

Fall 09-20-2021

Collaborative Assessment: Using Self-assessment and Reflection for Student Learning and Program Development

Ellen L. Flournoy

Quest University, ellen.flournoy@questu.ca

Lauren C. Bauman

Quest University, lauren.bauman@questu.ca

Follow this and additional works at: <https://www.cjsotl-rcacea.ca>
<https://doi.org/10.5206/cjsotl-rcacea.2021.1.14207>

Recommended Citation

Flournoy, E. L., & Bauman, L. C. (2021). Collaborative assessment: Using self-assessment and reflection for student learning and program development. *The Canadian Journal for the Scholarship of Teaching and Learning*, 12(1).
<https://doi.org/10.5206/cjsotl-rcacea.2021.1.14207>

Collaborative Assessment: Using Self-assessment and Reflection for Student Learning and Program Development

Abstract

As program-level assessment increasingly becomes an integral part of the higher-education landscape, so does the debate regarding the efficacy of current assessment methods. Traditionally, students do not participate in assessment—neither of their own learning nor of institutional or program efficacy. Our assessment process presents an alternative to traditional program-level assessment and is meant to improve student learning in two ways: (1) by asking students to reflect on their achievement of learning outcomes using evidence-based methods; (2) by providing assessment practitioners with authentic, contextualized data on which to make claims about curricula. This collaborative assessment process was designed to address the complex needs of a cross-curricular rhetoric program but responds to many general concerns about traditional assessment methods.

Au fur et à mesure que l'évaluation de programmes fait de plus en plus partie intégrante du paysage de l'enseignement supérieur, il en va de même du débat sur l'efficacité des méthodes d'évaluation actuelles. Traditionnellement, les étudiants et les étudiantes ne participent pas à l'évaluation, ni à celle de leur propre apprentissage ni à celle de l'efficacité des programmes ou de l'établissement. Notre processus d'évaluation présente une alternative à l'évaluation traditionnelle de programmes et a pour but d'améliorer l'apprentissage des étudiants et des étudiantes de deux manières : 1) en demandant aux étudiants et aux étudiantes de réfléchir à la manière dont ils et elles ont atteint les résultats d'apprentissage, à l'aide de méthodes basées sur l'évidence, et 2) en fournissant aux praticiens de l'évaluation des données authentiques et contextualisées sur lesquelles ils peuvent se prononcer sur les programmes d'études. Ce processus d'évaluation en collaboration a été conçu pour répondre aux besoins complexes d'un programme de rhétorique transdisciplinaire mais il répond également à de nombreuses préoccupations générales concernant les méthodes d'évaluation traditionnelles.

Keywords

program assessment, collaborative assessment, reflection, self-assessment; évaluation de programmes, évaluation en collaboration, réflexion, auto-évaluation

Why Are We Doing This? The Assessment Conversation

To whom are we, post-secondary educators, accountable? How do we measure success? What are acceptable standards for that measurement? Does the measurement itself influence the learning process? Program and institutional assessment talk revolves around questions such as these. Not surprisingly, the many answers and opinions can confuse rather than clarify, most often because they are bound up in other long-standing issues that can be quite emotional, such as faculty review and workload. This article intends to do two things: (1) rhetorically clear the air and respond to some criticisms of assessment in general and (2) present some alternative ideas and processes for outcomes-based program assessment. We will start with some ongoing, general criticisms of assessment and then summarize our own program assessment process, explaining as we go along how our process might address some of the aforementioned criticisms. After that, we will show how our process fits within some important conversations about learning and share some student work from one of our data collections. Finally, we will make some suggestions for implementation and explain the limitations of our project as well as what we are planning for our future efforts.

An author of a recent article in *The Chronicle of Higher Education* excuses postsecondary faculty who manufacture data for course-level assessment because “it turns out that the assessment program your college imposed on you was probably never going to improve anything” (Gilbert, 2018, para. 3). Set up as a polemic exposé of sorts, the article’s argument relies on a conspiracy theory that “people in the assessment world have known for some time that their work was not producing results” and “have blamed the failings of their ideas on faculty members” (para. 6). It is clear that this argument is forging an identification with frustrated faculty, confirming their worst collective fears, and most of all, indulging their revenge fantasies against assessment authorities by exposing them as peddlers of methods akin to “Rorschach tests” (Gilbert, 2018). Although there is much to criticize in the industrial complex of post-secondary education, this author’s approach poorly addresses the real issue of accountability: to taxpayers, to our communities, to ourselves, and most importantly, to our students.

David Eubanks, the “assessment insider” from whom the above author draws his critique, explains that his “conclusion, after seeing hundreds of real assessment reports . . . supervising assessment programs, and talking to many other assessment practitioners, is that *it is difficult to use assessment results because the methods of gathering and analyzing data are very poor* [emphasis in the original]” (2017, p. 4). However, even after explaining the prevalence of these data issues as well as decrying the bureaucracy that leads evaluators to checklists as opposed to deeper analysis, he offers a solution that has nothing to do with ending assessment practice. Rather, he calls for re-defining the “standards of practice” professionals use to evaluate the efficacy of assessments. We would like to share a non-traditional, program-level, collaborative assessment process that uses students’ reflection on their own achievement of learning outcomes to deepen their learning and provide data for program assessment. We propose that our nascent process also addresses some of Eubanks’ concerns.

How Can We Reflect and Collect?

Traditionally, students are silent, unwitting participants in assessment—either of their own learning or course, institutional, or programmatic efficacy. They respond to assignments or write tests without knowledge that their work is used for other purposes in addition to the calculation of grades. Their examples of work are also assessment “artifacts” used by faculty or administrators

to make broader arguments about the efficacy of programs meant to help them learn. We propose students' silence in traditional assessment is a wasted opportunity—for deeper learning, for modelling reflective practices, for faculty collaboration with more engaged students. Our proposed process attempts to take advantage of those opportunities by re-visioning data collection as a collaborative and reflective feedback loop that includes student engagement. Our goal is to deepen student learning and make them active collaborators in assessment processes for program development—as opposed to voiceless suppliers of “assessment artifacts.” This collaborative process asks students to reflect on their achievement of concrete, discrete, and measurable learning outcomes using evidence-based methods. Simultaneously, it provides institutional representatives with authentic data on which to make claims about and revisions to curricula. Although we designed this process to develop viable assessment practices to meet the complex needs of a cross-curricular rhetoric program at Quest University¹, many courses, programs, majors, and degrees across the disciplines rely on student learning outcomes. Because of the widespread use of student learning outcomes, we think this approach might be useful for a wide variety of educators. Figure 1 depicts our programmatic assessment process. Each step is explained in detail below.

