Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Assessing Basic Cognitive Skill, Transferable Skill and Critical Thinking Development in College Students from Admission to Graduation

Pamela Ingleton and Tim Fricker

Published by:

The Higher Education Quality Council of Ontario

88 Queens Quay West, Suite 2500
Toronto, ON
Canada, M5J 0B6

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca

E-mail: info@heqco.ca

Cite this publication in the following format:

Ingleton, P. & Fricker, T. (2021) *Assessing Basic Cognitive Skill, Transferable Skill and Critical Thinking Development in College Students from Admission to Graduation.* Toronto: Higher Education Quality Council of Ontario.

Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Table of Contents

List of Tables

List of Figures

Executive Summary

From 2017–2019, Mohawk College undertook a research project as part of HEQCO's Learning Outcomes Assessment Consortium (LOAC) to measure the essential skill development of two- and three-year diploma students. This project aimed to build a profile of a Mohawk graduate that covered basic cognitive skills (reading, writing and mathematics), higher-order cognitive skills (specifically critical thinking) and transferable attitudes toward learning (e.g., motivation, confidence and career clarity). This project also sought to validate new and existing assessment tools, while updating and building upon the routine practice of assessing the skill levels and individual characteristics of incoming students.

This project involved the re-administration of reading, writing and mathematics tests originally administered post-admission, as well as portions of Mohawk's Student Entrance Survey (now a Student Exit Survey), to a sample of students in their final term of study. Critical thinking was assessed as demonstrated in embedded assignments in first- and final-year courses using the Critical Thinking VALUE rubric, one of 16 rubrics created and validated through ongoing assessment research by the Association of American Colleges & Universities (AAC&U). Early exploratory work on best practices from the literature, current practices across the Ontario college sector and institutional capacity and preparedness informed the development of the research protocol.

Almost 3,000 students across four academic areas (business, engineering, health and community, justice & liberal studies) were invited to participate in at least one of the research pilots. We invited 911 students to test, registered 90 and ultimately tested 32 (a total of 85 tests) — a participation rate of 3.5%. We invited the same 911 students to complete the Student Exit Survey; 116 submitted surveys, 92 consented to their use for research (or 10%). We collected a total of 636 assignments out of a possible 2,570, representing a 25% participation rate (cohort participation ranged from 10–83%).

Unfortunately, due to a disruption to our project funding, concluding this work by completing our summative data analysis was not possible. However, having completed all of our intended activities with the exception of that analysis, our processes and protocols, outcomes of our early discovery process, lessons learned and future recommendations have been assembled in this report to support future, related research and practice. This includes:

- A list of 10 key themes gleaned from a conceptual **literature review** of large-scale assessment projects; literacy, mathematics, and critical-thinking assessment; and studies into the reliability and validity of assessment tools
- An **environmental scan** to learn more about the current state (as of early 2018) of pre- and post-admission assessment tools and processes in Ontario, including student entrance surveys
- An **assessment of our operational capacity** to deliver our current assessment tools (math, reading, writing) online and for the graduating cohort
- An **inventory** of all current, final-semester experiential learning activities undertaken at the college (with a focus on final-semester **capstones**) and how they link to our first-

semester courses and Institutional Learning Outcomes (ILOs), as well as the means by which they are assessed

- A **research protocol** for exploring differences in cognitive and critical-thinking skills and their underlying competencies for incoming and exiting students

From the perspective of our research process, we do feel positioned to conclude that **the re-administration of voluntary post-admission tests in a student's final term is likely not a viable approach to institutional skills assessment**. When asked to complete this supplementary, low-stakes testing, the majority of our student participants were unwilling and/or uninterested, even with significant financial incentive at stake. The option to test online with the support of remote proctoring — implemented in the second term of our pilots — proved more of a deterrent than a motivator due to cumbersome and problem-ridden registration and login processes.

However, students were significantly more willing to consent to the use of their assignments for rubric scoring, an approach that also resonated with faculty and administrators. This **curriculum-embedded assessment was viewed as having more direct applicability to and impact on curriculum development and student learning than the re-administration of standardized tests**. While assignment scoring could be challenging, it was also found to be deeply impactful and informative, leading to immediate curricular adjustments and more purposeful, instructional scaffolding. To this end, since the completion of our research with HEQCO, Mohawk has continued to explore and pilot modified approaches to curriculum-embedded skills assessment, improving upon its efficiency and actionability, most notably by integrating skills assessment with existing quality assurance practices.

Introduction

From 2017–2019, Mohawk College undertook a research project as part of HEQCO's Learning Outcomes Assessment Consortium (LOAC) to measure the essential skill development of two- and three-year diploma students. This project aimed to build a profile of a Mohawk graduate that covered basic cognitive skills (reading, writing and mathematics), higher-order cognitive skills (specifically critical thinking) and transferable attitudes toward learning (e.g., motivation, confidence and career clarity). This project also sought to validate new and existing assessment tools, while updating and building upon the routine practice of assessing the skill levels and individual characteristics of incoming students.

This project involved the re-administration of Mohawk's "Assessments for Success." About 70% of new students at Mohawk College participate in a post-admission assessment and survey process consisting of:

- Reading and writing assessments delivered through ACCUPLACER
- A math assessment developed internally by Mohawk faculty
- Mohawk's Student Entrance Survey.

We also assessed critical thinking as demonstrated in embedded assignments in first- and final-year courses using the Critical Thinking VALUE rubric, one of 16 rubrics created and validated through ongoing assessment research by the Association of American Colleges & Universities (AAC&U).

Our key research questions included:

1. Do students' basic cognitive skills — reading, writing and mathematics — and critical-thinking skills develop between admission and graduation?
2. Do students' attitudes toward learning change between admission and graduation?
3. Is there a correlation between cognitive skill development and students' attitudes toward learning?
4. Can existing post-admission assessment practices be leveraged to facilitate institutional skills assessment?

Unfortunately, due to a disruption to our project funding, concluding this work by completing our summative data analysis was not possible. However, having completed all of our intended activities with the exception of that analysis, our processes and protocols, outcomes of our early discovery process, lessons learned and future recommendations have been assembled in this report to support related research and practice in the future.

This report comprises two sections:

- The first section summarizes our initial, foundational research that informed the development of our research methodology. This section includes:
    - A conceptual literature review of institutional-level assessment

    - An environmental scan of pre- and post-assessments across the Ontario college sector
    - An institutional assessment of our capacity to administer large-scale and online assessments at Mohawk College
    - A summary of a capstone inventory completed at Mohawk College.
- The second section outlines our implementation methodology and preliminary conclusions and includes the typical results summary expected from a research report of this nature. However, it also provides some tentative conclusions in response to our fourth research question, as well as lessons learned and recommendations for future research and practice.

Despite its incomplete nature, we hope these insights into our research implementation and early findings will still prove useful to and help inform the ever-evolving discussions of and debates around skills instruction and assessment, particularly at the institutional level.

Literature Review

Introduction

We completed a literature review to help ground the many facets of our research in key work and ideas on assessment best practices, literacy, numeracy and critical-thinking assessment, (standardized) testing, curriculum-embedded assessment[1] and rubrics, among other topics. We wanted to know more about the history and current landscape of learning outcomes assessment, particularly at a large-scale or institutional level. We wanted to explore how assessment practices and tools vary depending on the specific skills being assessed, as well as how these practices and tools have been or could be demonstrated to be reliable and valid. It was also important for us to ask questions of particular relevance to the Ontario college context, articulating how the themes that emerge across this work might guide our efforts at Mohawk and at colleges across the province. In recognition of a project that is simultaneously global (i.e., piloting a large-scale, institutional assessment process for possible future replication) and local (i.e., building a profile of a Mohawk graduate), our review started big and gradually became more granular, moving from international, national and provincial precedence to the particular skills and instrumentation that comprise the focus of our study and that were shown in our environmental scan to be among the most commonly implemented across the sector. It was by no means intended to be comprehensive, with a number of concepts and practices still to be explored and many questions left unanswered.

What we discovered in completing this review is that, while a number of large-scale assessment initiatives have been attempted over the past two decades (Wagenaar, 2008; Lennon et al., 2014; OECD, 2018a; OECD, 2018b), the success and ease of implementing these initiatives has been inconsistent and often indeterminate, with some projects wrapping prior to their completion (Altbach & Hazelkorn, 2018), this one included. Broadly speaking, large-scale assessment projects are consistently motivated by a desire to produce a standardized means of measuring and reporting on student skill development, but are likewise consistently hemmed in by concerns over their very "standardization," their capacity for representativeness and how their collected data might be misused. This tension has historically played out both internally, within higher educational contexts, and externally, in its interaction with related questions of career preparedness and perceived gaps between the skills graduates are expected to possess and employers' own assessments of the skills they *do* possess as new employees entering the workforce. To some degree, these tensions can be resolved through a shared commitment to the prioritization and improvement of learning (or training, in the case of industry and employers), which suggests, perhaps, that skills/outcomes assessment is always best served with that goal in mind.

At the level of skills, the past two decades have marked a trajectory of first trying to articulate what postsecondary graduates should know and be able to do (i.e., learning outcomes) and then finding ways to assess that knowledge and those skills in a comprehensive, efficient and reportable fashion. Despite the significant work published on the matter, the implementation of large-scale assessment initiatives remains a challenge, even in today's culture of assessment. While there is still a great deal of momentum around learning outcomes and assessment practices both small and large, resistance to these initiatives, as in the case of the suspension

[1] In-course assignments are used as samples of student work/learning for assessment of student skills, often with the use of a validated rubric (sampling student coursework as evidence of their demonstration of skills).

of the AHELO project (Altbach, 2015), could impact future commitment to these ventures worldwide.

The findings of our literature review make clear that some practices seem to work better in certain contexts than others, making it difficult to determine an assessment "gold standard." We hope that by piloting different types of assessments, our research will help inform an understanding of the benefits and challenges of each assessment type and a greater understanding of skills assessment in Ontario colleges in general, which is underrepresented in the literature. For anyone undertaking this work moving forward, it would be useful to explore more targeted research on the assessment of specific skills (especially literacy, mathematics/numeracy and critical-thinking assessment) and best practices that provide developmental opportunities for staff and faculty to "connect accountability efforts with the everyday work and interests of faculty members" (Hutchings et al., 2013).

One gap of particular note in this review relates to the re-administration of post-admission testing with outgoing students. No literature addressed the reliability, validity or even the possibility of using placement tests in an end-of-program, post-assessment capacity. The closest comparator (Bunyan et al., 2015) involved the re-administration of placement tests at the end of a streamed, first-semester course. We have attempted to account for this oversight by taking additional measures to validate our results (triangulating test scores with VALUE rubric assessments and GPAs) and collaborating with the vendor of our tests (College Board). This approach is consistent with the view of validity as being related to the use of the tool and not (necessarily) to the intrinsic properties of the tool itself (Messick, 1988; Cumming & Miller, 2018), acknowledging that psychometric standards like validity and reliability are non-dichotomous, non-neutral (Baird et al., 2017) measures most productively thought of in terms of degree (Cumming & Miller, 2018).

There are still significant questions to address with regards to envisioning and eventually implementing skills assessment at the institutional level. There is the question, for instance, of how exactly to interpret collected data and what threshold(s) might signify sufficient change over the course of a student's time in college or university (Pascarella et al., 2011; Mathers et al., 2018), if any. There is also the question of whether "[g]eneralizing outcomes scores to college learning or even to the quality of higher education" is appropriate (Liu et al., 2012). However, while there is still much that can be learned before the execution of this project, as Blaich and Wise (2011) recommend, "it is important to 'iterate.'" Unresolved issues and questions will provide direction for future phases and research. After all, "large-scale assessment is still in beta mode" (Brumwell et al., 2018).

This reading has helped us to better articulate not only the crucial role of assessment in higher education but also the significant, complex issues that challenge its implementation, which we speak to more directly in the conclusion to this report. In an effort to support our faculty and any other postsecondary institutions interested in pursuing similar praxis, we share our findings as a "top 10" list of practical lessons and recommendations interpreted from the literature. It is clear that the path to institutionalizing valid and reliable skills-based assessment is ambitious and difficult, but in localizing our attempts at the college level and learning from the decades of work that have come before, we can be more prepared to tackle challenges (known and unknown) from an informed, evidence-based perspective.

10 Key Themes Interpreted from the Literature

1. Assessment is important.

Within the literature, assessment is variously positioned as a reporting practice of quality assurance and accountability (Blaich & Wise, 2011); an instructional practice *of*, *for* and *as* learning, increasingly framed in terms of its "authenticity" (Lock et al., 2018); a curricular and instructional design practice informed by outcomes-based education (Tam, 2014), backwards design (Wiggins & McTighe, 2005) and constructive alignment (Biggs & Tang, 2011); a measurement practice guided by psychometric standards (e.g., reliability, validity and fairness) (Cumming & Miller, 2018); and a responsive practice that aims to "close the loop" (Banta et al., 2009; Banta & Blaich, 2011) on evidence-based findings produced through scholarly teaching and, more formally, the Scholarship of Teaching and Learning (SoTL). Assessment matters because its output is valuable both to institutions and students: it helps us know what is working, what needs work and where to best target our time and resources. Assessment also provides students with an opportunity not only to grow within their learning, but also to learn to identify what they do and do not know for themselves — an ability as crucial as the particular skills we seek to measure. If we agree that assessment is important, then assessment of essential skills is doubly so, particularly when more dynamic, adaptable skills and attributes are integral to future success (Weingarten & Hicks, 2018), it is our responsibility to provide our learners with evidence-based information on how these skills develop.

2. Reliable data on postsecondary skill development is in high demand.

Commentaries on a presumed skills gap and the impact of such a gap on the job market have become fairly ubiquitous in recent years (Boden & Nedeva, 2010; Miner, 2010; Arum & Roksa, 2011; Jackson & Chapman, 2012; Dion & Maldonado, 2013; Dion, 2014; Miner, 2014; Weingarten et al., 2015; Social Capital Partners and Deloitte, 2015; Hora et al, 2016; Hora, 2017; Business Council of Canada & Morneau Shepell, 2018; RBC, 2018). While there are some questions regarding the verity of these claims (Brumm et al., 2006; Jackson, 2013; Joy et al., 2013; Strachan, 2016; Craig & Markowitz, 2017; Lackner & Martini, 2017; Goodwin et al., 2019; Kovalcik, 2019) and where the responsibility for such a "gap" might lie (i.e., successfully gaining skills versus articulating them; postsecondary education versus on-the-job training), significant commitments have been made to support increased understanding of how and where learning is happening — and not happening — in our programs of study and workplaces. At present, the skills gap "remains ill-defined, even while the solution is being outlined in considerable detail" (Harrison, 2017). The skills of interest are those featured prominently on the majority of job postings, the variously termed generic, essential, foundational and/or core skills, including the ability to communicate effectively, think critically and problem solve efficiently. Traditionally, institutions have been "more focused on subject- or discipline-specific learning outcomes than the essential or higher-order cognitive skills" (MacFarlane & Brumwell, 2016); however, it is becoming increasingly clear that assessment practices at the course, program and institutional level must move beyond exclusive discipline-specific evaluations to track (in part to better teach) those skills considered essential to future educational and career success. As highlighted repeatedly in HEQCO's reporting on Phase I of the Learning Outcomes Assessment Consortium, "non-disciplinary skills acquisition matters now in a way that it did not a few decades ago" (Deller, 2018).

