiate Learning Assessment Plus (CLA+)[1]	• 90-minutes • Online • Distributed by the Council for Aid to Education	• Critical thinking • Problem solving • Analytic reasoning • Effective communication • Scientific reasoning • Quantitative reasoning • Critical reading	**Short answers:** ("Performance tasks"): open-ended questions about real-world, problem-based situations **Multiple choice questions:** analyzing documents and selecting appropriate descriptors	Available 4-6 weeks following assessment **Test taker:**[2] total score, level, rank and sub-scores **Institution:**[3] mean results, effect size, value-added scores
Critical Thinking Assessment Test (CAT)[4]	• 1-hour • Paper-based • Distributed by Tennessee Technological University	• Evaluating information • Creative thinking • Learning and problem solving • Communication	**Short answers:** prompting questions about real-world, problem-based situations	Up to three months to be returned[5] **Test taker:** total score, sub-scores **Institution:** mean total scores and sub-scores
HEIghten® Test[6]	• 45-minutes • Online • Distributed by ETS	Critical thinking, specifically: • Analytic skills • Synthetic skills	**Short answers:** prompting questions assess social and conventional knowledge	Available as soon as the test has been competed **Test taker:** total score, level, comparison group scores

[1] http://cae.org/images/uploads/pdf/CLA_Plus_Technical_FAQs.pdf
[2] http://cae.org/images/uploads/pdf/Sample_CLA_Plus_Student_Score_Report.pdf
[3] http://cae.org/images/uploads/pdf/Sample_CLA_Plus_Institutional_Report_Cross_Sectional.pdf
[4] http://www.queensu.ca/qloa/assessment-tools/cat-test
[5] https://www.tntech.edu/cat/about/
[6] https://www.ets.org/heighten/about/critical_thinking/

Tool	Administrative Details	Skills Measured	Tasks	Results
			Long answer: informational passage prompts argumentative essay	**Institution:** mean scores and sub-scores, comparison groups scores
Education and Skills Online (ESO)[7]	• 90-minutes • Online • Distributed by the OECD	• Literacy • Numeracy • Problem solving	**Adaptive scenario-based questions:** increase or decrease in difficulty depending on test-taker's performance	Available as soon as the test has been competed[8] **Test taker:** scores, levels, comparison group scores **Institution**: mean scores, levels, comparison group scores
VALUE Rubrics[9]	• Downloadable, customizable • Developed and led by the AAC&U	• Intellectual and practical skills (e.g., critical thinking) • Personal and social responsibility (e.g., ethical reasoning) • Integrative and applied learning	**Applied to a student's own authentic course work**	Varies depending on several factors (i.e., experience with rubrics, number of assessors etc.)

[7] http://www.heqco.ca/SiteCollectionDocuments/HEQCO%20EASI_Status%20Report%20on%20College%20Pilot_March2017_English.pdf
[8] http://www.oecd.org/skills/ESonline-assessment/assessmentadministration/results/
[9] https://www.aacu.org/value-rubrics

About this report

In the winter of 2018, we conducted six, one-hour, semi-structured interviews with members of staff, faculty and administration at Queen's University. Each interviewee was involved with the implementation and/or integration of one or more of the assessments described above in Table 1 as part of the LOAC, PAWS or EASI projects.

This report summarizes some of the common themes that emerged during our interviews, framed as lessons to inform the selection of appropriate measurement tools in the context of skills-assessment projects at Ontario postsecondary institutions.

Considerations for Selecting an Appropriate Skills-assessment Tool

Each of the Queen's representatives we interviewed explained why a given tool did or did not work well in a specific context, and summarized their impressions of what each tool measured and how effectively. Interviewees also offered advice for other institutions looking to administer skills assessment tools. Three decision-making factors emerged from the interviews as being particularly important when selecting an appropriate skills assessment tool:

1. Clarity of purpose
2. An understanding of stakeholder wants and needs
3. An inventory of process requirements

We describe each factor in more detail below.

Clarity of purpose: What skills are being assessed and why?