1. In our effort to include students in program assessment, our program-level student learning outcomes are crafted to be accessible and useful to students. They are simple, and whenever possible, concrete and measurable.

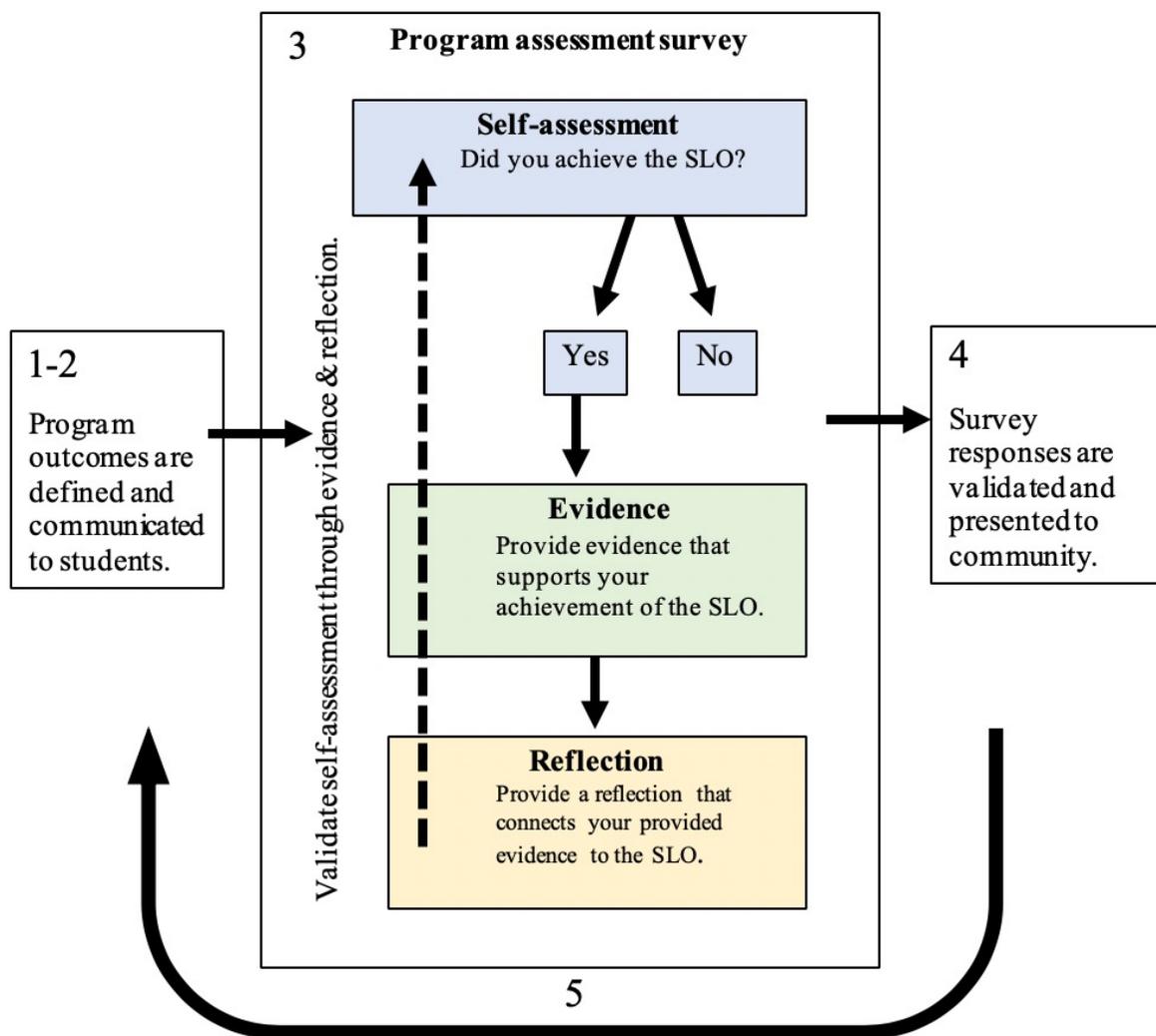
2. At the beginning of a course in which outcomes are to be assessed, students are given all of the student learning outcomes. Ideally, they are reviewed in conversation with the instructor.

3. At some later date, program representatives visit each participating class and lead a discussion about the assessment survey and facilitate student completion of the survey². Because the surveys are incorporated into class time, they are short and include questions that cover only two to three of the program-level learning outcomes. Data on each learning outcome is gathered in three to four questions that ask students to self-assess their own achievement of the learning outcome, provide textual evidence of that achievement, and finally, reflect on the relationship between their claim of achievement and the offered evidence. (See the Appendix for sample.)

¹ Quest University is a small, undergraduate, liberal arts and sciences university located in Squamish, BC. The student population is approximately 700, and there are no more than 20 students per class. The institution offers students an interdisciplinary undergraduate education in which each student is required to take 16 foundational classes across many major disciplines. The latter part of the degree program offers students the opportunity to design their focus area of study with 16 concentration-level courses. Quest University only offers one degree, a Bachelor of Arts and Sciences. The institution has no institutional research manager or department but has a faculty collective devoted to reflective teaching. This project was designed and executed within the institution's Rhetoric Across the Curriculum program, a cross-curricular program by which faculty intentionally embed rhetorical content in almost every offered course. Logistically, this means that faculty are teaching rhetoric “out of field,” which provides rich opportunities for inter- and trans-disciplinary learning. The described assessment took place in two mandatory courses, the required composition and capstone courses.

² The data collection described here has been approved by Quest University's Research Ethics Board, which required the authors to gather consent from participating students to use their survey answers for external purposes. All of the student work included here has been included with this consent.