3. Institutional skills assessment is challenging to implement.

Facilitating culture change is never easy and attempting to shift even some of the focus in higher education away from disciplinary learning to generic skill development and its measurement is bound to be a slow, tentative process. Large-scale skills assessment requires something of a reprioritization, not only at the level of practice, but also in terms of the approach to education and understanding of its purpose. If the results of previous attempts at implementing this work are any indication, this adjustment "remains a difficult and complex task" (Goff et al., 2015), in part because of the dichotomous rhetoric of accountability versus improvement (Baird et al., 2017; Cumming & Miller, 2018) that underlies its exploration. "Accountability" has emerged as a point of perceived tension (Blaich & Wise, 2011) between "top-down" accountability procedures (e.g., internal and public reporting) and more bottom-up practices like scholarly teaching and SoTL. However, as some key figures have noted, the meeting of accountability and (quality) improvement measures holds "potential for positive synergy" (Hutchings et al, 2013) when the interests of both align. This synergy is most successfully fostered through a shared prioritization of "demonstrably improving student learning by assessing it and using the findings to revise programs accordingly" (Banta & Blaich, 2011). This accountability to improve renders assessment much more than either a standardized process of checks and balances or the instigator of future curricular adjustment; it becomes an "ethical and moral commitment" and "pedagogical imperative" (Shulman, 2003).

4. Pre-assessment of skills is most effective when informed by multiple measures

Large-scale assessment initiatives often prioritize post-assessment and the desire to know where our students end up at the end of their diplomas and degrees. However, equally crucial are methods of pre-assessment, which do not only serve as benchmarks for comparison; they have also historically been used to determine student placement and curricular trajectory. Achieving a better understanding of the strengths and limitations of existing assessments is crucial, and the literature on this topic is increasingly clear: Skills-based tests should never be considered in isolation, and while their results may be usefully employed in the targeting of instruction and delivery, remediation practices that disrupt student progress are risky, with potentially significant impacts on retention (James, 2006; Fisher & Hoth, 2010; Medhanie et al., 2012; Scott-Clayton, 2012; Qin, 2017).

5. Indirect measures of student learning are useful but insufficient; direct measurement is imperative.

We like to think that we centre student experiences and feedback within existing practices. In the college sector, for instance, KPI surveys are administered annually, as are numerous internal surveys asking students to qualify their overall satisfaction with their postsecondary experience and provide input on present and future institutional initiatives. However, beyond in-course grading, higher education as a sector has not committed to a structured, embedded process for measuring student learning, and especially the learning of transferable skills needed for gainful employment. Students' satisfaction with their overall educational experience will always be important, but there is no substitute for direct measurement (Sitzmann et al., 2010; Gonyea & Miller, 2011; Patry & Ford, 2016; Brumwell et al., 2018; McCormick & Kinzie, 2018), particularly since self-assessments of the educational experience have been shown to, at best, only weakly correlate to learning gains on standardized achievement measures (Sitzmann et al.,

2010; Bowman, 2010, 2011; Calderon, 2013; Porter, 2013; Lennon & Jonker, 2014; MacFarlane & Brumwell, 2016; Mayhew et al., 2016; Horn & Tandberg, 2018).

6. Standardized tests can help provide reliable benchmarks, but they are not without their limitations.

Standardized tests allow for reliable, quantitative measurement and comparison within and between institutions (Beld, 2015; Cumming & Miller, 2018) and are characterized as being relatively straightforward to implement (Brumwell et al., 2018). However, the use of standardized testing has increasingly incited opposition (Scott et al., 2018), and been cited as ”polarizing,” “toxic” and “controversial” (Brumwell et al., 2018). Standardized tests are additional exercises beyond the scope of student coursework, the stakes for which (when high) can either induce anxiety, potentially impeding performance (Cassady & Johnson, 2002; Chapell et al., 2005; Szafranski et al., 2012; Vitasari et al., 2010; Gerwing et al., 2015), or (when low) risk limiting participation, motivation and effort (Haladyna & Downing, 2004; Banta, 2008; Wise & DeMars, 2005, 2010; Wise et al., 2006; Liu, 2011; Liu et al., 2012; Finn, 2015; Finney et al., 2016; Musekamp & Pearce, 2016; Wise & Smith, 2016; Scott et al., 2018). Furthermore, given the tests’ relative isolation from in-class learning and/or practical environments, many have expressed concerns around their effectiveness and authenticity, questioning their applicability and relevance (Litchfield & Dempsey, 2015; Banta & Pike, 2012; Barrie et al., 2012; Rhodes, 2012; Goff et al., 2015; Scott et al., 2018) and ability to contribute to evidence-based curriculum decision-making (Mathers et al., 2018). Faculty have also expressed concerns about the capacity of standardized tests to support (unintentional) ranking (personal and institutional) and corresponding pressures to “teach to the test” (Scott et al., 2018; Watkins & McKeown, 2018). These tests might then be more useful in or appropriate for certain contexts (e.g., pre-assessment of basic cognitive skills to inform delivery) than others.

7. Rubric-scored, curriculum-embedded assessment promotes active reflection on skills assessment at every level of the curriculum.

While standardized tests have been prominent assessment fixtures for some time, the assessment of existing course materials using tools like the VALUE rubrics[2] may be among the most promising authentic assessment tools for facilitating both small- and large-scale assessment moving forward. Leveraging existing assignments to facilitate rubric scoring eliminates the need for supplemental assessment (Pusecker et al., 2012; Cumming & Miller, 2018) and has been shown to be more cost-effective than vendor-based tests (Goff et al., 2015; Simper et al., 2018). This is not only due to the cost of the tests themselves, but also because there is no need to incentivize participation (Pusecker et al., 2012). Students are already more likely to be motivated to perform, since their work is also assessed as part of their grade (Goff et al., 2015; Cumming & Miller, 2018) and, ideally, aligns with the expectations of future employment and professional contexts (Gulikers et al., 2004; Goff et al., 2015; Litchfield & Dempsey, 2015). Students are also positioned to learn more about both the skills being assessed and assessment itself by participating in this work; embedded assessment practices promote metacognition (Litchfield & Dempsey, 2015). Finally, when used to assess the demonstration of skills in existing course assignments, VALUE rubrics provide immediate and targeted feedback on the specific criteria or component skills with which students succeed or

[2] The Valid Assessment of Learning in Undergraduate Education (or VALUE) rubrics were developed by the American Association of Colleges and Universities (AAC&U).

struggle most consistently, allowing for responsive interventions within curriculum and delivery (Pusecker at al., 2012; Beld, 2015).

8. Assessment results must be understood in context and made actionable both from the top down and the ground up.

More direct, more authentic measures allow for more accurate assessment results in alignment with course, program and even institutional learning outcomes (Pusecker et al., 2012; Goff et al., 2015; Cumming & Miller, 2018). When discussing lessons learned from years of implementing learning outcomes assessment research, Scott et al. (2018) stress the importance of both the support of senior leadership and the need to understand and value local learning contexts, as well as build and support relationships within these contexts. All parties must have a clear sense of how assessment efforts — which can require considerable time and commitment — will eventually impact their day-to-day experiences, and especially how these efforts will impact student learning. Good assessment data should, in equal measure, provide the evidence needed for large-scale cultural change and contextual, pedagogical responsiveness within individual learning environments.

9. Faculty involvement and support are critical to successful, impactful assessment.

The applicability and alignment of authentic assessment practices resonate with faculty and promote faculty engagement (Scott et al., 2018). Likewise, faculty engagement is critical to successful implementation of assessment processes (Deller, 2018). To commit the time and effort required to participate in assessment projects (whether by supporting testing with dedicated class time or actively participating in skills assessment via rubric scoring), faculty need to believe in the work and be convinced of its value to their teaching and to their students' learning. And for this work to matter, its results must be applied directly to faculty interaction with students in the classroom; "[a]n assessment project that does not engage instructors at the course level is unlikely to lead to transformative change in the teaching culture" (Scott et al., 2018, p. 49). Therefore, faculty must be involved at all stages of assessment, from gathering data, to its analysis and application.

10. Crucial questions remain unanswered.

It is still early days for skills assessment, particularly at a larger scale. We are still in the process of figuring out if, how, where and when this assessment can and should occur, and how collected data can be optimized for future improvement. For our part, we are interested in learning more about the (mis)alignment of particular skills and assessment approaches, as well as the teaching and learning of these skills, too. We are also continuously committed to improving assessment processes and the interpretation of assessment results. For instance, when we claim we want graduates to possess certain skill levels necessary for employment, what does this look like, and are these levels consistent for all disciplines and fields? Is any growth across an academic program significant? Or, if we aim for skill thresholds, how and by whom are such thresholds determined, and might they oversimplify the complexity of these skills and the centrality of lifelong learning by suggesting certain attainment is sufficient? By prioritizing skills assessment, might we inaccurately or inappropriately be generalizing skills, or isolating them from their contexts in ways that could be misrepresentative? Are some skills inherently disciplinary and should they therefore be assessed only within disciplinary contexts? What are the resource demands for skills assessment and how can it be supported on an

ongoing basis? Can individual skills truly be isolated or are different skills (e.g., critical thinking and motivation) inextricably linked? These questions are complex but productive, inviting us to consider student learning in increasingly rigorous ways.

Pre- and Post-admission Assessment Practices at Ontario Colleges

Background

In addition to reviewing the literature, we built on Fisher and Hoth's (2010) review of the practices and instruments used in language proficiency assessment in Ontario to complete an environmental scan in early 2018 that determines the specific pre- and post-admission assessment tools and practices currently employed at all 22 English-language Ontario colleges.[3] The goal of this scan was to expand on and update this work, in part by including both mathematics/numeracy and literacy assessment, with a particular focus on determining which pre- and post-admission assessments of the wide range available (see **Appendix A**) are used most consistently across the sector. With the advent of the COVID-19 pandemic in March 2020, it is now more likely than ever that these tools and practices have continued to change since this scan was completed, and perhaps significantly, due to the need for online and remote operations. However, understanding the status of these activities prior to the pandemic is useful for understanding how the sector is changing over time.

Methodology

Over 70 staff, faculty and administrators at testing centres and registrar and admissions offices were consulted during the preparation of this report. Respondents were asked about assessment tools and practices at their institution within a pre-/post-admission framework, using the following questions as a guideline:

- What pre-admission testing is administered (if any)? Which assessments are used?
- What post-admission testing is facilitated (if any)? Which assessments are used?
- Are there any specific assessments used for ESL students? If so, which assessments and when are they administered?
- Are there any program-specific assessments used for particular academic areas? If so, which assessments and when are they administered?
- Are there any other institution-wide assessments administered?

These questions intentionally targeted testing tools as these were of primary interest to the research team given the project's scope. Researchers followed up with more targeted queries to clarify ambiguous responses; additional details were confirmed with information publicly available on institutional websites. Over time an additional framework of four potential assessment categories (placement, equivalency, ESL and program-specific assessments) was adopted.

[3] Collège Boréal and La Cité were excluded from this scan as the researchers felt that having a predominantly Francophone demographic would impact the choice and use of assessments at these institutions in a way that greatly differs from other colleges in the province.

Key Findings

Pre- and Post-admission Assessment

The resulting output of these discussions is summarized in Tables 1 and 2. All 22 English-language colleges are represented in the cumulative data. The results of the environmental scan were supplemented by previously assembled data on student entrance surveys collected in the 2017 Institutional Research Department Inventory undertaken by the Heads of Institutional Research (HIR) Coordinating Committee.

Table 1: Number of Colleges Using Pre- and Post-Admission Assessments by College Size

Assessment Type	Small (7)	Medium (7)	Large (8)	All Colleges (out of 22)
Pre-admission	**6**	**7**	**8**	**21**
Equivalency	6	7	8	21
ESL	1	3	3	7
Program-specific	-	6	5	11
Post-admission	**-**	**2**	**8**	**10**
ESL	-	1	5	6
Placement	-	2	7	9
Student entrance survey	**4**	**6**	**6**	**16**
No Assessments	**-**	**1**	**-**	**1**

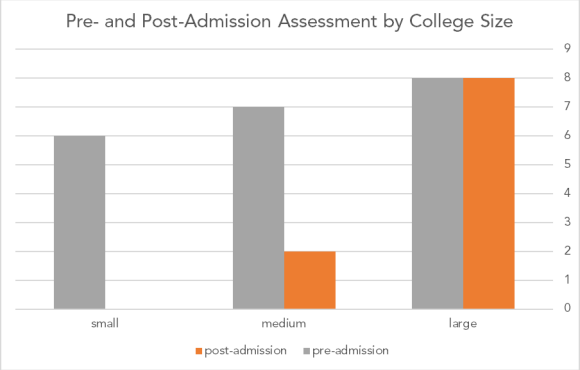
Note: College size groupings refer to categorization used by ONCAT

For the purposes of this report, pre-admission assessments are those administered between application and admission. Post-admission assessments are those administered after successful admission and before classes begin and include assessments most commonly used to stream or place students based on their skill level (i.e., placement tests).

Based on the results of the scan, 21 out of 22 English colleges conduct some form of pre-admission testing. Some of this testing is implemented for mature students and/or students who do not meet admissions requirements, which we refer to as “equivalency testing” throughout this report. Pre-admission testing can also be program-specific, administered most commonly for competitive programs. According to our data, this type of testing is not administered at any small Ontario colleges but is implemented for at least one program at half of the medium and large colleges. Finally, pre-admission testing can be used to assess ESL students who apply without specified English-language proficiency certification. This testing is more prominent at large institutions.

By contrast, post-admission assessment is used half as often as pre-admission assessment (see Figure 1). While post-admission assessment occurs at all eight large institutions, seven of which specifically make use of placement testing, only two of seven medium colleges report using placement testing as a component of post-admission assessment. Post-admission assessment does not occur at any small institutions.

Figure 1: Number of Colleges Using Pre- and Post-admission Assessments by College Size



These initial findings could suggest that small institutions lack the capacity to facilitate more widespread and comprehensive assessment, opting to prioritize equivalency testing over placement testing. Stakeholders at small colleges described a number of alternatives to post-admission assessment including:

- In-course pre-assessment practices facilitated by faculty and coordinated at the program, school or faculty level in lieu of more formalized testing
- Replacing testing with post-admission surveys that incorporate self-assessment measures of skills and competencies.

A few colleges also discussed recent decisions to phase out standardized post-admission assessments in consideration of cost-benefit analyses and anecdotal evidence questioning the effectiveness and impact of this testing.

With only nine colleges currently conducting placement testing with larger percentages of incoming students, there is the question of how well the proposed methodology for integrating an institutional assessment framework would scale at other institutions across the province. To facilitate such assessment, the majority of colleges would either need to introduce post-admission assessment in the months leading up to the start of the school year or would need to integrate such assessment within individual first-semester courses (a method this project also intends to explore). Additionally, no colleges made mention of any skills-based exit testing,[4] so any effort to incorporate these types of assessments into large-scale, institutional assessment frameworks would likely need to be built from the ground up.