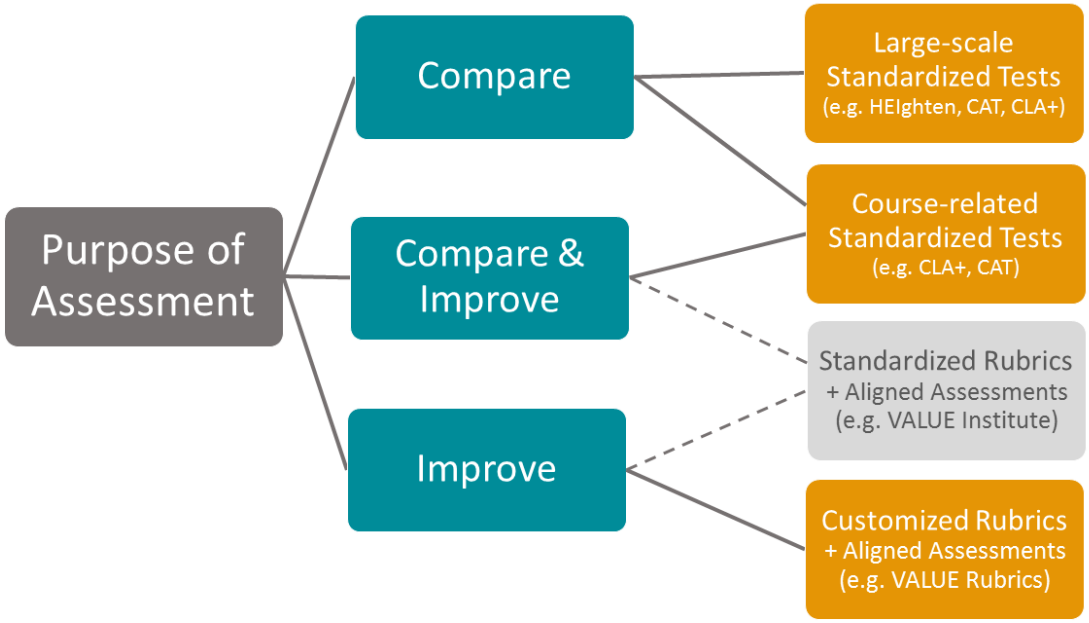
"You have to be very clear about your outcomes. Why are you doing this? What do you want to find out? What is your capacity? Is your institution assessment ready?"

An institution selecting a skills assessment tool undoubtedly intends to measure skills, but it's important to be clear about *which* skills and *why.*

Our team at HEQCO has been working with Queen's University since 2012 to assess students' learning outcomes. In some cases, we wanted to get a sense of how students in Ontario are performing relative to each other; in other cases we wanted to understand how Ontario students are faring relative to other jurisdictions. In each instance we (HEQCO) sought to inform our understanding through comparisons. Our research partners at Queen's shared our interest in understanding comparative skills development, while also being motivated by an interest in improving their courses and programs.

Figure 1, below, illustrates how a clear articulation of purpose can narrow the scope for selecting appropriate assessment tools. (Note that the goals reflected in teal are not inclusive of all the goals an institution might have). The orange boxes offer examples of tools that can fulfill the goals. "Standardized Rubrics" are in grey because the example we know of, the VALUE Institute, is still in development.

Figure 1: Assessment Tool Selection Flowchart



Our interviewees spoke about the goals depicted above and considered how each of the tools outlined in Table 1 served to advance them.

Compare. One possible intention behind the implementation of a skills-assessment tool may be to understand how students' skills developed in one context compared to those in another. Interviewees recognized that standardized tests enable comparisons across programs, institutions and jurisdictions. The tests provide an understanding of student achievement and in that way facilitate conversations about improvement at a system or institutional level. Standardized tests are also extremely valuable from an accountability standpoint and as a tool for validating other locally developed assessments. While there is clearly value in comparison, our interviewees noted it must be done carefully. One interviewee cautioned that doing an assessment for the sole purpose of comparison will likely receive resistance from stakeholders, as for many people it implies ranking and punitive measures.

Improve. Our interviewees also spoke of the potential for skills assessments to inform changes in curriculum and pedagogy and, in turn, students' skills development. They explained that while rubrics, such as the VALUE[10] rubrics developed by the AAC&U, offer less comparability between institutions, they provide an opportunity to look more closely within a single institution. In that way, rubrics offer promise for program level improvement and for understanding an individual student's progression. Rubrics can be aligned with a specific course's content, and the process of customizing and aligning rubrics and assignments within a course can be helpful for clarifying learning outcomes and for facilitating a reflection on teaching and assessment practices. As one interviewee explained, when compared with large-scale standardized tests, "Rubrics offer more transformation and commitment from the instructor."