Figure 1
Summary of our Program-level Assessment Process



Note. Numbers 1 – 5 correspond to explanatory text.

4. Program representatives code, validate, analyze, and report survey responses. We choose to code and quantify student responses to analyze general patterns, such as the number of students who achieved the given learning outcome. Quantifying our data helps us communicate our observations to our community. However, we also recognize the qualitative nature of our data set and use student responses to contextualize our findings. Using both qualitative and quantitative data addresses some common assessment concerns which we unpack later on.

5. This process is repeated until data has been gathered on all the program-level learning outcomes.

This proposed collaborative assessment process has the potential to deepen student learning as well as address multiple issues frequently cited in conversations about common assessment

practices. Before we share some examples of student learning, we will establish some common ground and refer to some pedagogical ideas that support our process.

What Are Our Key Terms?

First, we will define our use of “assessment” as it refers to multiple, intersecting processes up for discussion. Then, we will justify our assessment process within previous research to support our claim that our assessment process benefits students. Because we are developing a process to assess a rhetoric program, it is helpful to consider White et al.’s (2015) explanation of different kinds of writing assessment, which they organize from the micro to macro. Assessment can be responding to student work (for White, Elliot, and Peckham, work is writing) or evaluating how well a program meets its outcomes. Assessment can also refer to the ability of one program to prepare students for success in other articulated programs or on national tests. A fourth purpose has emerged as part of the movement for accountability in post-secondary education, and that is determining students’ workplace readiness.

We are mainly concerned with program assessment for the purposes of development—how well a program meets its outcomes and what that means for curriculum—as opposed to assessment of students. The following definition applies to our version of program assessment:

Assessment is a process in which rich, usable, credible *feedback* from an act—of teaching or curriculum—comes to be *reflected* upon by an academic community, and then is *acted* on by that community—a department or college—within its commitment to get smarter and better at what it does. (Marchese, 1997, n.p.).

We especially appreciate this definition because it describes the fluidity of the process by which academic programs can be adjusted based on reflection of practices and evidence.

However, one of the central features of our offered assessment practice is that we gather data from students’ self-assessment, which is their evaluation of their own work and learning. Using students’ self-assessment for program assessment supports our effort to increase student empowerment and independence by transparently communicating program outcomes so that students can evaluate their own learning as well as engage in program evaluation and planning. Because self-assessment de-centres authority and the evaluation of student work, it is easy for uninitiated educators to dismiss it over concerns about student judgment. However, times are changing, and educators have to reconsider the role of the student, who was once “a passive, powerless, often oppressed subject who is mystified by the process” but is now an “active participant who shares responsibility in the process, practices self-evaluation, reflection, and collaboration and conducts a continuous dialogue with the teacher” (Birenbaum, 1996, p.7). This shift away from testing toward the collaborative assessment of student work is supported by three decades of research that legitimates self-assessment (e.g., Keinänen et al., 2018; Sluijsmans et al., 1998; Topping, 2003) and presupposes calls for life-long learners who can self-evaluate in academic and professional contexts (Dochy et al., 1999; Falkichov & Boud, 1989; Peltier et al., 2005; Schön, 1987). Assessment of others and of self is a valuable skill that should be taught and studied. Even when a students’ self-assessment does not match the teacher’s assessments of them, self-assessment is still a valuable learning tool because it encourages autonomy, provides powerful feedback and the skills for effective self-evaluation for lifelong learning (Ross, 2006; Sluijsmans

et al., 1998). In addition, it can acculturate students into academic and professional contexts (Falkichov & Boud, 1989).

Self-assessment has even evolved for some into “sustainable assessment,” which can be defined as follows: “assessment that meets the needs of the present without compromising the ability of students to meet their own future learning needs” (Boud, 2000, p.151). It is a pedagogical practice by which student learning is evaluated in collaboration with the students themselves, who are engaged in processes by which they will learn to evaluate their own performance through self-reflective adjustment beyond any one course, discipline, or degree (Boud, 2000). Sustainable assessment is an effective framework to justify the unique assessment practices for which we are advocating because ours is a collaborative process by which students’ self-assessments and reflections are mined to create program-level data for the purposes of assessing curricular efficacy.

Further, self-assessment is beneficial because it requires students be able to maintain a robust practice of reflection. It is hard to argue against reflective practices in learning (Dewey, 1910, 1933; Schön, 1983). Reflection has been shown to increase students’ critical thought, making them more aware of their own learning processes, potential strategies for improving them, and the need to develop strategies to address difficulties (Colomer et al., 2013). Data suggests that reflection helps students increase their awareness of what they have learned and what they have left to learn (Colomer et al., 2013). Reflection is “central” to student learning (Yuek Ming & Manaf, 2014, p. 974) and is widely accepted as a necessary part of the contemporary education landscape: “Perhaps no other concept offers higher education as much potential for engendering lasting and effective change in the lives of students” (Rogers, 2001, p.55).

Writing as reflective practice is shown to be a transformative learning force (Gorlewski & Greene, 2011), especially when written reflections are guided (Yuek Ming & Manaf, 2014). Using written reflection in assessment of student work also has traction in pedagogical circles because of the benefits to students, mainly because assessments that require reflection make it more likely that students will continue to reflect after the assessment is over (Yancey, 1998). Furthermore, assessing students using their own reflective texts increases the validity of any assessment instrument because it requires students to self-assess and connect their own learning to the assessment itself (Yancey, 1998). Not only can our guided reflections help increase student learning and gather data for program assessment, these reflective writing activities offer reflective opportunities for faculty, too: “students’ written products provide a window into learning whereby instructors can refine their own instructional practices” (Black et al., 2000, p. 86). If we want students to learn the skills of self-reflection and self-assessment, then using self-reflection as a form of assessment is valuable because it provides a win-win opportunity for cyclical modelling of effective learning practices. Data collection should be explained as the mechanism by which faculty gather evidence on which they can reflect when making curricular decisions. Not only can this feedback loop benefit everyone through a practice of reflective adjustment, it can also reveal learning outcomes that aren’t apparent using traditional assessment methods, such as standardized tests (Yuek Ming & Manaf, 2014).

