Triangulated authentic assessment in the HEQCO Learning Outcomes Assessment Consortium

Jake Kaupp, Natalie Simper and Brian Frank
Queen’s University
kaupp@appsci.queensu.ca

Abstract – The Higher Education Quality Council of Ontario (HEQCO) has established a consortium of institutions committed to the development of useful learning outcomes assessment techniques and to their wide-scale implementation in their institutions. Queen’s University is one of three universities and three colleges of the consortium, and the Faculty of Engineering and Applied Science (FEAS) is participating due to familiarity with assessing learning outcomes as part of accreditation. The specific learning outcomes that are of interest to Queen’s are Critical Thinking, Problem Solving, Communication and Lifelong learning.

The goal of this three-year project is to assess the aforementioned general learning outcomes and cognitive skills using three assessment methods simultaneously: embedded course assessment, using meta-rubrics to score student artifacts, and using standardized tests/surveys. The study will document cost and time required to access each of these methods in specific courses, analyze correlation between scores from the three methods, and evaluate developments of the generic learning outcomes over the duration of a program. We aim to ensure that the work of outcomes assessment is sustainable, works within standard course contexts, and can be integrated into regular course activities. The paper identifies the goals of the project, current approach, and an example of data collection in one first-year engineering design course.

Keywords: Problem Solving, Critical Thinking, Communication, Lifelong Learning, Assessment, Learning Outcomes, Cognitive Skills

1. INTRODUCTION

The development of specific thinking, reasoning and communication skills is considered to be a hallmark of higher education. These skills and abilities are often referred to by a variety of terms: generic or general learning outcomes, transferable learning outcomes, transferable complex cognitive skills, and higher order thinking skills. The ability to solve problems, think critically and communicate effectively are of vital importance for students, employers, and policy makers[1]-[4]. The assessment of general learning outcomes is a high priority for institutions in response to stakeholder questions regarding the quality and value of higher education. This has led to a shift in higher education, in particular professional programs such as engineering, towards outcomes-based assessment and continuous improvement practises to demonstrate the development of these critical outcomes across a program or institution, and use the information to improve and enhance curriculum.

Despite this focus, there are considerable questions regarding how to assess the longitudinal development of general learning outcomes in an authentic, reliable and sustainable manner at the program and institutional level. To this end, the Higher Education Quality Council of Ontario (HEQCO) established a consortium of institutions, referred to as the Learning Outcomes Assessment Consortium (LOAC) with the following three commitments:

1. Develop valid and reliable assessment tools & techniques to measure general learning outcomes and cognitive skills.
2. Commit to the widespread, sustainable implementation of these assessment tools in their institution.
3. Share approaches and learnings with the consortium and other higher education institutions.

Three universities and three colleges were selected to the HEQCO LOAC: Queen’s University, University of Toronto, Guelph University, George Brown College, Durham College and Humber College. As part of the LOAC, each institution was directed to target general learning outcomes of specific interest to its institution. Queen’s University selected problem solving, critical thinking, written communication and lifelong learning as
the general learning outcomes. These outcomes were determined to be of critical importance to the institution as well as the students, as indicated in an exit poll of recent graduates (Figure 1) that revealed critical thinking, problem solving, leadership, writing and self-confidence were the five most important skills developed at Queen’s[5].

![Figure 1 - Student ranking of skills developed in a Queen’s education](Image)

The Faculty of Engineering and Applied Science is a very active participant in the project due to expertise in outcomes-based assessment with the CEAB accreditation mandate[6]-[8], and assessment of general learning outcomes in engineering[9],[10]. Three out of the four general learning outcomes selected by Queen’s are also part of the CEAB graduate attributes, and a significant argument can be made that elements of critical thinking are woven throughout all twelve graduate attributes. By participating in the project, we hope to continue to build upon the momentum and practises established to comply with the CEAB accreditation mandate, and offer expertise for a wider-scale rollout across all programs, and allow the university to develop internal processes for the implementation, management, and assessment of university-wide learning outcomes that recognize and enhance disciplinary expectations.