Compare and Improve. The intentions behind large-scale skills assessment might also be two-fold: to compare students' results and to improve skills development. Assessments selected to fulfill this dual purpose can facilitate conversations about students' achievements at an institutional and/or system level. Our interviewees identified particular standardized tests as useful for comparing results with locally developed assessments and, in turn, improving the validity and reliability of existing assessment practices. One interviewee also pointed out that there may soon be potential for the VALUE rubrics to act as a comparative assessment tool, through the VALUE Institute. The institute will establish a repository of student work, assessed by external scorers, to benchmark learning outcomes at a national scale.[11]

Understanding your stakeholders: Who are they and what do they want?

"One of the first steps is having conversations with all stakeholders from the top, middle and bottom."

After clearly articulating the underlying assessment goals, it's important to consider who will be included in the assessment process (e.g., students, faculty, IT staff, etc.), what their interests are, and how their involvement can be made both easy and meaningful.

Each of the Queen's representatives we spoke with highlighted the importance of ensuring assessments are relevant for stakeholders. Without considering and addressing key stakeholder wants and needs, confounding variables such as low motivation or effort can detract from results. In one interviewee's words, "The instructor needs to care about the assessment and the students need to see it linked to the course or connected to course goals."

One of the most important stakeholder groups to consider when selecting skills-assessment tools is students. When students are not motivated to perform well on an assessment, the validity of the results is put into question. As Lui, Bridgeman and Adler explain, "Highly motivated students tend to perform better than less motivated students" (2012). To address the issue of low motivation, interviewees emphasized the

[10] VALUE stands for Valid Assessment of Learning in Undergraduate Education
[11] https://www.aacu.org/VALUEInstitute

importance of selecting suitable assessments, clearly communicating the purpose(s) and benefit(s) of a selected assessment to students, and raising the stakes or appropriately incentivising assessments.

In speaking about ensuring the suitability of the assessment for students, one interviewee shared an example of a test, originally designed for American students, with demographic questions that didn't resonate for Queen's students. It left students feeling frustrated and unmotivated to give their best effort. Another interviewee spoke about ensuring the assessment tool is appropriately challenging, saying, "It seemed that the students did not find [one of the assessments] as rewarding, but frustrating, because it took longer than it had been advertised and many of the questions were too easy." Similar problems could arise if the assessment is too difficult.

It's equally important that students see value in the assessment. Interviewees explained that unless students (upper-year students in particular) perceive alignment between the test and their academic programs or recognize that the test might help them in the future, they won't try very hard. Certain assessments may offer benefits for students beyond the scope of their courses, and it's important to think about and communicate those. One interviewee heard from several students that one of the assessments felt like good practice for their upcoming LSAT or GRE. Other students said they planned to include their results of an internationally recognized test as part of their graduate school applications. Students who were involved in the longitudinal studies were also curious to compare for themselves how they had performed in the previous years, as well as against their peers.

Institutions can motivate students by raising the stakes or using incentives. Interviewees noted that weighting assessments toward students' course grades was a good way to incent effort. On a related note, all of our interviewees expressed a preference for using a validated rubric to assess assignments that are aligned with the courses and counted toward students' final grades. Monetary incentives are also an option: In one case at Queen's, Amazon gift cards were used as incentives. However, studies suggest these are only effective when extra payment is tied to student performance, as opposed to solely participation (Lui et. al., 2012; Wise & DeMars, 2005).

Faculty are another key stakeholder group that must be considered when selecting an assessment tool. Instructors are unlikely to incentivize students (with a course grade or encouragement) or support the assessment process unless they consider the results to be useful. Factors such as the subject matter of the test and the length of time it takes to provide test results to students and instructors can impact the perceived utility of the information and, in turn, motivation.

It's also important to think about the time and resources required of an instructor. Multiple interviewees recounted the challenges of customizing a rubric and aligning it with an appropriate assignment. Interviewees recommended that institutions offer support to instructors in the form of an assessment facilitator who can assist with adapting rubrics, designing course content to scaffold relevant skills, and aligning assessments with courses and rubrics. Multiple interviewees also suggested having a library that instructors can access and that includes authentic tasks that align with commonly used rubrics (e.g., VALUE rubrics) and examples of how a rubric might be applied in a particular subject area.

Process requirements: How will this work?

"Even if what you're doing is small, you are going to learn so much."

With an understanding of which skills are being measured and why, and knowledge of stakeholder interests, it's important to consider the process requirements of an assessment before implementing it.