Traditionally, assessment of all kinds was removed from student experience until the communication of results, i.e., grades, but assessment practices as learning activities are now promoted in pedagogical conversations across the disciplines: “the role of assessment is shifting. Assessment currently perceives as a means to promote learning rather than monitor it, hence assessment . . . should be a tool for promoting deep learning activities” (Utaberta & Hassanpour, 2012, pp. 228-229). Like proponents of participatory assessment (Bruch & Reynolds, 2012), we

believe that this idea of student-centred assessment for learning should extend to the assessment of curricular programs that promote specific types of student learning, such as Quest University's Rhetoric Across the Curriculum Program.

In short, we are arguing that our students benefit from their participation in our program assessment because they self-assess their achievement of the program learning using an evidence-based, reflective method. Once the participating students have self-assessed their achievement of a learning outcome, they provide a piece of textual evidence to support their self-assessment, and then they write short reflections on the connections between their self-assessment and the evidence they chose to provide. However, we recognize that students' self-assessment alone is not a reliable gauge of their learning for the purposes of program assessment. Thus, we mitigate that concern by validating the accuracy of the self-assessment and its connections to the provided evidence and reflection.

How Can Self-Assessment, Evidence, and Reflection Work Together?

Challenging Assumptions in Student Learning

In many traditional assessment settings, anonymous student artifacts are often pored over by departmental faculty. After being scored somehow—usually through the use of rubrics—they become evidence to indicate the level of success of a course or program based on students' mastery of the intended learning outcomes. This kind of traditional assessment process supports some implicit assumptions that may or may not be true. For example, this kind of process further codifies the traditionally common belief that students are not qualified to participate in the evaluation of their own learning or evaluation of courses and programs. Another assumption built into that process is that one-off evidence of learning outcome achievement is enough to assume that the student has achieved deep learning and has (a) understood the skill and the concurrent requirements embedded in that learning outcome, (b) could transfer and apply that skill in different settings, and has therefore (c) met the learning outcome. If decontextualized student artifacts are indicators of the mastery of learning outcomes, then these assumptions are essential to the traditional assessment process.

We challenge these assumptions based on the data collected through our reflective self-assessment survey. Instead of faculty pulling student artifacts and searching for evidence of outcomes achievement, we ask students to find their own evidence and reflect on their achievement of the learning outcomes. This reflection often illuminates the artifact provided by the student in ways that would not be possible with simple faculty identification, thus revealing challenges and successes more deeply than one-off evidence alone.

Superficial Learning: A First Example

One student learning outcome of Quest University's Rhetoric Across the Curriculum program is that students will be able to "demonstrate the ability to summarize and paraphrase academic writing." We assessed this learning outcome in our survey by asking students three questions that requested that they self-assess, provide evidence, and reflect. First, they self-assessed their completion of the learning outcome by responding to the following question: 1. Did you summarize and/or paraphrase a text-based resource in your Rhetoric class? Then, they were asked to provide evidence to support their self-assessment: 2. If so, provide evidence by copy and

pastings an example from an assignment. Finally, they were asked to provide a reflection: 3. Reflect on your learning by explaining how someone else might guess that the lines you copied and pasted here are summarized or paraphrased material.

One student responded “yes” to question one and provided textual evidence of an article summary from an annotated bibliography assignment. Based on *only* the excerpt they provided, we verified the student performed a summary. In a traditional assessment setting, we would conclude that performance indicated that the learning outcome had been met. We would assume that because the student demonstrated their ability to summarize, they know both how to do it *and* have learned the requisite skills required for the task. As a result, we assume they would be able to demonstrate the deeper learning required for application and would be able to summarize future text-based sources. However, this assumption is challenged by the above student’s reflection on their learning. In response to question three: “I know I summarized this article efficiently because of how many times I read it over. I understood the material, as well where the information came from.”

This reflection indicates superficial learning for several reasons. First, the student’s reflection does not include any indicators of understanding what is required in a summarized or paraphrased section, such as a mention of the inclusion of the author or a shortened version of the cited author’s ideas in different words from the original. They connect their understanding of the material they read with their apparent ability to summarize. Although understanding a source is an important step in the research process prior to summarizing a source, understanding of the source content alone does not indicate the deep understanding of what makes a summary. Based on this reflective response, which adds context to this student’s evidence, we become less confident of the student’s understanding and ability to transfer the skill of summary to different contexts.

Deeper Learning: Another Example

In contrast to assessment processes that employ decontextualized evidence to demonstrate learning, the reflective self-assessment survey has the potential to show deeper student learning and not merely one-time achievement of a skill. For example, another student reported their achievement of the summarizing/paraphrasing learning outcome by responding “yes” to question one and provided the appropriate evidence in question two. This student then reflected on their learning:

This writing sample exhibits my ability to effectively paraphrase and summarize evidence from an outside resource. I can tell this section included summarized and paraphrased material because I mention the name of the original author without using quotation marks but still have a citation to show that the ideas come from someone else other than me. I re-state her argument in my own words (and much shorter!) to summarize the main idea I want to use as evidence.

This student emphasizes indicators of summarized/paraphrased material in the reflection about their evidence, mentioning that their sample cites the author and the text even though direct quotations were not used, re-states an argument in different words, and condenses a lot of information and/or big ideas into shorter statements. Taken together with their provided evidence, the contextualized reflective response shows the student’s understanding of the intended learning outcome and their deeper comprehension of summarizing and paraphrasing. Because of this

reflection's connections to the included evidence, we have more evidence to conclude that the student really does understand what summarizing and paraphrasing entails and will be able to apply the skill outside of this one instance as well as be able to recognize summarized and paraphrased material in others' work. Furthermore, and perhaps most importantly, the student's self-assessment and reflective process have likely deepened their learning (Colomer et al., 2013). Instead of learning how to summarize and/or paraphrase through trial and error by being asked to "incorporate outside resources," we ask the student to think about what is required for summarizing and paraphrasing outside of the one context seen in their chosen example.