2. QUEEN’S UNIVERSITY APPROACH

The primary goal of this project is to pilot assessment of general learning outcomes and cognitive skills in selected courses within specific departments in the Faculty of Arts and Science and the Faculty of Engineering and Applied Science at Queen’s University. This project involves the following activities, all of which have been approved by the General Research Ethics Board at Queen’s University:

A. Assessing general learning outcomes and cognitive skills using the following three methods:

1. Using grades from course activities aligned with the targeted learning outcomes (type A in Figure 2 below)
2. Externally scored course artefacts (type B in Figure 2 below): artefacts aligned with the learning outcomes in the project will be scored using meta-rubrics like *Valid Assessment of Learning in Undergraduate Education* (VALUE) Rubrics, and other internally generated program-specific rubrics
3. External instruments (type C in Figure 2 below) including qualitative disciplinary think-aloud problems in small groups, and quantitative standardized tests of critical thinking and communication, including the CLA+ and *Critical Thinking Assessment Test* (CAT), and internally developed tools.
The study will examine the correlation between internal course assessment and these standardized external measures of general learning skills. Our expectations are that correlations will be more robust for curricula that are rated a priori higher in terms of effectiveness of general learning skills integration. If this is not the case, we have a strong foundation from which to make changes to our general learning skills integration standards.

B. Comparing and evaluating software tools to support assessment and evaluation of learning outcomes;

Faculty resistance to outcomes-based assessment is frequently related to additional workload involved in assessing, storing, and analyzing data from learning outcomes. A sustainable process for assessing and using learning outcomes must include processes that support and minimize additional workloads on faculty. A key support comes in the form of software, so this project will assess tools to manage outcomes-based assessment such as Desire2Learn and other assessment and management systems.

2.1 Assessment methods

Each of the three assessment methods in Figure 2 has specific benefits and drawbacks. The use of course specific criterion-referenced scoring, or rubrics, using course deliverables (A) provides excellent opportunities for formative feedback to improve student learning, as well as providing information to the instructor for the purposes of course improvement. Assessing general learning outcomes in this fashion also encourages faculty to consider how these outcomes are developed in their course and consider alignment between the curriculum and general outcomes. The drawback of this approach is a limited ability to assess longitudinal development over multiple years and concerns about interrater consistency and reliability.

The use of stand-alone standardized instruments (C) provides a means for the longitudinal assessment of learning outcomes over multiple years as well as an opportunity for intra and inter-institutional comparison. The validity and reliability of these instruments are well established, providing robust measures for learning outcomes at select times in a program. The drawbacks of standardized instruments are related to their cost, authenticity of the assessment, limited breadth of outcomes assessed, and issues regarding student motivation and test completion rates.

The use of general criterion-referenced scoring using key or signature assessments within a course (B) provides a means to assess outcome development over multiple years. This method uses rubrics that are intended not to assess outcomes in individual students, but assess performance in a specified cohort or program. These rubrics were developed by the Valid Assessment of...
Learning in Undergraduate Education (VALUE) project, a set of rubrics to measure essential learning outcomes using broadly shared criteria and performance levels for assessing student learning[11]. These rubrics can be used to longitudinally assess development over multiple years, avoiding motivation and completion rate issues of standardized instruments by using signature assessments or deliverables from a course experience. Adoption of these rubrics encourages instructors to establish alignment between curriculum, course outcomes and program outcomes outlined within these general criterion-referenced rubrics. Despite these traits, the use of these instruments present some drawbacks and challenges.

The VALUE project is relatively new and there are a relatively small number of validated rubrics. Adoption of this approach requires additional time and resources are required to score the course activities in this fashion, as they are on top of regular course grading. Lastly, as with any rubric, interrater consistency and reliability are a concern.

3. STUDY INSTRUMENTS

3.1 VALUE Rubrics

The VALUE Rubrics were initially developed through a project launched by the American Association of Colleges and Universities (AAC&U) to provide a valid assessment of learning in undergraduate education. These rubrics are broad, discipline-neutral descriptions of selected essential learning outcomes of undergraduate education. For each rubric, a panel of experts identified common themes and worked towards the development of discipline-neutral performance criteria. Their efforts focused on:

1. Performance criteria focuses on positive demonstration of outcomes rather than what was lacking
2. Performance criteria can be used to assess non-traditional modes of artifacts demonstrating student learning
3. Performance criteria are developed to assess summative displays of student learning rather than developmental or formative displays
4. Performance criteria are phrased in a manner as to be easily understood by non-experts

There are four levels of performance criteria, from the benchmark level of a student entering university to the capstone level of a student who has just completed their undergraduate experience. While the performance criteria and levels represent a consensus of experts and can be used in their original form, the rubrics require modification to foster alignment between course, program or institutional outcomes and to reflect the specific context in which they are used.