Successfully implementing large-scale assessment tools requires careful planning to make the process as seamless as possible. This means avoiding technical difficulties or glitches, being cognizant of time constraints and ensuring instructions are understood by the students being assessed. An interviewee spoke about the importance of a smooth process, saying, "You want to make sure the assessments are manageable and safe for the students…You should not disadvantage anyone by pushing a high-stakes test when something could go wrong technically."

Collectively, our interviewees encouraged institutions to keep the following questions in mind for implementing standardized tests and rubrics on a large scale:

For standardized tests

Primary Questions	Followup Questions
Is the assessment paper- or computer-based?	If the assessment is computer based, • Will students use their own computers or will the institution provide computer access? • Does the assessment require specific software? Can it be downloaded in advance? • Do other computer programs interfere with the testing interface? • Is technical support available from the test provider?
How much time do students need to write the test?	In addition to that, how much time do students need to: • Ask questions? • Understand the goals of the project? • Complete demographics survey? • Is the student aware of any system tests that need to be performed on their devices (e.g., computer or laptop) beforehand?

Primary Questions	Followup Questions
How many students should be assessed together?[12]	Does the test need to be proctored on campus or self-proctored? • If proctored on campus, what is the ratio of students to proctors?
What does the scoring process entail?	• Who will calculate scores? • How quickly will students and instructors receive the results? • In what form will results be disseminated?
Is there an incentive?	• Were students consulted about what the incentive should be? Were there options? • How will the incentive be distributed?

For rubrics

Primary Questions	Followup Questions
Which course(s) will be selected for assessment?	• Has the course already begun? • Does the curriculum, as designed, teach to the skills being assessed?
Which assignment(s) will be assessed?	• Are all of the constructs in the rubric applicable to the assignment?
Who will be applying the rubric to the assignment?	• Do they require training? • How much time will they need to complete the assessment?

These considerations are just some of the details to think through prior to administering large scale assessments. Once a tool has been selected, our interviewees recommended reaching out to other institutions with experience implementing that tool to learn about their process. Interviewees also suggested pilot testing the assessments before implementing them at a large scale. As one interviewee said, "Don't try to go too big too quickly."

[12] Note that our interviewees found that small group sizes (approximately 35 students) work best. In larger groups, regardless of the student-to-proctor ratio, interviewees found a higher incidence of misunderstanding, or of students simply not hearing instructions, which contributed to more technical and comprehension issues.

Concluding Thoughts

Ongoing skills assessment is critical to the sustainability of a high quality education system, and essential for curriculum and pedagogy improvement. Through their involvement with multiple HEQCO-funded projects, our research partners at Queen's University have gained unique insight about the selection and implementation of large-scale, skills-assessment tools. We saw an important opportunity to learn from and document their experience.

This report summarizes interviews with our research partners at Queen's, outlining three important decision-making factors and associated questions to address before undertaking large-scale assessment work:

1. **Clarity of purpose**
   - What types of skills do you want to measure (e.g., critical thinking, problem solving, numeracy, etc.)?
   - Why? Do you want to compare test takers with students in other jurisdictions? Are you aiming to administer a validated tool alongside another assessment to compare results and test validity? Do you want to align your program and assignments with constructs from a validated tool to improve pedagogy or curriculum?

2. **An understanding of stakeholder wants and needs**
   - Who is included in the assessment?
   - What are their interests? And how can their involvement be made both easy and meaningful?

3. **An inventory of process requirements**
   - For standardized tests, what are the logistics (e.g., time, incentive, type, number of students, etc.) of implementing the test?
   - For rubrics, what will be assessed in which course(s)? Who will be involved?

With a variety of tools at their disposal, institutions should be very intentional about selecting an assessment tool that advances a clear goal, invites meaningful stakeholder involvement, and minimizes practical and logistical hurdles for implementation at a large scale.

Acknowledgement

We would like to thank our research partners at Queen's University for sharing their time and insights to inform this report.

References

Lui, O. L., Bridgeman, B., Adler, R. (2012) Measuring Learning Outcomes in Higher Education: Motivation Matters. *Educational Researcher*, *41*(9), 352–362. http://journals.sagepub.com/doi/pdf/10.3102/0013189X12459679

Wise, S. L., & DeMars, C. E. (2005). Low Examinee Effort in Low-stakes Assessment: Problems and Potential Solutions. *Educational Assessment, 10*(1), 1–17. doi: 10.1207/s15326977ea1001_1

Higher Education Quality Council of Ontario

An agency of the Government of Ontario