The use of post-admission formal assessment of language proficiency appears to have decreased significantly compared to Fisher and Hoth's findings (2010). According to the data assembled in this scan:

[4] Fisher & Hoth (2010) did report on "exit testing" in their inventory of college-level literacy practices. They found that overall, 25% of colleges in Ontario utilized some form of exit testing, with that number increasing to 57% in large colleges (though "[o]nly four colleges reported rigorous practices in exit testing that also replicated their formalized entry-level processes"). However, they define exit testing as "a measure or indicator of language proficiency following some form of language training," (i.e., testing that occurs after the completion of transcript or modified communications courses or some combination thereof).

- 45% of colleges use some form of post-admission formal assessment of language proficiency, compared to 62% in 2010
- 14% of colleges assess using writing samples only, compared to 33% in 2010
- 5% of colleges assess using computerized assessment of reading and/or sentence skills only, compared to 20% in 2010
- 27% of colleges assess using multiple measures (both writing samples and computerized reading assessments), compared to 47% in 2010.

The cause of this shift could be attributed to methodological factors like different surveying practices and categorization. Table 1 also includes a tally of the number of colleges that currently make use of a student entrance survey. Qualitative commentary indicates that there is considerable variance in the content of these surveys, from base-level demographics to more extensive self-surveying of learning attributes and competencies. No data was collected on the use of corresponding exit surveys.

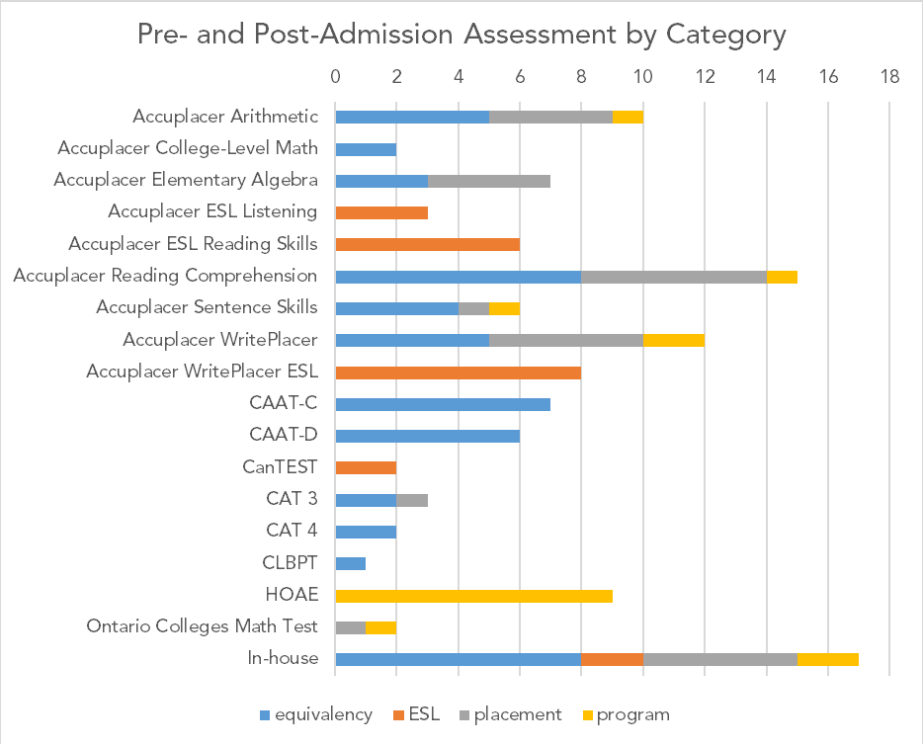
Table 2: Number of Colleges Using Specific Testing Tools by College Size

Test Name	Small (7)	Medium (7)	Large (8)	All Colleges (22)
In-house tests	**-**	**3**	**8**	**11**
ACCUPLACER	**2**	**2**	**7**	**11**
Arithmetic	1	1	4	6
College-level Math	1	-	1	2
Elementary Algebra	-	1	4	5
ESL Listening	-	-	3	3
ESL Reading Skills	1	2	3	6
Reading Comprehension	2	2	6	10
Sentence Skills	1	1	2	4
WritePlacer	1	1	6	8
WritePlacer ESL	1	2	5	8
Bennett Mechanical Comprehension Test	-	**-**	**1**	**1**
Cambridge Michigan Language Assessments	-	**-**	**1**	**1**
Canadian Adult Achievement Test	**1**	**2**	**1**	**4**
CAT 3	1	1	1	3
CAT 4	-	2	-	2
Canadian Adult Achievement Test	**5**	**3**	**4**	**12**
CAAT-C	4	2	1	7
CAAT-D	2	1	3	6
Canadian Language Benchmarks Test	**-**	**-**	**1**	**1**
CanTEST	**-**	**1**	**1**	**2**
Health Occupation Aptitude Examination	**-**	**6**	**3**	**9**
National Literacy Secretariat Writing Sample	**-**	**1**	**-**	**1**
Ontario Colleges Math Test	**-**	**-**	**2**	**2**

Assessment Instruments and Vendors

The most commonly used tests at Ontario colleges are the Canadian Adult Achievement Test (both the CAAT C and CAAT D, collectively used by 12 colleges primarily for equivalency testing), the suite of literacy and mathematics/numeracy tests produced by ACCUPLACER and assessments created in-house by individual institutions. The Health Occupation Aptitude Examination (HOAE), a mandatory assessment of academic aptitude, spelling, reading comprehension, science and vocational aptitude for many health programs, also features prominently. Tallies of all tests/tools by college size and testing category or purpose can be found in Table 2 and Figure 2.

Figure 2: Number of Colleges Using Pre- and Post-admission Assessments by Category



The distribution of individual tests by college size is consistent with the overall distribution of pre- and post-admission assessment practices. Program-specific testing and placement testing are conducted almost exclusively with ACCUPLACER and in-house tests, with the exception of the use of the CAT-3 by one college and the Ontario Colleges Math Test (OCMT)[5] by two colleges.

In-house assessments are most frequently used to assess mathematics/numeracy skills, which is consistent with current practices at Mohawk College. Of the 11 colleges that use in-house assessments as a component of their pre- and post-admission testing, nine use in-house tools for mathematics/numeracy assessment. Writing assessments were the next most common in-house assessment (six of 11 colleges), with varying practices for demonstrating and grading

[5] The OCMT was designed by Humber College in collaboration with Vretta and therefore could be considered both an in-house and vendor-based assessment. For the purposes of this data collection, it has not been included as an in-house assessment.

written work. Figure 3 outlines the distribution of in-house assessments by subject for all Ontario colleges. This distribution is roughly mirrored by the overall subject/content distribution of testing across all colleges, as shown in Figure 4.

Figure 3: Distribution of In-house Assessments by Subject for all Ontario Colleges

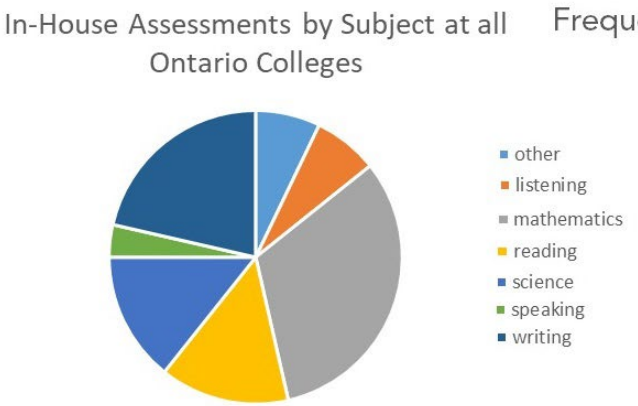
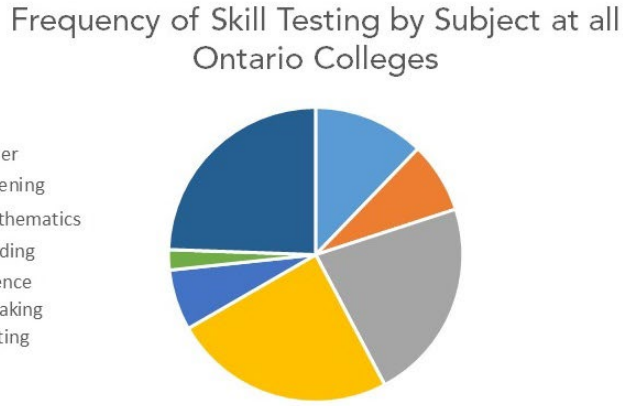


Figure 4: Frequency of Skill Testing by Subject for all Ontario Colleges



The snapshot presented above of pre- and post-admission practices in the Ontario college sector could be further refined by soliciting more qualitative data to determine why institutions are increasingly opting to forgo post-admission assessment; why in-house assessments are more commonly implemented to assess for math or numeracy levels than they are for skills like reading; whether assessments are administered online or in person or both; and, in the case of online assessment, whether test administrators and students have encountered similar challenges with the registration, scheduling and technology as outlined in the final section of this report.

Operational Assessment

Background

With a better understanding of assessment research and best practices at the international, national and provincial levels and an updated inventory of assessment tools and practices being implemented at Ontario colleges, we proceeded with internal reviews of current assessment practices at Mohawk, including an assessment of our operational capacity to deliver math, reading and writing tests online and for the graduating cohort.

One of the underlying motivations for this project was to investigate ways to facilitate a more streamlined, efficient and student-friendly survey and assessment experience (particularly the reading, writing and math assessments, plus the Student Entrance Survey). In anticipation of Phase II pilots, we felt it was important to first take stock of existing survey and assessment processes and examine their capacity for modification and/or adaptation.

Key Findings

As mentioned earlier in this report, pre-admission testing is mainly reserved for mature students, students who do not meet admission requirements or those applying to select programs such as

Nursing. Once accepted into a program, the majority of students are required to complete some combination of reading, writing and mathematics assessments to determine their level of comprehension and/or the specific course(s) for which they are eligible to register (when relevant to placement decisions). About 70% of Mohawk students participate in this process, completing a combination of pre-placement tests known as Assessments for Success (or AFS); degree and certificate students are not required to participate. With the exception of two remedial tests, the current reading, writing and math placement exams are available in a digital format, but are not accessible without prior registration, a voucher number or login credentials. Only a select number of programs require students to complete pre-admission testing.

Table 3: Pre- and Post-Admission Assessments at Mohawk College in 2018–2019

Learning Outcome	**Tests Available**	**Vendor/Delivery Method**	**Type**
Reading	CAAT-C (reading, spelling and vocabulary	Paper & Scantron	Pre-admission
	ESL Listening Test	ACCUPLACER	Pre-admission
	ESL Reading Skills	ACCUPLACER	Pre-placement
	Reading Comprehension	ACCUPLACER	Pre-placement
Writing	WritePlacer	ACCUPLACER	Pre-placement
	WritePlacer ESL	ACCUPLACER	Pre-placement
Mathematics	Grade 10 Math (PSW, trades)	Paper & Scantron	Pre-admission
	Comprehensive Tech Math (15 modules)	MapleTA	Pre-admission
	Technical Math (10 modules)	MapleTA	Pre-placement
	Business Math (first 5 modules)	MapleTA	Pre-placement

ACCUPLACER

Currently, Mohawk uses a suite of tests from ACCUPLACER to assess reading and writing post-admission. These tests are primarily adaptive computer tests that select questions based on the taker's previous responses. ACCUPLACER places restrictions on the use of their tests, including the need for proctoring.[6] As well, institutional credentials must be used to access and activate the tests. Currently, placement decisions are made based on benchmarks or cut scores, which are determined by a validated algorithm that is internally and iteratively developed. Enrolment recommendations for first-semester communications courses are generated based on a combination of students' WritePlacer and Reading Comprehension scores. Details regarding the benchmark scoring employed at Mohawk for both first-language and second-language testers are provided in Table 8. In terms of adapting the use of ACCUPLACER placement tests for summative, end-of-program assessment, Mohawk's benchmarks do include levels of exemption from enrolment in communications courses, which could provide a useful threshold for interpreting exiting students' language test scores.

[6] At the outset of this project, this was limited to in-person, on-site proctoring by an "authorized test administrator." However, on account of the partnership between ACCUPLACER and Examity, we were able to explore remote proctoring of these assessments for the first time as part of this project.

Table 4: ACCUPLACER Benchmarks at Mohawk College

First-language Benchmarks		
COMM 11040 Placements	COMM LL041 Placements	COMM LL041 Exemption
WritePlacer 1, 2, 3	WritePlacer 4 and Reading Comprehension ≥ 80	WritePlacer 7 and Reading Comprehension ≥ 110
WritePlacer 4 and Reading Comprehension < 80	WritePlacer 5 and Reading Comprehension ≥ 60	WritePlacer 8
WritePlacer 5 and Reading Comprehension < 60	WritePlacer 6 and Reading Comprehension ≥ 50	
WritePlacer 6 and Reading Comprehension < 50	WritePlacer 7 and Reading Comprehension < 110	

Second-language Benchmarks	
COMM LL043 Placements	COMM LL044 Placements
WritePlacer ESL 1, 2, 3	WritePlacer ESL 4 and ESL Reading Comprehension ≥ 90
WritePlacer ESL 4 and ESL Reading Comprehension < 90	WritePlacer ESL 5, 6

Although the process of writing AFS is well established for incoming first-year students, one of our challenges was to find a way to ensure that graduating students write the AFS in a proctored environment. The ACCUPLACER tests have been used in previous HEQCO research (Bunyan et al., 2015) to assess students at the end of courses (in this case at the end of developmental communication courses for which students were streamed or placed). Based on this experience, discussions with staff at the testing centre and consultations with ACCUPLACER/College Board, the research team developed a strategy to administer tests to the graduating cohort. Essentially, our options were to have graduating students write in a live-proctored classroom, the testing centre or a remote-proctored environment, which still requires coordination through the testing centre in order for each student to gain access to the test. Ultimately, we adopted the latter two options when faculty were not able to afford the time needed to have students complete the tests in class. Further details about the challenges of administering this test are outlined in the final section of this report.

In-house Math Assessment

Mohawk's math assessment, created by Mohawk faculty, was designed as an in-person, proctored assessment delivered through internal software (MapleTA). However, if students are registered in advance and given a username and password, the test can be administered online and off-site by adjusting IP settings and coordinating access and automatic credential verification through our Learning Management System (LMS). In partnership with the Centre for Teaching and Learning, we investigated various systems that can be purchased to proctor online tests and ultimately committed to trialing Examity, which was also used to remotely administer the ACCUPLACER assessments, albeit through a separate system.

Similar adaptive work would need to be completed with the math assessment to determine how scores typically used for placement purposes could be interpreted as evidence of students' skill levels and skill growth in a non-placement, post-assessment context.

Student Entrance Survey

Mohawk has been administering some form of the Student Entrance Survey since 2005. The survey was moved online in fall 2009. The majority of students complete the survey during AFS, though they also have the opportunity to independently complete it online during the first few weeks of classes. The survey recently underwent extensive revisions in fall 2017 and again in winter 2018. Among these revisions were the inclusion of attitude and self-perception statements based on the Ryff Scales of Psychological Well-Being (Wabash) and alignment of the survey's demographic questions with the Ontario Institute for Studies in Education's standard.