As part of the Queen’s approach, the original VALUE rubrics pertaining to critical thinking, problem solving, written communication and lifelong learning will be modified in the future to preserve alignment between course, program and standardized measures of each specific outcome as well as reflect the disciplinary expectations in which they are applied. Initially, significant work was put forth in using the VALUE rubrics to describe general indicators for each specific outcome.

These indicators were invaluable in communicating with course instructors by providing:

1. Greater understanding of what we were attempting to measure.
2. Whether or not the indicators were developed or assessed in their course.
3. Selection of a suitable student artifact to assess with the VALUE rubrics.

The indicators are shown below in Figure 3. The VALUE Rubric for lifelong learning was heavily modified in order to better reflect disciplinary and institutional priorities and to foster alignment between standardized measures, the VALUE Rubric and course expectations.
3.2 Standardized Instruments

3.2.1 Collegiate Learning Assessment (CLA+)

The Collegiate Learning Assessment (CLA+) was developed and administered by the Council for Aid to Education (CAE). It is an updated version of the collegiate learning assessment and addresses many critiques and concerns of the original. The CLA+ is reported to possess greater reliability and validity at the student level, and can be administered at the cohort level due to refined and increased granularity. This is an improvement from the previous version, which possessed poor student level validity and reliability, and was only recommended to be used as pre-post pairing when entering and exiting an institution. The CLA+ is a web-based instrument with students logging into a secure browser to take the test. The test takes 90 minutes to complete, with 60 minutes allotted to the performance task and 30 minutes allotted to the selected response questions.

The core of the CLA+, the performance task, is used to measure critical thinking, problem solving and written communication and remains the key focus of the CLA+. The performance task presents a real-world situation in which students assume an appropriate role to address the problem, identify a solution, or provide conclusions and recommendations resulting from careful analysis of the provided evidence. The student responses in the performance task are graded via rubric and are scored by an automated system on the following scales:

1. Analysis & Problem Solving
2. Writing mechanics
3. Writing effectiveness
4. Scientific & Quantitative Reasoning
5. Critical Reading & Evaluation
6. Critique an Argument

Overall, the performance task outcomes of the CLA+ are well aligned with the VALUE rubric outcomes for critical thinking, problem solving and written communication. While the additional selected response outcomes will provide a greater understanding of student reasoning, analytical and communication skills for both courses and programs.

3.2.2 Critical Thinking Assessment Test

The Critical Thinking Assessment Test (CAT) was developed collaboratively by Tennessee Technological University (TTU) and faculty from participating institutions with funding provided by the National Science Foundation (NSF). The primary goals of this collaboration were to develop a faculty-driven assessment tool to engage the faculty in meaningful, authentic
assessment of higher order skills and improving student learning through assessment. The CAT is based around a consensus core set of five skills that comprise critical thinking across a variety of disciplines and 12 representative areas in which to assess these skills[12]:

1. Evaluating Information
   1.1. Separate factual information from inferences
   1.2. Interpret numerical relationships in graphs
   1.3. Understand the limitations of correlational data
   1.4. Evaluate evidence and identify inappropriate conclusions
2. Creative Thinking
   2.1. Identify alternative interpretations for data or observations
   2.2. Identify new information that might support or contradict an hypothesis
   2.3. Explain how new information can change a problem
3. Learning and Problem Solving
   3.1. Separate relevant from irrelevant information
   3.2. Integrate information to solve problems
   3.3. Learn and apply new information
   3.4. Use mathematical skills to solve real-world problems
4. Communication
   4.1. Communicate ideas effectively

The CAT consists of 15 questions requiring both quick-response items (multiple choice, binary scale) and short-answer essay responses that take students 60 minutes to complete. Currently, the CAT is a paper-based test, with plans to move to a digital delivery system in the near future. The participating institutions own faculty score the CAT, through a “train-the-trainer” system and following a detailed scoring rubric. The CAT is a reliable and valid assessment, with authors providing studies into the psychometric properties of the test, rigorous comparison with other vetted instruments, and several published studies that have used the CAT. The CAT represents a well-aligned standardized instrument for the Queen’s approach, possessing a high degree of alignment with the VALUE rubrics and offering a range of information to assess student development for both course and program interests.

3.2.3 Motivated Strategies for Learning Questionnaire

The Motivated Strategies for Learning Questionnaire (MSLQ) is a survey instrument designed by Paul Pintrich to assess students' motivational orientations and their use of learning strategies. It consists of fifteen scales that can be used singly or together, with a total of 81 separate items answered on a Likert scale that ranges from “True of me” to “Not true of me at all”. The available subscales and the number of question that comprise each are illustrated in figure 4.