How Can You Think Ahead? Intentional Setup Contributes to Program Assessment

Institutions that want to build an assessment process may feel like they are behind when in actuality, they are in a prime position. They have the opportunity to stack the deck against top-down approaches that may result in the "random shuffling of educational practices" and "*post-hoc* justification of a change that is desired for another reason" (Eubanks, 2017, p. 9). Ultimately, assessment processes are most useful when they germinate a desire for improvement on a grass-roots level, and therefore, institutions like ours, who are just beginning to think through assessment processes are in a privileged position to "build in assessment from the beginning" (International Network of WAC Programs, 2014, p. 4). Although it is not impossible to retroactively adjust, it is simply easier to begin as you mean to go on. Two ways to effectively stack the deck, either retroactively or from the beginning, are by using disciplinary frameworks and matching purposes and methods.

Disciplinary Frameworks

Using disciplinary frameworks is another way to focus and guide program creation from inception. Disciplinary frameworks are research-based schemas that can locate both outcomes as well as assessment practice to prioritize rigour without creating insurmountable barriers to assessment—even in times of scarce resources. Some of the most experienced practitioners of writing program assessment claim consensus-based disciplinary framework documents in Rhetoric and Composition have changed the way we think about the "writing construct" (White et al., 2015, p. 17) and improved the assessment of student writing, assessment of writing programs, and outcomes-based curricula.

Our program is akin to a writing program that also includes other modes of expression, such as verbal and visual modes, so we relied on germinal disciplinary frameworks for Rhetoric and Composition found in the Council of Writing Program Administration's *WPA Outcomes Statement* (2014) and *Framework for Success in Postsecondary Writing* (2011) to build our cross-curricular rhetoric program's outcomes as well as find potential solutions to apparent gaps in students' ability to demonstrate achievement of program outcomes. Using a disciplinary framework to guide our program assessment was an effective starting point and one that helps prioritize more nuanced disciplinary ideas of programmatic success over other ideas that were not as meaningful to us, such as value-added arguments or the isolated requirement that statistical relevance be a criterion for all assessment exercises.

Matching Purpose and Method

Because our institutional culture already incorporates reflection and self-assessment in many ways, including embedding it in pedagogy, curriculum, and faculty development and review, it seemed most appropriate to align our program assessment process with these cultural values. Furthermore, aligning our purposes and methods ensured that we were initiating a process that was more likely to result in something we could use, as Rossi and Freeman argue:

the evaluator should choose the best possible design from a methodological standpoint, having taken into account the potential importance of the program, the practicality and feasibility of each design, and the probability that the design chosen will produce useful and credible results. (Rossi & Freeman as cited in Upcraft & Schuh, 2002, p. 19).

We were trying to avoid a situation in which we collected data and decided its purpose afterward, which leads to the random shuffling and *post-hoc* practices mentioned by Eubanks. As explained in our previous section, our assessment model is intended to provide students an opportunity to deepen their own learning through self-assessment and guided reflections. We prioritized their active collaboration in program assessment, and our survey is an inexpensive way to start that work in *media res* until we can achieve our aspiration of an embedded rhetorical portfolio process.

We also took great care to align our methodology with our intended plans for using the data. We contend that the rich student responses to the reflective questions in our survey are more appropriate to our purposes for student learning and program development, even though the qualitative data set cannot be statistically scrutinized in the same way as a pre-test/post-test type of assessment. We are not using this data to support any “value-added” arguments about our program, specific hypotheses, or arguments about faculty or student performance. We are using our data to initialize and formalize a cross-curricular program using disciplinary best practices and research.

Furthermore, our institution is new to the assessment game, and this process models a student-centred reflective adjustment process and provides an entry to increased dialogue around assessment amongst faculty and students. Rather than communicating about conclusive data and summative figures, we plan to use our data to support conversations about future program development and increased accountability. Program review will come later once there is evidence of alignment of program standards and curriculum.

How Did You Arrive at that Figure?

General Assessment Concerns and Our Own Limitations

Eubanks, our aforementioned “assessment insider,” explains common data gathering problems lead to data sets that are too small and decontextualized, a lack of statistical rigour, and data analysis corrupted by “common-sense inference.” Not surprisingly, some of our assessment limitations align with some of these broader concerns about assessment, such as the lack of cross-disciplinary statistical validity and contextualized data analysis. We find it meaningful to address general assessment concerns as we address our assessment’s limitations because it helps us create an assessment discourse for our institution. Furthermore, considering our limitations within the context of common concerns about assessment allows us to negate or confirm our original

assumptions about our process, as well as be transparent about the potential challenges of implementing such a system.

Small Data Sets

Eubanks takes issue with small data sets used to make generalizations about large groups of students. Traditionally, the process of collecting, compiling, and searching students' rhetorical artifacts for evidence of learning outcomes achievement is labor-intensive and time consuming. Institutions usually lack the resources to assess a full interdisciplinary portfolio for every student and are faced with a trade-off: big sets of superficial data or small sets of deeper data. This trade-off is less than desirable as both setups lead to the loss of crucial information.

Our model uses electronic surveys that do not require extensive labour output until the validation and analysis stages. After validation of student responses, we quantify general patterns, such as how many students achieved a given learning outcome, in an effort to explain and communicate our observations. However, the survey framework still allows us to analyze the qualitative data (e.g., students' reflections) which contextualize these larger patterns. Further, we are able to collect data from a large percentage of enrolled students because the survey is embedded into the curriculum as a valuable learning opportunity. Not only does the survey save time and resources, but the collaborative model in which students are active participants requires that the students themselves direct those validating the data to the pieces of evidence they, as the authors of the artifacts, feel is most applicable to demonstrate their learning as opposed to systems that requires faculty hours to identify appropriate evidence amongst collected artifacts. Packaging the activity as a survey further saves time as responses are exported as an organized bundle of thoughts ready for analysis.

Assessment or Research: Statistical Limitations

Eubanks' concerns about data analysis is a warning about "common sense inference," the analysis of disparate number sets for the purpose of justifying change without incorporating the nuanced information available in each specific case. According to Eubanks, common sense inference is oversimplification in which analysts blindly compare data sets year after year, justifying numbers based on reasonable assumptions but with little methodology to determine validity and reliability of results. He argues for statistical rigour and goes on to explain that a lack of good data leads assessors to make arguments based on bad data (Eubanks, 2017).