Mohawk does not currently administer an exit survey, other than the Key Performance Indicators (KPI) survey administered at all Ontario colleges. This project piloted the re-administration of key components of the entrance survey (e.g., the Ryff Scales mentioned above) as a Student Exit Survey with graduating students in order to compare how students' self-assessment of their attitudes and approaches to learning change over time.

Capstone Inventory

Background

In addition to an operational assessment, we also worked collaboratively with stakeholders across the college to generate an inventory of all current, final-semester experiential learning activities undertaken at the college (with a focus on final-semester capstones) and how they link to our first-semester courses and Institutional Learning Outcomes (ILOs), as well as the means by which they are assessed.

We pursued this data because capstones have been identified as high-impact practices (Hauhart & Grahe, 2015) that provide well-aligned, authentic opportunities for assessment. Capstones require students to integrate, synthesize, apply and refine knowledge and skills learned across a program of study (Ahlawat et al., 2012; Appleby et al., 2016; Fairchild & Taylor, 2000; French et al., 2015; Kerka, 2001; Kiener et al., 2014), particularly skills like critical thinking and communication (Haave, 2015) — the focus of our research. Capstones have also been shown to support students' transition from academic study to the workforce (Fairchild & Taylor, 2000; French et al., 2015; Kiener et al., 2014; Kinzie, 2013). Using capstone assignments in the context of institutional learning outcomes assessment and curriculum-embedded assessment in particular is also consistent with best practices. For instance, the AAC&U recommends scoring capstone-adjacent "signature assignments" or "signature student work," i.e., "work used to assess students' cumulative learning at the end of a semester or at graduation" (Drezek McConnell et al., 2019).

Methodology

In collaboration with stakeholders working on capstone projects across campus, a working definition of "capstone"[7] informed by a literature review was developed. This definition was used

[7] Our working definition aligns well with the official definition of "capstone" recently adopted by the college as part of the work of the Centre for Community Partnerships and Experiential Learning: "A cumulative activity in the final semesters of a program that is based significantly on knowledge and skills acquired in earlier course work. It involves a creative, iterative and often open-ended process using problem-based learning to address a project challenge. Students spend a significant amount of time, working independently or in a team environment, throughout the semester and translate their results using written reports, oral presentations or poster presentations. Projects can involve qualitative or quantitative research."

to frame and direct a program-level audit of capstones at the college. Collectively, we defined capstone as:

> "a culminating experience that integrates learning at the end of a program of study. It is a way for a student to demonstrate the breadth and depth and of their learning and the skills they've developed during the program of study. A capstone supports transition to career by connecting academic experience to future professional experience. It may:
>
> - Be an assignment, a group of assignments, or a full course
> - Be a paper, a presentation, a project or an industry-related product
> - Be an individual or group assignment
> - Include industry connections or an experiential learning component."

This definition was shared with all 13 associate deans at Mohawk, along with pre-populated, template spreadsheets for all of their programs of study. They were asked to provide the following information in each spreadsheet:

- **Course Lead:** Name of the faculty leading the course (currently)
- **Includes a Capstone:** Indicate with a "Y" the courses that include a capstone (based on the definition provided above)
- **Link to Institutional Learning Outcomes (ILOs):** List the most prominent of the five ILOs — communicator, critical thinker, continuous learner, collaborator and responsible citizen — that are assessed in the course
- **Capstone Assessment Details:** Provide any relevant information regarding the assessment methods used in the course (e.g., student artifacts created, assessment tools used)
- **Potential Link in First Semester:** Indicate the course code of any first-semester courses where a similar assessment tool/method could assess the same ILO (in order to measure that metric of learning from intake to graduation).

Associate deans were encouraged to work with program coordinators and other faculty members to complete this work and help identify capstone courses across their curriculum. They were given a timeline of two weeks to prepare the data; most units required more time. Much of the data represented here was the product of conversations that occurred in the ensuing weeks after the initial request was distributed.

Key Findings

Table 5: Number of Capstones by Department; Percentage Distribution of Capstones (as identified by the academic areas)

Department	Number of Capstones	Percentage of Capstones
Allied Health	7	4%
Building, Construction, Electrical & Energy	10	5%
Business and Media Graduate Studies, Applied Research, Entrepreneurship	19	10%

Business	31	16%
Chemical, Mechanical and Aviation	4	2%
Community Studies	8	4%
Computer Science and Information Technology	20	10%
Construction and Building Systems	6	3%
Health Studies	3	2%
Liberal Studies	6	3%
Media and Entertainment	62	32%
Nursing	2	1%
Social Services and Justice Studies	18	9%
Total Number of "Capstones"	**196**	

Together, associate deans, program coordinators and faculty identified 196 capstone projects or courses across Mohawk's programs of study. While Table 5 provides a breakdown of the number of capstones per department, as well as the percentage distribution of capstones across departments, this data has since been refined in partnership with Mohawk's Centre for Community Partnerships and Experiential Learning to reveal that only 18% of these projects/courses qualify as true capstones based on ministry definitions and our own unique nomenclature, with the other 82% representing other final-semester experiential learning opportunities within the curriculum. However, we present our collected data unaltered, since, for the purposes of this research, non-capstones with experiential learning components and the opportunity for culminating, end-of-program assessment were just as valuable for potential piloting as "true" capstones.

Table 6: Frequency of Institutional Learning Outcome Assessment in Capstones by Department

Department	Communicator	Collaborator	Critical Thinker	Continuous Learner	Responsible Citizen	All
Allied Health	100%	100%	100%	0%	0%	0%
Building, Construction, Electrical & Energy	0%	0%	50%	20%	0%	20%
Business and Media Graduate Studies, Applied Research, Entrepreneurship	74%	68%	68%	47%	11%	0%
Business	45%	16%	45%	23%	6%	19%
Chemical, Mechanical and Aviation	0%	0%	100%	0%	0%	0%
Community Studies	25%	0%	25%	38%	38%	0%
Computer Science and Information Technology	0%	70%	100%	0%	0%	0%

Construction and Building Systems	67%	33%	67%	0%	0%	33%
Health Studies	33%	33%	0%	67%	0%	0%
Liberal Studies	50%	0%	0%	33%	0%	17%
Media and Entertainment	27%	15%	50%	27%	0%	0%
Nursing	0%	0%	50%	50%	0%	0%
Social Services and Justice Studies	22%	17%	39%	6%	0%	17%
All Departments	**34%**	**28%**	**55%**	**22%**	**4%**	**7%**

Respondents were asked to identify the single most prominent ILO for each capstone; however, many indicated that multiple ILOs (and sometimes all) were assessed in capstones, which is not surprising since capstones are intended to be cumulative, comprehensive final assessments. The most commonly identified ILO was critical thinker, assessed most prominently in 55% of capstones. The second mostly frequently cited ILO was communicator (34%), followed by collaborator (28%) and continuous learner (22%). Responsible citizen was the least represented of all the ILOs, assessed most prominently in only 4% of capstones. The percentages for all individual departments are included in Table 4. The distribution of ILO assessment across departments has also been provided in Table 5. (Note: the percentage distribution is skewed by the dominance of certain programs.)

Table 7: Distribution of Institutional Learning Outcome Assessment by Department

Department	Communicator	Collaborator	Critical Thinker	Continuous Learner	Responsible Citizen	All
Allied Health	11%	13%	6%	0%	0%	0%
Building, Construction, Electrical & Energy	0%	0%	5%	5%	0%	14%
Business and Media Graduate Studies, Applied Research, Entrepreneurship	21%	24%	12%	20%	29%	0%
Business	21%	9%	13%	16%	29%	43%
Chemical, Mechanical and Aviation	0%	0%	4%	0%	0%	0%
Community Studies	3%	0%	2%	7%	43%	0%
Computer Science and Information Technology	0%	26%	19%	0%	0%	0%
Construction and Building Systems	6%	4%	4%	0%	0%	14%

Health Studies	2%	2%	0%	5%	0%	0%
Liberal Studies	5%	0%	0%	5%	0%	7%
Media and Entertainment	26%	17%	29%	39%	0%	0%
Nursing	0%	0%	1%	2%	0%	0%
Social Services and Justice Studies	6%	6%	6%	2%	0%	21%

Perhaps the most challenging aspect of this work was identifying an assignment in first term where critical thinking is (as) prominently demonstrated and assessed as it is in the final term — in other words, determining how — and whether — we could assess the same skill in both the first and final year. This proved to be a tricky proposition, with only 44% initially identifying a natural linkage, and even fewer linkages identified upon closer examination. As we dug deeper into the curriculum and more carefully examined specific assignments, we discovered that, in many cases, critical thinking is a skill reserved for final-term curriculum and not necessarily scaffolded throughout programs of study. In order to select an "incoming" work sample, we sometimes had to look to second term or stretch our expectations in order to include a first-term assignment not as well aligned with the performance criteria as would have been ideal. These challenges, however, were met with an eager, forward-looking desire to develop curriculum with skills in mind — a significant outcome of this research and one the college will continue to pursue beyond it.

Finally, Table 8 represents a preliminary attempt at categorizing the different types of capstones collected in the inventory, which, in fact, align quite well with different categories of experiential learning since recognized in Mohawk's official nomenclature. Portfolios or related reflective, goal-setting work was the most commonly identified category of capstone, followed closely by various project-based learning. Somewhat surprisingly, placements or practicums were among the least cited capstones, occurring only 8% of the time. (Note: Some capstones are represented more than once if more than one type of component is included in the assessment. Again, the distribution is skewed by the dominance of certain programs.)

Table 8: Distribution of Capstone Type at Mohawk College across All Departments

Type of Capstone	Distribution
Portfolio/Goal Setting/Interview/Mentorship	16%
Project	15%
Presentation/Performance	13%
Report/Article/Story/Publication	11%
Experiential (Solo or Group)	9%
Film/Television	9%
Plan/Proposal/Pitch/Campaign	9%
Placement/Practicum	8%
Simulation/Scenario/Role Play/Case Study	8%

As alluded to above, many of the faculty, program coordinators and associate deans involved in the collection of this data struggled with the capstone definition and the process of identifying capstones within their curriculum. The idea of a capstone is still unclear for many people, especially given how applied much of the learning is in college. The cause of the most confusion and perhaps the most common misattribution of the “capstone” descriptor were various experiential learning assignments embedded throughout the curriculum, which many faculty members found difficult to distinguish from “true” capstones. Many respondents seemed to inherently associate capstones with hands-on, practical, authentic learning, which, while not untrue, does not necessarily constitute a proper capstone if this learning is not occurring nearer to the end of the program and does not constitute a “culminating” experience. We also learned, not surprisingly, that the type of culminating activities varied greatly across the college.

However, the debates and discussions around (mis)identifying capstones were found to be productive and engaging. Looking ahead to future impacts (beyond the scope of this project), it should be noted that the reflective process of analyzing curriculum for capstones draws attention to the presence — or absence — of capstones in current curriculum, which could lead programs to consider incorporating capstones more purposefully — a process that would be best facilitated during cyclical program review (which happens every five years at Mohawk). This was an unintended outcome of this work that could have a great impact on student learning, with academic units engaging in discussions on the role of capstones within the curriculum, the current and evolving definitions of capstone within specific disciplines, the value of capstones to assessment across programs and the willingness and capacity to engage capstone courses or courses with capstone projects as potential sites for pilot-testing of various assessment methods (e.g., adaptation of validated rubrics, testing, etc.).

Research Protocol: Phase II Pilots

We developed a framework to allow for the large-scale measurement of student learning outcomes at entry to and exit from the college. A sample of participants was drawn from four academic areas across 11 courses. Data generated from Mohawk’s Assessments for Success (AFS) would be compared to collected data for outgoing students in two and three-year diploma programs retaking the reading, writing and mathematics tests. This comparison would provide some preliminary information about cognitive differences between entering and exiting college students, as well as insight into the reliability, validity and feasibility of re-administering placement tests prior to graduation. Additionally, the critical thinking of incoming and exiting

students would be measured by having trained scorers assess student work samples using the Critical Thinking VALUE rubric. These work samples would also be assessed using the Written Communication VALUE rubric as a means of triangulating the reading and writing test scores, further validating our results.[8] Overall, this protocol was designed to be forward-looking, incorporating a diverse set of pilots to compare the relative value and institutional fit of different assessment practices, while also laying the groundwork for future generic skills assessment initiatives.

Background

Research Questions

1. Do students' basic cognitive skills[9] — reading, writing and mathematics — and critical-thinking skills develop between admission and graduation?
2. Do students' attitudes toward learning change between admission and graduation?
3. Is there a correlation between cognitive skill development and students' attitudes toward learning?
4. Can existing post-admission assessment practices be leveraged to facilitate institutional skills assessment?

Project Design

Phase II of this project involved cross-sectional studies of both new and existing assessment processes toward the future development of an institutional generic skills assessment process. Given the one-year timeline for this phase of the research, we tested a cross-section of an incoming cohort (in this case across two separate intakes, September and January) and a different (concurrent) outgoing cohort of students. Mohawk's current post-admission assessments — AFS — were leveraged to provide measurements of incoming students' reading, writing and mathematics skills, with the same tests re-administered on a pilot basis to outgoing students as a measurement of their equivalent skills prior to graduation. As outlined in the "Operational Assessment" section, Mohawk uses the ACCUPLACER Reading Comprehension and WritePlacer tests (as well as ESL Reading Skills and WritePlacer ESL) to assess incoming students' reading and writing skills and an in-house assessment designed and refined by our mathematics faculty led by Professor Craig Cooke to assess students' mathematics skills. Incoming students also complete Mohawk's Student Entrance Survey, which incorporates attitude statements from the Ontario College Student Engagement Survey (OCSES), as well as the Ryff Scales of Psychological Well-being, a theoretically grounded tool used in the Wabash National Study. A portion of the Student Entrance Survey was re-administered to students participating in the pilots to help track the development of those

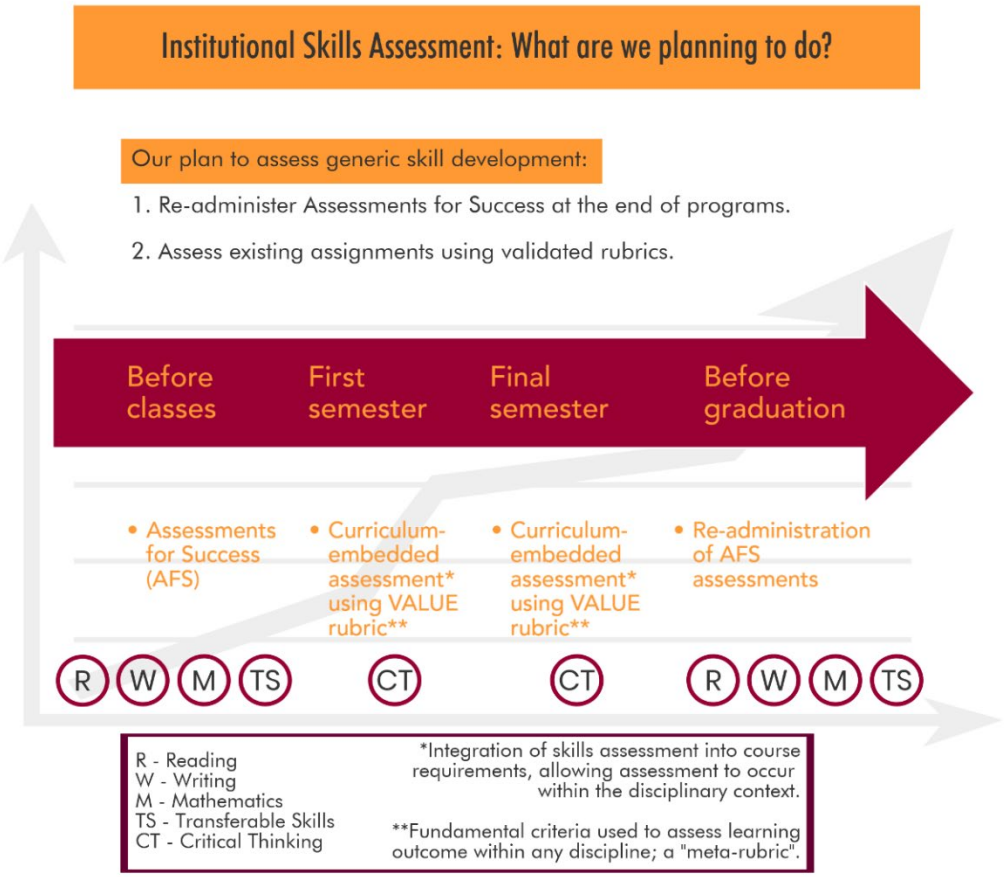
[8] Since our assessment tools (i.e., the ACCUPLACER tests and in-house math assessment) were designed to assess students' readiness to enter college and not their skill levels upon completing college, it is important to validate the new application of these tools; the additional assessment data provided by curriculum-embedded assessment of literacy is intended to assist in this effort.