![Figure 4](image_url)

Figure 4 - Components of the Motivated Strategies for Learning Questionnaire (MSLQ). The highlighted scales are included in the Queen’s version of the MSLQ, the rest were excluded.
Upon careful analysis and reflection, select scales of the original MSLQ were removed from the final study instrument. This was done to establish and maintain alignment between the MSLQ and disciplinary priorities regarding learning orientation and self-regulation, and between the MSLQ subscales and the VALUE rubrics outcomes. Scales were also removed to eliminate duplicate measures (critical thinking) and items that did not align with the VALUE rubric outcomes (task value, test anxiety, effort regulation and peer learning). The final version of the MSLQ used of this study has 7 scales and a total of 58 questions. Removal and alteration of these scales does not affect the previously established validity and reliability of the MSLQ, nonetheless future plans include independent psychometric analysis on the scales in the final instrument.

4. INVOLVEMENT OF FACULTY OF ENGINEERING & APPLIED SCIENCE

With this being the inaugural year of the Learning Outcomes Project, data is being collected from first and fourth year course experiences in order to measure the incoming cohort and provide a benchmark for future comparison. The study includes courses from engineering, physics, psychology, and drama. Researchers met with each course instructor to inform them about the project, relate to them the goals and requirements of the project, and determine a sustainable approach for the course. These discussions:

1. Identified and mapped the study outcomes to course outcomes and assessments.
2. Identified which elements of the VALUE rubrics for each study outcomes were developed in the course.
3. Compared course outcomes with the VALUE rubric outcomes, modifying the rubric language where appropriate.
4. Identified the appropriate assessments to use as key/signature assessments for VALUE rubric scoring.
5. Determined appropriate times to embed introduction to the study, soliciting volunteers and administering standardized instruments

For the purposes of this paper, we are focusing on the implementation of the Queen’s approach in the first year engineering course, APSC 100 Module 1: Problem Analysis & Modeling. This course was selected because the course learning outcomes align well with the goals of this project.

4.1 First Year Engineering Design Course

Module 1 is a semester-long integrative experience that uses concepts from engineering sciences, natural sciences, and mathematics courses to solve complex open-ended problems. During the first week of the semester students are enrolled in a workshop to introduce them to time and study management skills, to help acclimate students to the expectations of a team-based design course. The module is structured around model-elicitating activities (MEAs)- problems used in class that are set in a realistic context that requires the learner to document not only the solution to the problem, but also their process for solving it[13]. The situations described in the MEAs require students to create and use a mathematical model of a physical system using MATLAB, and deal with professional issues including ethical dilemmas, conflicting information, and incorrect/missing information. This module focuses on developing problem solving, critical thinking, and communications skills[10].

The MEAs, being the central pedagogical and assessment instrument of the course, have their own specific outcomes. These outcomes focus, in higher detail than the course outcomes, on critical thinking, problem solving and communication within the context of solving realistic, open-ended, complex engineering problems. There was a high degree of alignment between the VALUE rubric, course and task (MEA) outcomes, which is outlined in Figure 5. This is primarily due to the efforts of the instructor, having already embedded critical thinking, problem solving and written communication outcomes in the course, and carefully maintaining the alignment between instruction and assessment, as demonstrated in the course and task outcomes.

Regarding the learning orientation and self-regulation outcomes, the dedicated workshops in the first week introduce students to effective time management and study skills and the instructor believes that the flipped nature of the course alongside the team and independent learning required by MEAs will foster student development in these areas.
Figure 2 above showed the three assessment methods used in this study. The section below shows examples of how these methods were used in APSC100 Module 1.

4.1.1 Assessment in APSC100

After the initial discussion with the instructor, it was agreed upon that the course rubrics used for the assessment of the MEAs would be adjusted to more accurately reflect the language used in the VALUE rubrics. Specifically, the performance criteria in the MEA rubrics was adjusted with the intent of the criterion level for the achievement of the outcomes in the MEA rubrics (or meeting expectations) informing the wording of the “benchmark” performance criteria for the VALUE rubrics within engineering.

With the alignment preserved through the restructuring of rubric language and performance criteria, a suitable assessment point was required. In APSC 100 Module 1, students submit two MEAs for grading and also work on a mini-MEA between the graded activities for the purposes of formative feedback. The second MEA was selected as the summative piece of student evidence for the assessment of select outcomes and will be scored using the VALUE rubrics during the summer of 2014 for critical thinking, problem solving and written communication.