While we see the logic in Eubanks' analysis of the common-sense inference, we find the conversation more nuanced. Because we experience first-hand the factors that limit institutional data collection and analysis, we find pragmatic arguments more actionable at our institution. Our philosophy of assessment aligns with Upcraft and Schuh (2002), who explain the differences between assessment methods and research methods. In short, they argue that data collection and analysis practices exist on a spectrum, depending on disciplinary norms and the purposes of the data. Because administrators have to make decisions whether or not data exists, it is clear that a lack of assessment data can lead to decisions based on personal evidence, such as intuition, bias, prejudice, and personal proclivities. Thus, even if data cannot meet the statistical standards of all disciplines, it is most often true that data is more useful than an absence of data. However, assessment data must always be communicated along with its limitations so consumers of assessment results can better understand its implications. Under these considerations, Upcraft and

Schuh make it clear that an offered assessment process is seen as a supporting piece of evidence instead of a reproducible theory.

Although Eubanks focusses on all the ways that data can be “bad,” his argument ends with a conclusion about critical reflection: “The goal is not perfect measurement; the goal is to not fool ourselves” (p. 12). Constant and critical awareness of the purpose of assessment data in addition to communicating the limitations of the assessment processes are ways to avoid “fooling ourselves.”

We accept that the data we collected for program-level assessment and the subsequent data analysis involved a degree of uncertainty and error. Because the data was qualitative in nature and required us to develop a code that best represented the existing patterns, we used a combination of *a priori* and emergent coding techniques. Thus, our positions as assessors inevitably influenced our interpretation of student responses. To counteract our biases, we went through many rounds of independent coding, requiring discussion and consensus when coding was not aligned. Due to the qualitative nature of our data, the potential for measurement error in the quantitative results is difficult to quantify. By attempting to quantify our subjectivity, we would do more harm by implying a level of statistical rigour that is impossible to achieve in the context of our assessment program. Specifically, we identify validated large-scale patterns present in the data set created from student survey responses to questions about the rhetoric program outcomes, which are already steeped in disciplinary frameworks. Some student responses offer insight into strengths and potential problems students may encounter as they move through the program. We then use disciplinary best practices and research in conjunction with those locally-sourced data set patterns to make arguments about potential curricular and faculty development. The combination of disciplinary rigour and large-scale percentage patterns allow us to be more certain in the benefits of resulting proposed changes.

Decontextualized Data

Eubanks also cautions against decontextualized data, recommending that student data be connected to other information markers that create possibilities for more nuanced interpretation analysis. For example, he claims that cause and effect arguments and claims about changes in student learning can most effectively be made with information about student traits and experiences, such as previous grades and length of time between sequenced courses.

Without an institutional research department to collate student information and create a process by which to apply it to other data sets, contextualizing student data in traditional ways becomes a technological obstacle. Although we lack the ability to match individual student characteristics with collected data, we are adjusting our next data collection to contextualize student data in another way—by asking students who report that they did not achieve a learning outcome achievement why they cannot demonstrate their learning. Our hope is that this adjustment will give students the opportunity to explain their individual situations, such as previous gaps in education or absences, that prevented them from achieving a learning outcome. Although asking students to self-report without validation cannot result in ironclad data about individual students, their explanation of the absence of learning outcome achievement can shed light on certain aspects of the curriculum or their experiences of it. It will also be easier to verify patterns in those justifications, especially if those patterns are connected to a particular curricular location. For example, the development of research skills is challenging at our institution because the duration of intensive course is only three and a half weeks. Entry-level students who report

underachievement in the entry-level research outcome for personal reasons, such as illness or a lack in previous education, help assessors create arguments for the locations of entry-level research skills in multiple curricular locations and not simply in the introductory rhetoric class. We look forward to the data results of what we are calling the “contextualized no,” another attempt to avoid “fooling ourselves” with our data.

Drawing Conclusions

We are not suggesting our collaborative program-assessment process is ideal for every post-secondary environment or even claiming it is a complete evaluative picture of our own program. Rather, our process provides us an opportunity to reflect on the larger assessment conversations, collect assessment data, and encourage student learning. This paper shows our efforts to shift program assessment from a bureaucratic, “box checking” requirement to a valuable learning opportunity for both students and assessment practitioners. Although the process presented here cannot paint a complete picture of a program, we see opportunities for replication in institutions that want to begin to incorporate students into any part of the assessment cycle, from course-level to institutional-level assessment. Our institutional context includes flexibility in program design and evaluation but does not include collaboration with an institutional research team. What the successful completion of this multi-year assessment shows is that interested parties could attempt collaborative, reflective assessment methods or easily scale up this project with a small increase in assessment resources. We have also found evidence, as explained in the offered student examples above, that guided reflection can provide a context for students’ self-assessment of learning that allow us to hold student learning to a standard higher than is possible with more traditional, faculty-centred models.

There are also discussions to be had about the utility of in-class activities for data collection. Part of our argument is that students can benefit from collaborative assessment, so we feel justified in including the data collection as a required class activity. Students learn about program assessment and have an opportunity to self-assess and reflect on their learning. We also conduct assessment of our program outcomes on a rolling calendar, which means that only two to three outcomes are being assessed during any given cycle. That practice ensures short surveys that all students can complete in a short time, so students who do not feel that they are benefitting from the reflective opportunity can at least meet the requirement quickly. Ideally, collaboration would not be a coercive exercise, but systems for voluntary participation across the data collection process create difficult and sometimes expensive hurdles, like incentivization programs and complicated collection processes heavy with their requirements of people power. Soliciting voluntary participation without incurring substantial expense would be a conversation for the future in our collaborative model of assessment.