[9] Much of the discussion around the so-called "skills gap" in Ontario (and beyond) focuses on the basic cognitive skills of literacy and numeracy, even though "[b]asic cognitive skills are seen to be a prerequisite for PSE but not necessarily an outcome of it" (Deller et al., 2015). Despite considerable anecdotal evidence, "[f]ewer tools are available to measure the value-added of a postsecondary education" (Deller et al., 2015) as it relates to reading, writing and mathematics. While these skills remain central to future employability and careers success, they are not traditionally conceived of as curricular components of a "higher" education. The instruction of these skills is likewise often "unsystematic" and their importance is not always articulated, nor are they always assessed (Deller et al., 2015).

transferable skills that support disciplinary and generic skill development[10], along with a brief post-assessment questionnaire targeting student motivation, effort, confidence, commitment and comfort when taking the tests.

In addition to these "testing pilots," we also piloted the use of curriculum-embedded assessment of student work samples from first- and final-term courses using the Critical Thinking VALUE rubric. We opted to use an embedded assessment method in part because critical thinking is not currently assessed as part of AFS or otherwise at the institutional level, but also because the literature suggests that the use of "real" assignments for skills-based assessment and the application of a validated rubric allows for more authentic representation of those skills (Baird, 1988; Banta & Pike, 2012; Tremblay et al., 2012; Litchfield & Dempsey, 2015), particularly a complex skill like critical thinking. All Phase II "rubric pilots" took place in existing courses during the fall 2018 and winter 2019 semesters.

Figure 5: Distribution of Capstone Type at Mohawk College across All Departments



Project design decisions were informed by the results from our first phase of research, prioritizing feasibility, availability of interested faculty, scheduling and representative diversity of programs and were adapted on an ad hoc basis throughout implementation. In April and May of

[10] "Transferable skills" are inconsistently defined across the literature, used to denote everything from attitudes toward learning and non-academic competencies that support learning (self-assessed in the completion of Mohawk's Student Entrance Survey) to the entire swath of essential employability or generic skills (an equally indeterminate term and beyond the scope of this particular measurement).

2018, the project faculty lead met with representatives from 44 programs of study identified during the capstone inventory. The list of possible participants was narrowed based on a number of factors including availability of personnel, interest and scheduling. The greatest limiting factor for pilot selection was faculty interest/availability; administrators and/or faculty who demonstrated a particular interest in the data collection of Phase I were contacted and invited to discuss the possibility of joining the project in Phase II.

Following extensive consultations, four programs were selected to participate, representing four of Mohawk College's six schools or academic areas. Two additional programs were added in winter 2019 to help ensure we reached our commitments for student participation. The use of convenience sampling to determine participating programs will limit the degree to which our conclusions can be generalized to the larger student population, but was crucial to securing a team of dedicated, enthusiastic participants.

Following detailed curricular and assessment discussions, the project faculty lead, associate deans and participating faculty worked through the identification of and, in some cases, justification for, faculty selection, course selection, assignment selection, rubric preferences and scope of involvement as a component of a fully supported orientation process to help prepare them and their students for participation in these assessment pilots. Supporting documentation is provided in **Appendices B, C** and **D**.

Initial Hypotheses

In regard to our final research question targeting the process and, in particular, the efficacy of administering this type of assessment, we anticipated that, while our current post-admission practices could certainly be used as comparators for later, end-of-program assessments, the value of these comparisons would be limited by the purpose for which these assessments were originally designed (i.e., as placement tests to assess students' readiness to enter college, as opposed to more general assessments of skill level or learning). With this in mind, while we certainly hoped to identify growth in students' skill development and further refinement of their attitudes toward learning (i.e., research questions #1 and #2), we recognized that our instruments and methodology (i.e., cross-sectional study, convenience sampling) limit the degree to which these findings are generalizable.

Despite these limitations, we did anticipate finding evidence of skill development across students' time in college, while also acknowledging that *any* results would help identify areas for improvement and future development and inform our institutional practices and pedagogies moving forward. Finally, there exists extensive research on the verity of research question #3 regarding the correlation between cognitive development and attitudes toward learning; we anticipated finding similar correlation in our study.

Methodology

Participant Eligibility and Recruitment

The primary focus of Phase II was to pilot and begin to troubleshoot a process of administering skills-based assessment in anticipation of implementation on a larger scale. To that end, random sampling of participants was of lesser concern for these initial pilots than it would have been for future implementations. Thus, in part to promote administrator commitment and faculty involvement in Phase II, no probability sampling methods were employed; instead, convenience

sampling was used, informed primarily by faculty interest/availability, as well as course scheduling.[11]

Testing and Survey Pilots: Before Classes Begin

At Mohawk College, approximately 70% of new students are required to complete post-admission testing, which ultimately determines their placement in — or, for approximately 3% of test takers, exemption from — one of Mohawk's introductory Communications courses, as well as their selective enrolment in developmental math modules to upgrade skills before the start of their first term.

Table 9: Testing requirements for Phase II pilot programs

Program	Reading	Writing	Business Math	Technical Math
Business (two-year program)	YES	YES	YES	
Engineering Technology (two-year program)	YES	YES		YES
Engineering Technology (three-year program)	YES	YES		YES
Community, Justice & Liberal Studies (two-year program)	YES	YES		
Health (two-year program)	YES	YES		
Health (three-year program)	YES	YES		

Our sample is representative of the variations of post-admission testing administered at Mohawk. The majority of programs require students to write only the reading and writing tests. Students entering programs requiring post-admission mathematics testing complete one of two versions of the math assessment: either a five-section Business Math assessment or a ten-section Comprehensive Technical Math assessment.

Rubric Pilots: First and Final Semesters

Participant eligibility and recruitment for all Phase II research were determined by enrolment in the individual courses selected to participate in curriculum-embedded assessment of existing assignments. Students in the selected first-semester courses (and one second-semester course) served as the incoming students for this research and students in fourth- and sixth-semester (i.e., final-semester) courses (and one fifth-semester course) served as the project's outgoing students. Since Phase II was only one year, this study was not longitudinal, though the data from this phase could be used to further develop a longitudinal analysis in the future by following the "incoming" students through to the completion of their respective programs. For the purposes of Phase II research, only Mohawk's two-year diploma and three-year advanced diploma programs were eligible to participate.

Testing and Survey Pilots: Before Graduation

Students in their final year who were invited to participate in the rubric pilots were our preferred group for completing the re-administration of AFS since the assignments they submitted for critical-thinking assessment could also be assessed with a separate rubric for writing skills, providing additional validation of the use of the ACCUPLACER communication tests within the

[11] This methodological approach is consistent with that used in the recent Essential Adult Skills Initiative (EASI) project, also funded and led by HEQCO (Weingarten et al., 2018; Weingarten & Hicks, 2018).

context of our research. All students in the selected courses were eligible to participate in end-of-program testing (subject to their interest/consent), though only those who completed AFS post-admission could provide data relevant to this research; this was clarified on the consent form (see **Appendix E**). We did account for the possibility that this particular pool of students would not be large enough to ensure we reached our participation commitments. Based on low testing members in the fall term and similarly low participation numbers at the start of testing in the winter term, we opted to invite three additional business programs and one additional computer science program to test and complete only the Student Exit Survey. We were also able to offer online testing to all invited students in the winter term.

Students were offered financial incentive for their participation in the form of Amazon gift codes. They were compensated $15 for each test they completed, up to $45 (see Table 12). They were able to write as many or as few tests (for which they were eligible) as they wanted. Students were also able to complete the Student Exit Survey apart from the testing via Mohawk's LMS, which is consistent with its administration for incoming students.

Table 10: Incentives for Phase II Testing Pilots

Program Area	Reading	Writing	Business Math	Technical Math	Possible Total Incentive
Business	$15	$15	$15		$45
Engineering	$15	$15		$15	$45
Community, Justice & Liberal Studies	$15	$15			$30
Health	$15	$15			$30

Unlike incoming students who are required to complete AFS before starting their programs, outgoing students have no similar directive or requirement to participate in end-of-program testing, which is why extrinsic (i.e., financial) incentives were employed. As soon-to-be graduates and prospective employees, these incentives were intended to compensate for their time and intellectual labour at a rate consistent with minimum wage (in this case, approximately $15 per hour). During classroom visits and prior to testing, research team members also spoke to additional, intrinsic motivators for student participation, including the value of being able to articulate skills and skill levels when applying to jobs. Pilot students had the option of receiving their results immediately following the completion of their tests, as is standard with the AFS process.

Finally, we asked that all students complete a brief (five-question) post-assessment questionnaire on their motivation, effort, confidence, commitment and comfort level after completing each test (**Appendix F**). Taken together with the Student Exit Survey, these surveys were intended to enhance the analysis of testing data by providing insight into the transferable competencies underlying the tested skills.

Assessment of Critical-thinking Skills: Rubric Pilots

Instrumentation[12]

Critical Thinking VALUE rubric

The VALUE rubrics are "meta-rubrics" (Siefert, 2012; Cumming & Miller, 2018), which "can be tailored to reflect discipline-specific requirements" (Deller et al., 2015). Designed as part of the Liberal Education and America's Promise (LEAP) initiative led by the AAC&U, they are the most commonly used programmatic/institutional-level rubrics (Cumming & Miller, 2018) and are recognized as having high face, content and construct validity due to the large, interdisciplinary team of experts (faculty, employers and accreditors) who developed them (Finley, 2011; Goff et al., 2015; Cumming & Miller, 2018; Drezek McConnell et al., 2019). In one study, the critical-thinking rubric was found to have the "highest degree of agreement and reliability" of all 16 (Finley, 2011), though the reliability of any of the VALUE rubrics is dependent on the quality of the process of norming or calibrating scorers (Siefert, 2012; Szafran, 2017; Scott et al., 2018). Additional studies into the reliability of the rubrics found moderate to strong interrater reliability (r=0.50 - 0.84), with scorers concluding that "the rubrics could be used for judging quality of learning in different courses in different fields by faculty from different departments" as "valid measures of the learning being assessed" (McConnell & Rhodes, 2017). Studies into the reliability and validity of these rubrics are ongoing.

We had originally planned to administer the Academic Motivation Scale from the Wabash National Study to provide insight into student attitudes toward learning relative to performance on course assessments. However, due to concerns about the potential impact on participation numbers with the introduction of an additional assessment and inconclusive evidence of the tool's reliability and validity, we opted not to include it.

Planning and Faculty Development

Participating faculty, supervising associate deans and the project faculty lead met to workshop the terms and specifics of each pilot. Each pilot team was provided with two resources: "LOAC II Pilots: Information and Guidelines" (**Appendix B**), the Critical Thinking VALUE rubric and a visual overview of the project (**Appendix H**); after teams had the opportunity to review this information, they were sent the "LOAC II Pilots: Planning Document" (**Appendix C**), with detailed requests for information relevant to the eventual staging of the pilot. Each team was also asked to customize their pilot's administration by deciding whether or not to

- Adapt or translate the Critical Thinking VALUE rubric for use within their particular discipline.
- Adapt the selected assignments to better align with the assessed skill.
- Communicate the details of the project and/or skills-based, curriculum-embedded assessment to their students.

It was crucial that these processes not be fully prescribed and that the opinions, insights and experiences of individual faculty be taken into account. However, to ensure these variables were accounted for, any and all modifications to the administration of these assessments within individual pilots were carefully documented. We anticipated that this variability would allow us to learn more about faculty preferences, ease of administration and the student experience,

[12] A full table of instruments and their assessment constructs appears in **Appendix G**.

ultimately informing the analysis of our results. The majority of participating faculty/programs opted not to modify their assignments. Some students had no awareness that research was taking place until they were approached for their consent, which, in some cases, occurred following assignment submission. All variations were tracked in detail and will be used to categorize our results.

Assessment Administration

The research team visited each class to briefly introduce the research and its potential value to students, answer any questions, request students' consent to participate and collect the completed consent forms (**Appendix E;** classroom visit protocol details outlined in **Appendix I)**. Faculty were not directly involved in the consent process so as not to compromise their relationship with their students. Students were provided with additional opportunities to consent outside of class time with the use of online consent forms shared via email and the LMS.

Once students provided consent to use their assignments for the rubric pilots, they were under no further obligation to the project. Unlike the testing pilots, students were not directly compensated for the use of their assignments since they were originally submitted for credit as part of their program requirements and toward the completion of their diplomas. However, to promote participation, an incentive draw was introduced in both the fall and winter terms and students were able to request the results of the scoring of their assignments. For the majority of student assignments, submission was facilitated through eLearn, Mohawk's LMS, which ensured easy access throughout the work sample collection, rubric scoring processes and beyond.

Scoring and Scorer Calibration

Two faculty members were selected from each participating program to score the collected assignments/work samples using the Critical Thinking VALUE rubric. Teams met to discuss the assignment content, the Critical Thinking VALUE rubric (and its relation to any course or assignment rubrics, as applicable) and how the constructs of the rubric were represented in the assignment's instructions, questions and submitted work. Part of these discussions was to determine whether any of the assessment criteria were not represented in the assignments (in some cases this might not become apparent until assessment is underway); the team then decided whether it was appropriate to assess for that criterion or to leave it out. Once everyone was comfortable with both the assignment and the rubric, the project faculty lead and scoring faculty worked through a few "practice" assignments to get a sense of each other's scoring and begin the calibration process. Previous research has shown that scorer preparation and calibration is crucial to ensuring reliable results (Siefert, 2012). This was intended to be an iterative process that would proceed slowly at first, allowing faculty multiple opportunities to check in with each other and compare scores. A guidance document was created to support faculty's completion of this work (**Appendix D)**.