Collecting the scores from the MEA rubrics, as well as the consenting student submissions for MEA was done in such a way to minimize impact on students and minimize work for the instructor course personnel. Research personnel were granted access to the course management system (Desire2Learn) enabling them to collect the rubric scores and artifacts for consenting students.

The third assessment method used was standardized tests/surveys. Using standardized instruments requires a significant investment of time and resources of a course instructor. In APSC 100 all students have a two-hour block of studio time, in which they typically work on course concepts and MEA in a computer lab. All students were instructed to bring their laptops to the studio session if they had one, if they didn’t a computer was provided for them. This was an ideal arrangement for administering a standardized test, as the first two hurdles of having enough time and enough available computers were addressed. The instructor agreed to fully embed the standardized tests within the course itself, with students obtaining a completion grade for the tests. The instructor
also took the time in the initial lecture with students to speak about the study and talk to students about the importance of higher order skills in their course experience and their discipline. By endorsing the project as a worthy research activity and by embedding the instruments the typical problems associated with standardized tests, such as low volunteer rates and motivational issues, can potentially be avoided.

With APSC 100 being a common first year course in the faculty of engineering, the large number of students enrolled (762) in the course afforded researchers the ability to have three separate cohorts, each writing a different standardized test (either the CLA+, the CAT or the MSLQ). During the first week of the semester, the instructor set aside the studio time so researchers could come into the class, introduce themselves and the study, issue and collect consent forms for participation, and administer the standardized test.

By using the approach outlined above, a volunteer rate of 84% was achieved with 640 students volunteering to participate in the study. Using only the consenting students, the cohort distribution had 250 write the CLA+, 170 students write the CAT and 210 students wrote the MSLQ. Lastly, we had an 18%, 141 students, volunteer to be contacted regarding participation in activities outside of the course experience (think aloud activities)

5. PRELIMINARY RESULTS & DEVELOPMENTS

The focus in the 2014 – 2015 academic year has been identifying proper courses to participate in the study, working alongside instructors to maintain a positive experience and foster alignment between course and project outcomes, and data collection in 4 first year courses and 7 fourth year courses. The VALUE rubric grading along with the results from the standardized tests will not be available until the end of the summer of 2014. However, the data collected from the MSLQ was readily available from engineering as well as the other participating first year programs. Upon collection and analysis, a trend emerged that led the researchers to some preliminary conclusions:

1. Students did not properly understand their own motivations and learning strategies and answered questions on what they wanted to aspire to, rather than what accurately reflects them.

2. The survey was poorly worded or confusing to students, which complicated their understanding

3. The survey is best used at a different time in student’s university education, and not as a pre-post measure

4. The students, at the beginning of their university education, have not yet formed an opinion on their motivation and learning strategies, having not yet applied them.

5. The MSLQ is designed to use at the course level and asks students to answer questions that refer to the course (e.g. “In this course..”) or a similar course (e.g. “In a course like this”). Most incoming first year students have never taken a university level course before, which differ significantly from high school courses, thus have a little to compare to. This is even more exacerbated in the case of unique courses (design courses, capstones, flipped classrooms etc.).

After consulting faculty members participating in the study, and soliciting the opinion of educational psychologists, researchers decided that pursuing the development of a custom approach would potentially address many of these issues. This new approach will continue to use the MSLQ as its foundation, as the instrument provides valid and reliable measures of motivation, self-regulation, self-efficacy and learning strategies. Additionally, adopting MSLQ as a framework and adapting the scale items in the development of a new approach will help preserve the efforts made to ensure alignment between the VALUE rubrics and the MSLQ. The researchers are hopeful that through these improvements, the modified survey will develop into a valid, reliable, and meaningful approach to assessing learning orientation and self-regulation aspects of lifelong learning.

6. CONCLUSIONS

The researchers hope that this project will achieve multiple interrelated goals. It should provide data to evaluate the reliability of data drawn from in-class assessment tasks, which is the most common approach used to assess graduate attributes in engineering programs across Canada. It will also document the costs and time required to run standardized instruments and scoring of course artifacts external to academic grades in comparison with course grading. Efforts in this project will continue over the next three years, following the current first year cohort through the university experience.
ACKNOWLEDGEMENTS

The authors would like to acknowledge support from Higher Education Quality Council of Ontario (HEQCO), the Office of the Provost at Queen’s University, and the DuPont Canada Chair in Engineering Education Research and Development.

REFERENCES