References

- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & Filip J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-29). https://doi.org/10.1007/978-94-011-0657-3_1

- Black, R. S., Sileo, T. W., & Prater, M. A. (2000). Learning journals, self-reflection, and university students' changing perceptions. *Action in Teacher Education*, 21(4), 71-89. <https://doi.org/10.1080/01626620.2000.10462982>
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167. <https://doi.org/10.1080/713695728>
- Bruch, P. L., & Reynolds, T. (2012). Ideas in practice: Toward a participatory approach to program assessment. *Journal of Developmental Education*, 35(3), 12-22.
- Colomer, J., Pallisera, M., Fullana, J., Burriel, M. P., & Fernández, R. (2013). Reflective learning in higher education: A comparative analysis. *Procedia - Social and Behavioral Sciences*, 93, 364-370. <https://doi.org/10.1016/j.sbspro.2013.09.204>
- Council of Writing Program Administration. (2014). "WPA Outcomes Statement for First Year Composition: 3.0." <http://wpacouncil.org/positions/outcomes.html>
- Council of Writing Program Administration, National Council of teachers of English, & National Writing Project. (2011). *Framework for success in postsecondary writing*. <http://wpacouncil.org/files/framework-for-success-postsecondary-writing.pdf>
- Dewey, J. (1910). *How we think*. Heath & Co.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Heath & Co.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. <https://doi.org/10.1080/03075079912331379935>
- Eubanks, D. (2017). A guide for the perplexed. *Intersection*, 4-13.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395-430. <https://doi.org/10.2307/1170205>
- Gilbert, E. (2018, January 12). An insider's take on assessment: It may be worse than you thought. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/An-Insider-s-Take-on/242235>
- Gorlewski, J., & Greene, K. (2011). Research for the classroom: The power of reflective writing. *The English Journal*, 100(4), 90-93.
- International Network of WAC Programs. (2014). *Statement of WAC principles and practices*. <https://wac.colostate.edu/principles/statement.pdf>
- Keinänen, M., Ursin, J., & Nissinen, K. (2018). How to measure students' innovation competences in higher education: Evaluation of an assessment tool in authentic learning environments. *Studies in Educational Evaluation*, 58, 30-36. <https://doi.org/10.1016/j.stueduc.2018.05.007>
- Marchese, T. J. (1997). The new conversations about learning: Insights from neuroscience and anthropology, cognitive science and work-place studies. In *Assessing impact: Evidence and action* (pp. 79-95). American Association for Higher Education. <https://people.kth.se/~gunnarj/AAPORTFn/PED/ped010419.html>
- Peltier, J. W., Hay, A., & Drago, W. (2005). The reflective learning continuum: Reflecting on reflection. *Journal of Marketing Education*, 27(3), 250-263. <https://doi.org/10.1177/0273475305279657>
- Rogers, R. R. (2001). Reflection in higher education: A concept analysis. *Innovative Higher Education*, 26(1), 37-57. <https://doi.org/10.1023/A:1010986404527>
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, 11(10), 13.

- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1998). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*, 1(3), 293-319.
<https://doi.org/10.1023/A:1009932704458>
- Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87).
https://doi.org/10.1007/0-306-48125-1_4
- Upcraft, M. L., & Schuh, J. H. (2002). Assessment vs. research: Why we should care about the difference. *EBSCO Publishing, About campus*, 6.
- Utaberta, N., & Hassanpour, B. (2012). Aligning assessment with learning outcomes. *Procedia - Social and Behavioral Sciences*, 60, 228-235. <https://doi.org/10.1016/j.sbspro.2012.09.372>
- White, E., Elliot, N., & Peckham, I. (2015). *Very like a whale: The assessment of writing programs*. University Press of Colorado. <http://www.jstor.org/stable/j.ctt130hkzt>
- Yancey, K. B. (1998). *Reflection in the writing classroom*. All USU Press Publications.
https://digitalcommons.usu.edu/usupress_pubs/120
- Yuek Ming, H., & Manaf, L. A. (2014). Assessing learning outcomes through students' reflective thinking. *Procedia - Social and Behavioral Sciences*, 152, 973–977.
<https://doi.org/10.1016/j.sbspro.2014.09.352>

Appendix

Assessment Survey Questions

Rhetoric (Introductory Composition Course) Questions

1. ANSWER OUTCOMES QUESTION

Did you develop strategies to overcome research-related barriers* in order to find resources, such as articles, for your projects?

* A research-related barrier is anything that causes problems during research, such as difficulties finding effective database search terms, problems identifying academic articles, or a whole host of other things. Just think about anything that frustrated you while researching, and you'll have thought up a "research-related barrier."

Please note that failure to remember or implement proper citation format (MLA, APA, Chicago, etc.) does not count as a research-related barrier.

2. PROVIDE EVIDENCE

If your answer to the last question was "yes," please go back to your Rhetoric projects, and find a resource you located after struggling in some way. Copy and paste that full citation from that project's reference page here. Don't worry about formatting. Please do not give an example of an in-text citation but rather a full citation as seen in the example below.

Here is a sample response: Murphy, Raymond. "The media construction of climate change quiescence: veiling the visibility of a super emitter." *Canadian Journal of Sociology* 40.3 (2015): 331-54. Web. 9 Feb. 2016.

3. REFLECT ON YOUR LEARNING

Then explain in one to two sentences what problem (the research-related barrier) you encountered while locating that resource.

Here is a sample response: The problem I encountered was that I was looking for specific information for my research paper and I was not finding it through searching the library's catalogue.

4. REFLECT ON YOUR LEARNING

In another 1-2 sentences, explain how you overcame the problem you listed above.

Here is a sample response: I overcame the problem by approaching the process from a different angle. I started looking at related scientific articles and the references they used, and using the

command find option in my computer I skimmed dozens of papers until I found one that hit on key words I needed.

5. ANSWER OUTCOMES QUESTION

Did you write a Rhetoric Action Plan goal for improving the ways you incorporate feedback to revise your drafts or improve later projects?

A goal for "improving the ways you incorporate feedback to revise your drafts or improve later projects*" could be any goal about organizing feedback, implementing feedback, or the practice of changing a paper, image, or presentation in response to feedback.

*Please note that we are looking for a goal about your process for incorporating feedback and not a goal inspired by feedback you received during your Rhetoric block. Please see example in #6 for clarification.

6. PROVIDE EVIDENCE

If your answer to the last question was "yes," please copy and paste any goal from Step 3 of your Rhetoric

Action Plan that is about improving the ways you incorporate feedback to revise your drafts or improve later projects. Don't worry about formatting.