All identifying information was removed from assignments prior to scoring to reduce any potential familiarity or bias. Individual assignments were coded to ensure work samples could be re-associated once scoring was complete, calibrated and debriefed. The Institutional Research and Data Analytics Department will oversee the re-association of coded assignments and student indicators, as well as the linking of rubric-scoring data with the testing data (pre- and post-) and student academic and demographic information.

Assessment of Basic Cognitive Skills (Reading, Writing, Mathematics): Testing Pilots

Instrumentation

ACCUPLACER Classic[13] Reading Comprehension and WritePlacer tests (including ESL versions)
Mohawk Math Assessment Test
Reading VALUE Rubric
Written Communication VALUE Rubric
Post-assessment Questionnaire (**Appendix F**)

ACCUPLACER advertises the high reliability of their tests with reliability scores ranging between .84 and .96, with an average reliability of .907. ACCUPLACER's promotional material also emphasizes the importance of test validity; however, rather than claim universal validity,[14] ACCUPLACER recommends that individual institutions conduct their own validation studies for each specific use of the tests, supported by their Admitted Class Evaluation Service (ACES). Since their suite of placement tests is designed to assess students' readiness to enter college, ACCUPLACER relies on individual institutions to determine whether the content assessed by their tests matches the requirements of any introductory or remedial classes (ACCUPLACER, 2017).

The Mohawk Math Assessment Test (MAT) assesses key mathematics skills generally covered in the provincial math curriculum between Grades 4 and 10. The MAT has undergone several internal reliability and validity reviews over the course of its history.

End-of-Program Testing

Programs participating in the rubric pilots were encouraged to invite the same group of students to participate in the re-administration of AFS, as well as a portion of Mohawk's Student Entrance Survey. Testing blocks were established in coordination with our testing centre as close to the end of students' final term as deadlines and scheduling would permit. Great care was taken to ensure that the implementation of this testing did not affect the important day-to-day functions of the centre, including all make-up and accommodated testing for Mohawk students. Our testing centre staff have previously been involved in research administration and informed consent processes as part of other HEQCO-funded research projects, and were provided with an operations outline to further support their involvement. Students were informed in advance that they should be prepared to set aside 1.25 – 3.5 hours to test, depending on which tests they agreed to take. Table 14 provides a complete list of time restrictions for Mohawk's AFS assessments.

In the winter term, students were provided the option of taking tests remotely. Unfortunately, due to limitations with the selected online proctoring service (Examity), remote administration of the ACCUPLACER reading and writing tests needed to be implemented separately from Mohawk's Assessment Test. After consenting and indicating their testing preferences on the

[13] ACCUPLACER announced the replacement of their Classic assessments with ACCUPLACER Next Generation assessments in late 2018. The sunset of all Classic assessments took place on January 28, 2019. However, thanks to the cooperation of ACCUPLACER, Mohawk was permitted to retain the use of Classic tests for the duration of Phase II. Since all incoming Mohawk students up to and including September 2018 were assessed using these versions of the tests, it was important that any end-of-program testing employ the same instrumentation to ensure consistency.

[14] When discussing the broader validity of their tool, ACCUPLACER points to a meta-analysis of 47 studies, which found a moderate to strong relationship between ACCUPLACER scores and course success (Mattern & Packman, 2009). However, one of the primary limitations of this analysis is its reliance on "validity reports published by the test-makers themselves" (Scott-Clayton, 2018). As such, its conclusions have been frequently challenged (Belfield & Crosta, 2012; Fulton, 2012; Medhanie et al., 2012; Willett, 2013; Ngo & Melguizo, 2015).

electronic consent form, students were required to register for the service, schedule their own tests at a time convenient to them (24/7) and "attend" their scheduled test when appropriate. For the ACCUPLACER tests, this process was facilitated with a voucher code that was automatically emailed to students once they were added to the appropriate branching profiles (i.e., the appropriate combination of reading and/or writing, standard or ESL). For the math test, a unique eLearn page was created to provide students with resources to support their registration and scheduling and to allow them to access both Examity (the proctoring service) and MapleTA (the software that hosts the assessment) in a manner that authenticated their credentials.

Table 11: Test Completion Times for the Re-Administration of Assessments for Success

Assessment	**Time Allotted**
ACCUPLACER Reading Comprehension	Untimed (typically 50–60 minutes)
ACCUPLACER WritePlacer	1 hour
Mohawk Math Assessment Test	1 hour (same for business and technical math)
Student Entrance Survey (excerpt)	Untimed (5–15 minutes)
Post-assessment Questionnaire(s)	Untimed (1–5 minutes)

To help ensure the integrity of our data, we established thresholds for both student receipt of incentives and inclusion in the project's dataset. Any tests that did not meet the minimum requirements outlined in Table 12 would not be included in our analysis. Data from the completed Post-Assessment Questionnaires (**Appendix F**) would be valuable to our final analysis, providing additional context for the skills-based test scores. From the perspective of program-level assessment, it was important for faculty to have some indication of students' attitudes/approaches to better inform the collected data and attempt to control for unmotivated students in the analysis of the sample (Swerdzewski et al., 2011; Rios et al., 2014; Horn & Tandberg, 2018).

Table 12: Completion Thresholds for Financial Incentives and Inclusion in Research/Data Analysis

Assessment	**Incentive Threshold**	**Threshold(s) for Inclusion in Data Analysis**
ACCUPLACER Reading Comprehension	Minimum 15 minutes	15 minutes and minimum 50% completion; >1 in response to question B (i.e., effort) on post-assessment questionnaire
ACCUPLACER WritePlacer	Minimum 300 words or 30 minutes	Minimum 300 words or 30 minutes; >1 in response to question B (i.e., effort) on post-assessment questionnaire
Mohawk Math Assessment Test	Minimum 15 minutes	15 minutes and minimum 50% completion; >1 in

		response to question B (i.e., effort) on post-assessment questionnaire
Student Entrance Survey (excerpt)	Completed	
Post-assessment Questionnaire	Completed (one for each test taken)	

Rubric Validation/Triangulation

For those students who agreed to participate in both the rubric and testing pilots, we will attempt to conduct a validation/triangulation of our use of the ACCUPLACER reading and writing tests using rubric scores (determined by expert assessment using the VALUE Written Communication rubric), pilot testing scores (ACCUPLACER Reading Comprehension and WritePlacer) and students' course grades and/or their weighted GPA. The work samples originally used to assess critical thinking will be used for a second round of scoring of writing skills. Students have been informed of this additional assessment process on the rubric pilot consent form. A methodological process consistent with that described for the critical-thinking rubric pilots will be implemented here with the rubric scoring of writing. The program coordinators for Mohawk's introductory Communication courses will serve as our scorers.

Due to curricular limitations, we are unable to provide similar validation/triangulation using assignments demonstrative of the mathematics skills assessed by the MAT.

Assessment of Transferable Skills: Survey Pilots

Instrumentation and Administration

Mohawk's Student Entrance Survey

Mohawk's Student Entrance Survey includes items from the Ontario College Student Engagement Survey and the Wabash National Study's Ryff Scales of Psychological Well-Being. The process of validating these items is ongoing. The research team is undertaking a retrospective analysis of multiple years of Student Entrance Survey data to identify key student entry items from the survey, determine whether uni- or multidimensional constructs exist and confirm/explore the reliability of the scale in order to determine what student entry constructs are sufficiently represented in the survey in relation to various demographics, as well as the incoming skills assessed during AFS. Anticipated constructs identified through this work include program fit, educational commitment, career clarity and perceived family support — some of which could be considered transferable skills for analysis in the context of this study.

Participating students were asked to (re)complete a portion of the Student Entrance Survey (now the Student Exit Survey) either at the end of testing in the testing centre or independently through the college's LMS.

Preliminary Results

Due to a disruption to our funding, we were unable to complete our data analysis. However, we are able to share participation data to provide insight into student uptake for both end-of-program testing and consent to the use of existing assignments for research scoring.

In total, we invited approximately 900 outgoing students to rewrite the reading, writing and mathematics tests, as well as recomplete portions of the Student Entrance Survey as a Student Exit Survey. Additionally, 2,600 students were asked to consent to the use of their assignments. Table 9 outlines our final participation numbers for both the fall and winter terms. A more granular breakdown of participation in the rubric pilots is outlined in Table 10.

Table 13: LOAC II Pilot Participation: Commitments versus Actual

	Assignments	**Tests**	**Surveys**
Invited participants	~2600	~900	~900
Original commitment	100–200 incoming 100–200 outgoing	100 outgoing	100 outgoing
Confirmed participants	**636 total consents** 446 incoming* 190 outgoing** 320 scored assignments (anticipated)	**32 testers** (85 completed tests) 90 registered (34 in-person 56 online)	**116 completed** (92 consented)

* Includes one cohort of second-term students
** Includes outgoing students in four- and six-term programs, as well as fourth- and fifth-term students in six-term programs

On average, students consented to the use of their assignments 25% of the time. We experienced the highest consent rate with an outgoing cohort of health students (83% in the winter term) and lowest consent rate with an outgoing cohort of engineering students at (10% in the fall term). Generally, more students (both incoming and outgoing) consented in the winter term than in the fall term. We believe this could be the result of adjustments made to our in-class presentations in response to feedback and observations from our first implementation. Overall, these consent rates positioned us to exceed our participant commitments for the rubric pilots, with representative samples available for scoring across all of the academic areas.

Table 14: Consents and Work Sample Collections by Academic Area

	"Incoming" students **(1st and 2nd term)**	**"Outgoing" students** **(4th, 5th and 6th term)**
Business	**Fall term** 530 students invited 88 consents	**Winter term** 230 students invited

	Winter term 340 students invited 94 consents **182 consents** **(60 would be scored)**	**84 consents** **(all fully consenting groups would be scored)**
Engineering (Combined)	**Fall term** 395 students invited 95 consents **Winter term** 320 students invited 62 consents **157 consents** **(60 would be scored)**	**Fall term** 60 students invited 6 consents **Winter term** 75 students invited 36 consents **42 consents** **(all would be scored)**
Community, Justice & Liberal Studies	**Fall term** 200 students invited 44 consents **Winter term** 135 students invited 28 consents **72 consents** **(60 would be scored)**	**Winter term** 110 students invited **22 consents** **(all would be scored)**
Health (Combined)	**Fall term** 50 students invited 17 consents **Winter term** 60 students invited 18 consents **35 consents** **(all would be scored)**	**Winter term** 65 students invited **42 consents** **(all would be scored)**

By comparison, students registered to complete some portion of the testing only 10% of the time on average. This was due in part to the fact that we invited such a large number of students to ensure our registration numbers were within range to meet our participant commitments. However, the actual testing rate was significantly lower; with only 32 students following through on their commitment to test, resulting in an overall participation rate for the testing pilot of 3.5%. Since the majority of these students did not actually write any of the tests in the end, we fell well short of our 100-student goal.

We did roughly meet our commitment for the survey pilot, with 10% of invited participants (92 students) consenting to the use of their completed surveys. While students had the option to

complete the Student Exit Survey as a component of their testing suite in Mohawk's testing centre, the survey was also accessible through existing student accounts and the college's LMS, and required significantly less time to complete than the tests. Ease of access and reduced time commitment likely contributed to more participants from the same pool completing the survey than the tests.

Given the abrupt conclusion of our research and data analysis, we are not able to answer the majority of our research questions related to skills learning and the relationship between skills acquisition and transferable skills/competencies. However, we can assert tentative conclusions in response to our first research question: "Can existing post-admission assessment practices be leveraged to support the transition into, and efficient administration of, institutional generic skills assessment?" On account of the extremely low uptake by outgoing students to retake the reading, writing and math tests they first completed ahead of their first year of study, we conclude that the re-administration of post-admission tests is likely not a viable approach to institutional skills assessment. Without a viable comparator for exiting students, incoming test scores have limited value beyond their intended use to assess and often to stream students into first-year courses. With no "requirement" or impetus to retake the tests at the end of their program of study, students appear unwilling to and uninterested in participating, even with significant financial incentive at stake. Additionally, with post-admission testing decreasing across the sector (down 17% from 2010, according to the results of our environmental scan), fewer colleges can rely on having that incoming data on hand.

Students were significantly more willing to consent to the use of their assignments for scoring purposes than they were to commit additional time to taking tests. We found faculty and administrators to be enthusiastic, too, about the curriculum-embedded approach to skills assessment, which was viewed as having more direct applicability to and impact on their curriculum and their students' learning than the standardized tests. Selecting, sampling, scoring, calibrating and analyzing existing assignments to measure skill development can certainly be time and resource intensive (we speak to this and other challenges in our conclusion below and share insights and recommendations for future research and practice); however, for the above reasons, we believe the VALUE approach to institutional assessment is worthy of further exploration. Since the end of this study, we have explored and piloted modified approaches to this type of assessment, improving upon its efficiency and actionability, most notably by integrating skills assessment with existing quality assurance practices.

Conclusion: Recommendations for Future Research and Practice

While our work with HEQCO ended prior to the completion of this study, Mohawk is committed to continuing to pursue learning outcomes assessment with an aim to better align and scaffold curriculum within and across programs and to produce useable data on our students' skill development. This data is crucial not only at a high level, but also at the individual level, helping students to better articulate the skills and competencies they develop and refine throughout their postsecondary experience.

We anticipate that we will not be the only institution pursuing this work; with that in mind, we have outlined key challenges we encountered throughout our implementation, how we resolved them in the context of this research project and how we would improve upon our protocol and

processes in future implementations to ensure more effective and more efficient data collection and analysis and, ultimately, ensure greater benefit to student learning.

Recruitment Challenges

Challenge: Securing Student Consent

We quickly learned that our expectations related to students' willingness to consent to the use of course assignments were overly optimistic. While we were able to secure participation/consent in the range of 20–30% (even greater than 80% in one instance), these numbers were only achieved through considerable effort.

Solution: Multiple means of providing consent, incentives and widespread recruitment

We had originally intended to facilitate all informed consent electronically to allow for more efficient tracking; however, upon launching in the fall, we swiftly amended our REB application to allow for hard-copy consent forms that could be distributed during classroom visits, encouraging students to review the details and make an immediate decision on their participation while we were available to answer questions. Meanwhile, our electronic consent forms allowed us to reach students who were not present in class the day we visited, or who preferred to take more time to consider their participation and follow up afterward.

We also instituted a formerly unplanned incentive draw for rubric pilot students in recognition of the time they were taking to learn about our project and consent to the use of their intellectual property. We believe these changes were crucial to promoting student involvement based on the comparative rate of participation before and after they were instituted. Even with these changes, there were still some concerns about the ability to reach participation commitments; thus, we opted to expand the scope of our recruitment in the winter term to include an additional, related program from engineering technology (which allowed us to invite more outgoing students to participate) and an additional program from health, as well as January intakes for the incoming courses already invited to participate in the fall.

Solution: Faculty Endorsement

Ease of consent and financial incentive were not the only factors impacting participation rates. Many students expressed reluctance to and discomfort with their personal academic work being used in the context of research. Based on anecdotal evidence from over 50 classroom visits, students were far more likely to consent to participate when the professor of the class we were visiting had previously signaled our arrival, spoke to the project and its potential benefits and was engaged in our presentation. As has consistently been pointed out with similar recruitment efforts, the support of program faculty cannot be underestimated. They set the tone for both the delivery and the reception of the shared information, and in some exceptional cases, breed enthusiasm for the research and its implications.

Challenge: Efficient Administration

There is no doubt this type of work is labour intensive; we are acutely aware of the time and effort involved in recruiting programs and faculty, securing student consent, collecting, sorting, anonymizing, coding and scoring student work samples, and analyzing these scores. It has become increasingly clear that, in order for a scaled initiative of this sort to succeed, it would need to be integrated into and supported by existing institutional frameworks. After five years of implementing their own learning outcomes assessment projects, Queen's University drew similar conclusions; in a recent chapter highlighting lessons learned they likewise stress the

importance of program-level learning outcomes assessment data as a means of engaging instructors, the contextual value of authentic assessment methods and the need for sustainability of those methods, while cautioning that there is no perfect model for work that is so complex (Scott et al., 2018).

Solution: Leveraging Existing Technologies

It is our contention that learning outcomes assessment initiatives would be best supported by targeted developments of an institution's LMS in ways that help reduce the workload of both researchers/administrators and faculty/students. Effective upgrades could include:

- Online VALUE rubrics integrated into the LMS
- Implementation of a default disclaimer regarding use of assignments for quality assurance at the point of online assignment submission (as a substitute for onerous consent processes)
- Automated sampling and cataloguing of student work

Conestoga College has conducted related work to support the collection of work samples for PEQAB's external expert review process for degrees. In collaboration with their librarians, they designed a collection and cataloguing process and, ultimately, a repository to address the need for consistent, efficient, accessible and high-quality work sample collection (Weigel-Green et al., 2018).

We are also interested in standardizing and streamlining a similar work sample collection process for the purposes of skills assessment; specifically, we are exploring how it could be embedded in and supported by our LMS. Online assignment submission is utilized fairly consistently across the college; almost all of the assignments collected as part of Phase II were submitted and are managed through the system. This pre-existing database of samples can be leveraged to sample assignments, generate repositories and assign scorers. Obviously, the degree to which relevant activities can be integrated into an LMS is limited not only by the LMS software, but also by the pace of technological innovation. We recognize that the success of such initiatives would be reliant on extensive preparatory work, training and professional development for our faculty. However, we believe the efforts to better sample and catalogue in alignment with Institutional Learning Outcomes benefits not only the skills assessment process, but also the delivery of blended and online learning more broadly, and we consider this to be a key growth area in skills assessment moving forward.

Challenge: Increasing Institutional Capacity and Buy-in

Finally, while recruitment of research participants (in this case students) is always a primary challenge of any research project involving human subjects, implementing learning outcomes assessment and reporting also requires the recruitment and commitment of large numbers of faculty, staff and administrators.

Solution: Emphasizing Flexibility and Program-level Benefits

One of the key lessons we learned in Phase II is that administrator and faculty buy-in and participation are dependent on them seeing the potential benefits of skills assessment at a program level. While our research questions, the consortium and HEQCO's work more generally are tied to institutional reporting on student skill development, for our internal stakeholders and research partners, data is most usefully thought of and applied locally. All of the administrators and faculty with whom we worked on this project view these assessment

pilots as opportunities to learn more about what is working and not working in their programs. In response, we chose to locate our Phase II pilots within specific programs of study, and allow faculty on each project freedom to modify elements of the pilot protocol in response to their particular program's needs.

Online Assessment Challenges

This project allowed us to investigate alternative approaches not only to post-admission assessment, but also to the delivery of those assessments and the means by which they are proctored. Beyond the scope of this research project, Mohawk has been committed to exploring testing alternatives and accessible solutions to best support the diverse needs of our students. Our Environmental Scan (and subsequent followups, including as part of internal literacy strategy discussions) revealed that we are not alone in these efforts, with colleges across the province actively working to determine how best to assess their students on their way in, whether to assess their students on their way out and how, at the level of curriculum and targeted courses, to scaffold essential skill development within largely disciplinary paradigms.

As a newly introduced component of our winter pilots, we were able to trial two different approaches to remote proctoring, which, prior to this research, had not been attempted at the college. We were very curious and eager to pilot these services, which would allow students to complete both the ACCUPLACER reading and writing and internally developed mathematics tests online in their own time and space. We believed this increased accessibility would benefit not only the administration of research efforts like LOAC II, but could also have a significant impact on the ease of administration of a process like Assessments for Success, particularly for out-of-town and international students.

Challenge: Scheduling and Accessing Remote-proctored Online Assessments

Our experience with remote testing administration was considerably fraught. We were unable to allow students to complete our Math Assessment Test at the same time as their ACCUPLACER tests, since ACCUPLACER tests are proctored through an exclusive Examity contract that cannot be connected to other Examity testing. This meant developing/supporting two entirely different processes for communication, access, registration, scheduling and test-taking. Examity's existing processes for remote testing are already considerably complex, requiring multiple steps on the part of students well in advance of testing and at the point of testing (i.e., authentication, systems requirements, etc.). Particularly as part of an optional research project, we were concerned the multi-step process would result in loss of student participants at every stage, and it appears we were right. While almost 100 students registered to complete the tests, with the majority requesting remote administration, only a third actually wrote them, with only a small handful of students completing them online. A number of additional students made it through the registration/scheduling process, only to encounter a litany of technical issues upon showing up to test. And when we attempted to extend the expiry of existing vouchers to let interested students test after their exam period, we discovered that the dates associated with the vouchers could not be modified, and would need to be individually reissued — a process that only confused students further.

Overall, we learned that our operations were not well-positioned to transition to a fully online delivery model — the assessments, the technology and our processes each presented barriers that would need to be addressed. This holds true for the external assessments as well — the

technology and processes are not as user friendly (for administrators or students) as they need to be to facilitate efficient assessments.

Solution: TBD

While the promise of these online survey and testing interventions is certainly compelling, it is clear more work is needed to make these processes more efficient for students and institutions alike. However, as institutions across the country and the world moved to remote and online learning in 2020 in response to the COVID-19 pandemic, there are likely to be new and emergent tools and practices to explore, as well as an enhanced understanding of what is possible, reasonable and appropriate when it comes to assessing student work online.

Alignment Challenges

Challenge: Finding the "Right" Assignments

As highlighted by Queen's University in their reporting on similar learning outcomes assessment initiatives (Simper et al., 2018; Scott et al., 2018), in order to maximize the use of the VALUE rubrics, it is crucial that assignments are purposefully designed — or at least re-designed or aligned — with those rubrics before pilots or courses begin. We cannot responsibly assess students for skills they were either not explicitly asked to demonstrate or asked to demonstrate in ways that diverge significantly from the performance criteria outlined in the rubrics themselves.

Particularly in first-year courses, it was not uncommon for our faculty to struggle to find (substantive) assignments that explicitly required the demonstration and therefore assessment of critical thinking. And when critical-thinking assignments were identified, the degree to which it would have been considered reasonable for all VALUE rubric dimensions to be represented was inconsistent. Because these assignments were not necessarily designed to assess critical thinking, nor explicitly designed to assess critical thinking using VALUE-rubric criteria specifically, they were not ideally positioned to serve as representative measures of students' critical-thinking levels.

Solution: Developing Whole Curriculums with Skills in Mind

This is not to say all assignments should be designed exclusively with something like the Critical Thinking VALUE rubric in mind; such extreme alignment efforts could end up reproducing the "teaching to the test" approach consistently critiqued within more traditional assessment methods. Like any good pedagogical tool, the process of implementing and integrating the VALUE rubrics is a matter of deciding when, where and how (much) to use them — decisions that need to be made within each unique assessment context. It is our contention that rigorous discussions on what constitutes critical thinking, how its component skills and competencies can be clarified for students, and how students can be mentored through gradually more intricate demonstrations of those component skills and competencies not only serve to better align assessment; they also help improve curriculum and teaching. The discussions and work accomplished across the course of this project no doubt built internal capacity for future skills instruction and assessment work in the future, whether at the macro or micro level (or both). Many of our participating faculty have already adapted existing assignments (and not only those assignments used as part of LOAC II), implementing the VALUE rubrics as developmental tools within their own teaching. Overall, this work highlighted the need to scaffold skill instruction/development across programs, instead of leaving the demonstration of higher-order skills like critical thinking to final-term, capstone-like assessments. It is this work that has been

the most exciting and rewarding for all involved in the project, which is why Mohawk is now taking steps to integrate it into our quality assurance framework.

References

ACCUPLACER. (2017). Reliability and Validity. *College Board.*

Ahlawat, S., Miller, G., & Shahid, A. (2012). Promoting Liberal Learning in a Capstone Accounting Course. *American Journal of Business Education, 5*(1), 11–24.

Altbach, P. G. (2015, May 15), AHELO: The Myth of Measurement and Comparability. *University World News*. https://www.universityworldnews.com/post.php?story=20150515064746124

Altbach, P. G. & Hazelkorn, E. (2018). Can we measure education quality in global rankings? *University World News*. http://www.universityworldnews.com/post.php?story=20180814184535721

Appleby, K. M., Foster, E., & Kamusoko, S. (2016). Full Speed Ahead: Using a Senior Capstone Course to Facilitate Students' Professional Transition. *Journal of Physical Education and Dance*, *87*(3), 16–21.

Arum, R. & Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. Chicago: The University of Chicago Press.

Baird, L. (1988). Diverse and Subtle Arts: Assessing the Generic Outcomes of Higher Education. In C. Adelman (Ed.), *Performance and Judgment: Essays on Principles and Practice in the Assessment of College Student Learning*. Washington: U.S. Government Printing Office.

Baird, J., Andrich, D., Hopfenbeck, T. N. & Stobart, G. (2017). Assessment and Learning: Fields Apart? *Assessment in Education: Principles, Policy & Practice*, *24*(3), 317–350.

Banta, T. W. (2008). Trying to Clothe the Emperor. *Assessment Update*, *20*(3–4), 16–17.

Banta, T. W. & Blaich, C. (2011). Closing the Assessment Loop. *Change: The Magazine of Higher Learning*, *43*(1), 22–27.

Banta, T. W., Jones, E. A. & Black, K. E. (2009). *Designing Effective Assessment: Principles and Profiles of Good Practice*. San Francisco: Jossey-Bass.

Banta, T. W., & Pike, G. R. (2012). Making the Case Against – One More Time. In P. Ewell & R. Benjamin (Eds.), *The Seven Red Herrings about Standardized Assessments in Higher Education* (pp. 24–30). NILOA Occasional Paper #15. Champaign, IL: National Institute for Learning Outcomes Assessment.

Barrie, S., Hughes, C., Crisp, G., & Bennison, A. (2012). *Assessing and Assuring Australian Graduate Learning Outcomes: Principles and Practices Within and Across Disciplines*. Final report. http://www.itl.usyd.edu.au/projects/aaglo/pdf/SP10-1879_FINAL%20sydney%20barrie%20final%20report%20part%201.pdf

Beld, J. (2015). *Building Your Assessment Toolkit: Strategies for Gathering Actionable Evidence of Student Learning*. Atlanta: Southern Education Foundation.

Belfield, C., & Crosta, P. (2012). Predicting Success in College: The Importance of Placement Tests and High School Transcripts. *Community College Research Center*.

Biggs, J. & Tang, K. (2011). *Teaching for Quality Learning at University* (4th ed.). Maidenhead: Oxford University Press.

Blaich, C. & Wise, K. (2011). *From Gathering to Using Assessment Results: Lessons from the Wabash National Study*. National Institute for Learning Outcomes Assessment.

Boden, R., & Nedeva M. (2010) Employing discourse: universities and graduate 'employability', *Journal of Education Policy*, 25:1, 37–54.

Bowman, N. A. (2010), Assessing Learning and Development among Diverse College Students. *New Directions for Institutional Research*, *2010*(145), 53–71.

Bowman, N. A. (2011). Examining Systematic Errors in Predictors of College Student Self-Reported Gains. *New Directions for Institutional Research, 2011*(150), 7–19.

Brumm, T., Mickelson, S., & White, P. (2006). Integrating Behavioral-Based Interviewing Into the

Curricula. *NACTA Journal*, 50(2), 28–31.

Brumwell, S., Deller, F. & Hudak, L. (2018). The Case for Large-Scale Skills Assessment. In F. Deller, J. Pichette & E. K. Watkins (Eds.), *Driving Academic Quality: Lessons from Ontario's Skills Assessment Projects* (pp. 53–64). Toronto: Higher Education Quality Council of Ontario.

Bunyan, S., Jonker, L., & Dion, N. (2015). *The Effects of Developmental Communication Instruction on Language Skills and Persistence at Four Ontario Colleges*. Toronto: Higher Education Quality Council of Ontario.

Business Council of Canada and Morneau Shepell. (2018). *Navigating change: 2018 Business Council Skills Survey*.

Calderon, O. (2013). Direct and Indirect Assessment of Learning Outcomes: What Do We Actually Measure? *Journal of Social Work Education*, *49*(3), 408–419.

Cassady, J. C. & Johnson, R. E. (2002). Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, *27*(2), 270–295.

Chapell, M. S., Blanding, Z. B., Takahashi, M., Silverstein, M. E., Newman, B., Gubi, A., & McCann, N. (2005). Test Anxiety and Academic Performance in Undergraduate and Graduate Students. *Journal of Educational Psychology*, *97*(2), 268–274.

Craig, R. & Markowitz, T. (2017, March 17). The Skills Gap is Actually an Awareness Gap — And It's Easier to Fix. *Forbes*. http://www.forbes.com/sites/ryancraig/2017/03/17/the-skills-gap-is-actually-an-awareness-gap-and-its-easier-to-fix/#3ffd7d203ff4

Cumming, T. & Miller, D. M. (2018). *Enhancing Assessment: Putting Psychometrics to Work in Higher Education*. Sterling: Stylus.

Deller, F. (2018). Conclusion: Making the Leap from Assessment to Teaching and Learning. In F. Deller, J. Pichette & E. K. Watkins (Ed.), *Driving Academic Quality: Lessons from Ontario's Skills Assessment Projects* (pp. 53–64). Toronto: Queen's Printer for Ontario.

Deller, F., Brumwell, S. & MacFarlane, A. (2015). *The Language of Learning Outcomes: Definitions and Assessments*. Toronto: Higher Education Council of Ontario.

Dion, N. (2014). *Emphasizing Numeracy as an Essential Skills*. Toronto: Higher Education Quality Council of Ontario.

Dion, N. & Maldonado, V. (2013). *Making the Grade? Troubling Trends in Postsecondary Student Literacy*. Toronto: Higher Education Quality Council of Ontario.

Drezek McConnell, K., Horan, E. M., Zimmerman, B., & Rhodes, T. L. (2019). *We Have a Rubric for That: The VALUE Approach to Assessment*.

Fairchild, G. F., & Taylor, T. G. (2000, June). *Using Business Simulations and Issue Debates to Facilitate Synthesis in Agribusiness Capstone Courses*. Paper presented at the meeting of the Western Agricultural Economics Association, Vancouver, BC.

Finley, A. P. (2011). How Reliable Are the VALUE Rubrics? *Peer Review*, *13/14*(4/1), 31–33.

Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series*, *2015*(2), 1–17.

Finney, S. J., Mathers, C. E., & Myers, A. J. (2016). Investigating the Dimensionality of Examinee Motivation across Instruction Conditions in Low-Stakes Testing Contexts. *Research & Practice in Assessment, 11*(2016), 5–17.

Fisher, R. & Hoth, W. (2010). *College-Level Literacy: An Inventory of Current Practices at Ontario's Colleges*. Toronto: Higher Education Quality Council of Ontario.

French, E., Bailey, J., van Acker, E., & Wood, L. (2015). From Mountaintop to Corporate Ladder — What New Professionals Really Really Want in a Capstone Experience! *Teaching in Higher Education*, *20*(8), 767–782.

Fulton, M. (2012). Using State Policies to Ensure Effective Assessment and Placement in Remedial Education. *Education Commission of the States (ECS)*.

Gerwing, T. G., Rash, J. A., Allen Gerwing, A. M., Bramble, B., & Landine, J. (2015). Perceptions and Incidence of Test Anxiety. *The Canadian Journal for the Scholarship of Teaching and Learning, 6*(3), 3. http://dx.doi.org/10.5206/cjsotl-rcacea.2015.3.3

Goff, L., Potter, M. K., Pierre, E., Carey, T., Gullage, A., Kustra, E., Lee, R., Lopes, V., Marshall, L., Martin, L., Raffoul, J., Siddiqui, A. & Van Gastel, G. (2015). *Learning Outcomes Assessment: A Practitioner's Handbook*. Toronto: Higher Education Quality Council of Ontario.

Gonyea, R. M. & Miller, A. (2011), Clearing the AIR about the Use of Self-Reported Gains in Institutional Research. *New Directions for Institutional Research*, *2011*(150), 99–111.

Goodwin, J., Goh, J., Verkoeyen, S., & Lithgow, K. (2019). Can students be taught to articulate employability skills?, *Education + Training*, 61(2).

Gulikers, J. T. M., Bastiaens, T. J. & Kirschner, P. A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Educational Technology Research and Development (ETR&D)*, *52*(3), 67. http://doi.org/10.1007/BF02504676

Haave, N. C. (2015). Survey of Biology Capstone Courses in American and Canadian Higher Education: Requirements, Content, and Skills. *Bioscene: Journal of College Biology Teaching*, *41*(2), 19–26.

Haladyna, T. M., & Downing, S. M. (2004). Construct-Irrelevant Variance in High-Stakes Testing. *Educational Measurement: Issues and Practice*, *23*(1), 17–27.

Harrison, A. (2017). *Skills, Competencies and Credentials*. Toronto: Higher Education Quality Council of Ontario.

Hauhart, R. C. & Grahe, J. E. (2015). *Designing and Teaching Undergraduate Capstone Courses*. San Francisco: Jossey-Bass.

Hora, M. (2017) Beyond the Skills Gap. *NACE Journal.*

Hora, M., Benbow, R. & Oleson, A. (2016). *Beyond the skills gap: preparing college students for life and work*. Cambridge, Massachusetts: Harvard Education Press.

Horn, A. S. & Tandberg, D. A. (2018). Indicators of Educational Quality for Postsecondary Accountability Systems. In H.P. Weingarten, M. Hicks, & A. Kaufman (Eds.) *Assessing Quality in Postsecondary Education: International Perspectives* (pp. 3–26). Kingston: McGill-Queen's University Press.

Hutchings, P., Borin, P., Keesing-Styles, L., Martin, L., Michael, R., Scharff, L., Simkins, S. & Ismail, A. (2013). The Scholarship of Teaching and Learning in an Age of Accountability: Building Bridges. *Teaching and Learning Inquiry*. *1*(2), 35–47.

Jackson, D. (2013), Student perceptions of the importance of employability skill provision in business undergraduate programs. *Journal of Education for Business*, 88(5), 271–279.

Jackson, D. and Chapman, E. (2012), Non-technical skill gaps in Australian business graduates, *Education + Training*, 54(2/3), 95–113.

James, C. L. (2006). ACCUPLACER (TM) onLine: Accurate Placement Tool for Developmental Programs*? Journal of Developmental Education*, *30*(2), 2–8.

Joy, R., Shea, R. & Youden-Walsh, K. (2013) Advancing career integrated learning at Memorial. Paper presented at Cannexus, Ottawa, January 28–30.

Kerka, S. (2001). Capstone Experiences in Career and Technical Education. Practice Application Brief No. 16, ERIC Clearinghouse.

Kiener, M., Ahuna, K. H., & Gray Tinnesz, C. (2014). Documenting Critical Thinking in a Capstone Course: Moving Students toward a Professional Disposition. *Education Action Research*, *22*(1), 109–121.

Kinzie, J. (2013). Taking Stock of Capstones and Integrative Learning. *Peer Review*, *15*(4), 27–30.

Kovalcik, B. (2019). Developing Employability Skill Articulation in College Students: A Framework and Practitioner Approaches for Co-curricular Educators. *The Journal of Campus Activities Practice and Scholarship*. 1(2).

Lackner, C., & Martini, T. (2017). Helping University Students Succeed at Employment Interviews: The Role of Self-Reflection in e-Portfolios. *Teaching & Learning Inquiry*, *5*(2), 3–15.

Lennon, M. C. & Jonker, L. (2014). *AHELO: The Ontario Experience*. Toronto: Higher Education Quality Council of Ontario.

Lennon, M. C., Frank, B., Humphreys, J., Lenton, R., Madsen, K., Omri, A. & Turner, R. (2014). *Tuning: Identifying and Measuring Sector-Based Learning Outcomes in Postsecondary Education*. Toronto: Higher Education Quality Council of Ontario.

Litchfield, B. C. & Dempsey, J. V. (2015). Authentic Assessment of Knowledge, Skills and Attitudes. *New Directions for Teaching and Learning*, *142*(2015), 65–80.

Liu, O. L. (2011). Value-Added Assessment in Higher Education: A Comparison of Two Methods. *Higher Education*, *61*(4), 445–461.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring Learning Outcomes in Higher Education: Motivation Matters. *Educational Researcher*, *41*(9), 352–362.

Lock, J., Kim, B., Koh, K. & Wilcox, G. (2018). Navigating the Tensions of Innovative Assessment and Pedagogy in Higher Education. *The Canadian Journal for the Scholarship of Teaching and Learning*, *9*(1). http://doi.org/10.5206/cjsotl-rcacea.2018.1.8

MacFarlane, A. & Brumwell, S. (2016). *The Landscape of Learning Outcomes Assessment in Canada*. Toronto: Higher Education Quality Council of Ontario.

Mathers, C. E., Finney, S. J. & Hathcoat, J. D. (2018). Student Learning in Higher Education: A Longitudinal Analysis and Faculty Discussion. *Assessment & Evaluation in Higher Education*, *43*(8), 1211–1227.

Mattern, K. D. & Packman, S. (2009). Predictive Validity of ACCUPLACER Scores for Course Placement: A Meta-Analysis. *College Board*. 2.

Mayhew, M., Rockenbach, A. N., Bowman, N. A., Seifert, T. A. D., Wolniak, G. C., Pascarella, E. T., & Terenzini, P. T. (2016). *How College Affects Students* (Vol. 3). San Francisco: Jossey-Bass.

McConnell, K. D. & Rhodes, T. L. (2017). *On Solid Ground: VALUE Report 2017*. Association of American Colleges & Universities.

McCormick, A. C. & Kinzie, J. (2018). Quality as both State and Process: Implications for Performance-Indicator Systems. In H. P. Weingarten, M. Hicks & A. Kaufman (Eds.) *Assessing Quality in Postsecondary Education: International Perspectives* (pp. 27–48). Kingston: McGill-Queen's University Press.

Medhanie, A. G., Dupuis, D. N., LeBeau, B., Harwell, M. R. & Post, T. R. (2012). The Role of the ACCUPLACER Mathematics Placement Test on a Student's First College Mathematics Course. *Educational and Psychological Measurement*, 72(2), 332–351.

Messick, S. (1988). *Meaning and Values in Test Validation: The Science and Ethics of Assessment*. Princeton: Educational Testing Service.

Miner, R. (2010). *People Without Jobs, Jobs Without People: Canada's Labour Market Future.* Toronto: Miner and Miner Management Consultants.

Miner, R. (2014). *The Great Canadian Skills Mismatch: People Without Jobs, Jobs Without People and MORE* (Report No.II). Toronto: Miner and Miner Management Consultants.

Musekamp, F. & Pearce, J. (2016). Student Motivation in Low-Stakes Assessment Contexts: An Exploratory Analysis in Engineering Mechanics. *Assessment & Evaluation in Higher Education*, *41*(5), 750–769.

Ngo, F. & Melguizo, T. (2015). How Can Placement Policy Improve Math Remediation Outcomes? Evidence from Experimentation in Community Colleges. *Educational Evaluation and Policy Analysis*.

OECD. (2018a). Programme for International Student Assessment (PISA). http://www.oecd.org/pisa/

OECD. (2018b). Survey of Adult Skills (PIAAC). http://www.oecd.org/skills/piaac/

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How Robust Are the Findings of Academically Adrift? *Change: The Magazine of Higher Learning, 43*(3), 20–24.

Patry, D., & Ford, R. (2016). *Measuring Resilience as an Education Outcome*. Toronto: Higher Education Quality Council of Ontario.

Porter, S. R. (2013). Self-Reported Learning Gains: A Theory and Test of College Student Survey Response. Re*search in Higher Education, 54*(2), 201–226.

Pusecker, K. L., Torres, M. R., Crawford, I., Levia, D., Lehman, D., & Copic, G. (2012). Increasing the Validity of Outcomes Assessment. *Peer Review*, *13/14*(4/1), 27–30.

Qin, L. D. (2017). *Assessing the Relationship between Multiple Measures Placement and Student Academic Success at a Community College (*Ed.D. Dissertation). Retrieved from ProQuest. (10255925)

RBC. (2018). *Humans Wanted: How Canadian youth can thrive in the age of disruption.*

Rhodes, T. L. (2012). Show Me the Learning: Value, Accreditation, and the Quality of the Degree. *Planning for Higher Education*, *40*(3), 36–42.

Rios, J., Liu, O. L. & Bridgeman, B. (2014). Identifying Low-Effort Examinees on Student Learning Outcomes Assessment: A Comparison of Two Approaches. *New Directions for Institutional Research*.

Scott, J., Frank, B. & Simper, N. (2018). Ten Recommendations for Undertaking Institutional Assessment. In F. Deller, J. Pichette & E. K. Watkins (Eds.), *Driving Academic Quality: Lessons from Ontario's Skills Assessment Projects* (pp. 29–52). Toronto: Queen's Printer for Ontario.

Scott-Clayton, J. (2012). Do High-Stakes Placement Exams Predict College Success? *Community College Research Center*.

Scott-Clayton, J. (2018). Evidence-Based Reforms in College Remediation are Gaining Steam — and So Far Living Up to the Hype. *Brookings*. http://www.brookings.edu/research/evidence-based-reforms-in-college-remediation-are-gaining-steam-and-so-far-living-up-to-the-hype/

Shulman, L. S. (2003). No Drive-By Teachers. *Carnegie Perspectives*. http://carnegiefoundation.org/perspectives/no-drive-teachers

Siefert, L. (2012). Assessing General Education Learning Outcomes. *Peer Review*, *13/14*(4/1), 9–11.

Simper, N., Frank, B., Scott, J., & Kaupp, J. (2018). *Learning Outcomes Assessment and Program Improvement at Queen's University*. Toronto: Higher Education Quality Council of Ontario.

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure? *Academy of Management Learning & Education, 9*(2), 169–191.

Social Capital Partners and Deloitte. (2015). *Working together: Implementing a demand-led employment and training system*. Author

Strachan, L. (2016), Teaching employability skills through simulation games. *Journal of Pedagogic Development*, 6(2), 8–17.

Swerdzewski, P. J., Harmes, J. C., & Finney, S. J. (2011). Two Approaches for Identifying Low-Motivated Students in a Low-Stakes Assessment Context. *Applied Measurement in Education*, *24*(2), 162–188.

Szafran, R. F. (2017). *The miscalculation of interrater reliability: A Case Study Involving the AAC&U VALUE Rubrics*. *Practical Assessment, Research & Evaluation*, *22*(11), 2. https://pareonline.net/getvn.asp?v=22&n=11

Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test Anxiety Inventory: 30 Years Later. *Anxiety, Stress, & Coping: An International Journal*, *25*(6), 667–677.

Tam, M. (2014). Outcomes-Based Approach to Quality Assessment and Curriculum Improvement in Higher Education. *Quality Assurance in Education*, *22*(2), 158–168.

Tremblay, K., Lalancette, D. & Roseveare, D. (2012). *Assessment of Higher Education Learning Outcomes: Feasibility Study Report*. Vol. 1 Design and Implementation.

Vitasari, P., Wahab, M. N. A., Othman, A., Herawan, T., & Sinnadurai, S. K. (2010). The Relationship between Study Anxiety and Academic Performance among Engineering Students. *Procedia – Social and Behavioral Sciences, 8*(2010), 490–497.

Wagenaar, R. (2008). *Learning Outcomes a Fair Way to Measure Performance in Higher Education: the TUNING Approach*. Organisation for Economic Co-operation and Development. http://www.oecd.org/site/eduimhe08/41203784.pdf

Watkins, E. K. & McKeown, J. (2018) The Inside Story on Skills: The Student Perspective. In F. Deller, J. Pichette & E. K. Watkins (Eds.), *Driving Academic Quality: Lessons from*

Ontario's Skills Assessment Projects (pp. 81–92). Toronto: Higher Education Quality Council of Ontario.

Weigel-Green, T., Brown, J. & Burrows, S. (2018, October 11–12). *Aligning Evidence with Outcomes: Conestoga's Student Work Repository*. Paper presented at the Educating for the Future: Learning Outcomes and Experiential Learning Symposium. Toronto, ON.

Weingarten, H. P., Drummond, D. & Finnie, R. (2015). Canada Needs a New Skills Agenda [Blog post]. http://blog-en.heqco.ca/2015/10/harvey-weingarten-don-drummond-ross-finnie-canada-needs-a-new-skills-agenda/

Weingarten, H. P. & Hicks, M. (2018). *On Test: Skills: Summary of Findings from HEQCO's Skills Assessment Pilot Studies*. Toronto: Higher Education Quality Council of Ontario.

Wiggins, G. & McTighe, J. (2005). *Understanding by Design* (2nd ed.). Alexandria: Association for Supervision and Curriculum Development.

Willett, T. (2013). Student Transcript-Enhanced Placement Study (STEPS) Technical Report. *RP Group*.

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

Wise, S. L. & DeMars, C. E. (2010). Examinee Non-Effort and the Validity of Program Assessment Results. *Educational Assessment*, *15*(1), 27–41.

Wise, S. L. & Smith, L. F. (2016). The Validity of Assessment when Students Don't Give Good Effort. In G. T. L Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 204–220). New York: Routledge.

Wise, V. L., Wise, S. L., & Bhola, D. S. (2006). The Generalizability of Motivation Filtering in Improving Test Score Validity. *Educational Assessment*, *11*(1), 65–83.