Here is a sample response: RAP GOAL: I will take the time to organize the feedback I receive so I can make better decisions about using that feedback in future drafts or assignments.

7. REFLECT ON YOUR LEARNING

Then explain in two to four sentences how that goal will help you to incorporate feedback into drafts or improve later projects.

Here is a sample response: Most of the time, when I get feedback on a paper, I read through the teacher's margin comments without actively doing anything other than thinking about my grade and trying to make it match the number of negative comments. What that means for my revisions is that I only change how I write if I've seen the same comment over and over. From now on, every time I get feedback from anyone, I'll organize it into categories that I've created, and then when I want to revise or write something new, I can go back to those categories to figure out what I'm supposed to be working on so I don't make the same mistakes.

8. ANSWER OUTCOMES QUESTION

In your Rhetoric block, did you summarize* and/or paraphrase** a text-based resource, such as an article?

* To summarize something is to re-write it in a shorter version.

** To paraphrase something is to re-write it using different words.

9. PROVIDE EVIDENCE

If your answer to the last question was "yes," please copy and paste a few lines that show summarized or paraphrased information from a text-based resource.

Here is a sample response: Vale discusses the increased attention being given to grass-roots approaches to climate change mitigation because it's become clear coordinated international response is not working (12). She proposes that climate change policy is no longer a feasible issue for global collaboration, and therefore individual actors must shoulder more of the solution (15).

10. REFLECT ON YOUR LEARNING

Then explain in two to four sentences how someone else might guess that the lines you copied and pasted here are summarized or paraphrased material.

Here is a sample response: This writing sample exhibits my ability to effectively paraphrase and summarize evidence from an outside resource. I can tell this section included summarized and paraphrased material because I mention the name of the original author without using quotation marks but still have a citation to show that the ideas come from someone else other than me. I re-state her argument in my own words (and much shorter!) to summarize the main idea I want to use as evidence.

Keystone (Concluding Capstone Course) Questions

1. ANSWER OUTCOMES QUESTION

One of the goals of the Rhetoric Across the Curriculum program is that you learn to successfully negotiate genre conventions in a variety of contexts. Can you compose a list of appropriate genre conventions* appearing in the written portion of your Keystone project?

*Genre conventions refer to the "rules" that go along with certain categories of communication. For example, a letter is a genre and requires you to have a greeting, such as "Dear Jason."

2. PROVIDE EVIDENCE

If your answer to the last question was "yes," please name your chosen genre and then list 4-6 genre conventions appearing in the written portion of your Keystone project.

Here is a sample response: My genre is a humanities-style analysis of a film. Some of the genre conventions I included in the written portion of my Keystone are citations, section headers, shot descriptions, small sections of plot summary, and discussions of symbolism.

Use the text box below to write the following sentences, filling in the blanks with your own answers:

My genre is _____. Some of the genre conventions I included in the written portion of my Keystone are _____, _____, _____, _____, _____, _____.

If you answered "no" to question #1, please skip to question #4.

3. REFLECT ON YOUR LEARNING

Write 4-6 sentences that explain (guesses are fine) why any three of the conventions you listed are part of your chosen genre.

Here is a sample response: Citations are required in this genre because incorporating outside research is a convention that must result in providing bibliographic information on resources. Section headers are organizational pieces that help structure the argument and break up sections of text. Discussions of symbolism help to connect the formal elements of the film back to a deeper meaning.

4. ANSWER OUTCOMES QUESTION

One of the goals of the Rhetoric Across the Curriculum program is for students to intentionally utilize frameworks as tools. Can you compare the structures you chose for your written Keystone and presentation, and explain why the structures differ?

5. PROVIDE EVIDENCE

If your answer to the last question was "yes," please write 3-6 sentences that list some of the differences between the structure of the written portion of your Keystone and your Keystone presentation.

Here is a sample response: My written Keystone used the traditional science paper structure, so it was divided into the following sections: abstract, introduction, methods, results, discussion, conclusion, and references. On the other hand, my presentation was structured completely differently to keep a mixed audience engaged. I had a hook that attempted to establish common ground with audience members and didn't get much further than talking through why my topic was important and locating it in the conversations of the field. I also discussed my main results and reviewed what those results meant for the field.

6. REFLECT ON YOUR LEARNING

Then explain in 3-5 sentences why your written Keystone and Keystone presentation have different structures.

Here is a sample response: It would be quite difficult to keep a mixed audience engaged in a talk that chronologically walked through a scientific paper starting with the abstract and ending with

the references. That order has a lot of meaning for scientists in the field, and they would certainly be able to follow it, but mixed audiences need to hear more about why a study is important and why it should matter to them. A mixed audience would probably get lost in the level of detail orientation required by a results section written for disciplinary specialists and can usually absorb much less as they listen to a talk; sometimes a general overview and a sentence or two of “takeaway results” is all that can be quickly processed.

7. ANSWER OUTCOMES QUESTION

One of the goals of the Rhetoric Across the Curriculum program is that you understand rhetorical creation and argumentation as processes. Gathering feedback and implementing that feedback to make revisions should be part of any rhetorical process. One way you did that in the Keystone block was by completing a Rhetoric Action plan process, sharing your Keystone goals with the members of your Rhetoric feedback group, and asking for feedback targeted to those goals. Did you use your Keystone Rhetoric Action plan goals to solicit targeted feedback from your peers?

8. PROVIDE EVIDENCE AND REFLECTION

If your answer to the last question was "yes," please use the text box to write the following sentences, filling in the blanks with your own answers.

One of my Keystone Rhetoric Action Plan goals was _____. One piece of targeted feedback I received from a Rhetoric Feedback group member was _____. I was _____ with my ability to implement that targeted feedback into my written Keystone.

Here is a sample response: One of my Keystone Rhetoric Action Plan goals was to improve the incorporation of my chosen images. One piece of targeted feedback I received from a Rhetoric Feedback group member was to explicitly connect all images back to my argument by adding multiple sentences that addressed each image. I was only moderately pleased with my ability to implement that targeted feedback into my written Keystone because although I followed the advice and thought it improved my argument, I felt the approach ended up being too “blocky” but I ran out of time to continue revising.