Abstract

Developing program-level assessment systems creates an opportunity for faculty to think deeply about their student learning objectives while creating structures that provide meaningful and actionable feedback. Unfortunately, many programs lack the expertise and/or support at the programmatic or institutional level to create efficient and effective assessment systems. We describe the development of a learner-centered programmatic assessment system. A backward design process enabled faculty to discuss their curriculum and identify in-class signature assignments that provide assessment data while a systems thinking approach helped faculty integrate their course, program, and institutional assessments. The resulting programmatic assessment system facilitates longitudinal tracking of individual students and cohorts within six program learning objectives through the use of a rubric that maps signature assignments within courses onto the program learning objectives (the megalo-rubric). Each semester, stakeholders analyze the student learning evidence and use those data to drive curricula reform at the course and program levels.

AUTHORS

Geoffrey M. Bowers, Ph.D.
St. Mary's College of Maryland

Pamela S. Mertz, Ph.D.
St. Mary's College of Maryland

Kelly Y. Neiles, Ph.D.
St. Mary's College of Maryland

Daniel T. Chase, Ph.D.
St. Mary's College of Maryland

Andrew S. Koch, Ph.D.
St. Mary's College of Maryland

Randolph K. Larsen III, Ph.D.
St. Mary's College of Maryland

Shanen M. Sherrer, Ph.D.
St. Mary's College of Maryland

Troy K. Townsend, Ph.D.
St. Mary's College of Maryland

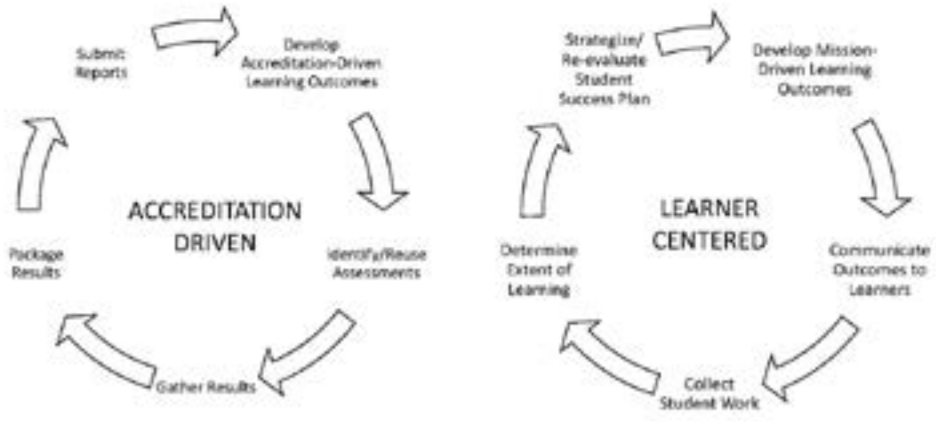
CORRESPONDENCE
Email
kyneiles@smcm.edu

A Programmatic Assessment System that Features Signature Assignments and a Longitudinal Rubric

Well-designed assessment systems provide data that programs use to (a) increase the equity of and (b) improve the student learning experience through curricular change (Kuh et al., 2015). Unfortunately, assessment systems are often created in response to administrative directives that provide little to no guidance on how to develop a useful and meaningful assessment system (Bowers et al., 2022). The results are often "check-box" assessment systems used primarily as a reporting mechanism to obtain/retain accreditation (Kuh et al., 2015). In this view, assessment is disconnected from the process of teaching and learning and becomes an outward-facing add-on, frustrating faculty who view assessment as meaningless and bureaucratic (Kuh & Hutchings, 2014; Jankowski & Marshall, 2017). Campus administrators have largely missed the opportunity to frame assessment in terms of deep programmatic reflection that promotes alignment of the courses within a program and understanding of how that program fits within the students' overall learning experience (Ewell et al., 2015).

To address these concerns, faculty perspective must change from an accreditation-centered to a learning-centered view of assessment (Jankowski & Marshall, 2017). In an accreditation-centered approach, institutions report the data that accreditors want to hear (Figure 1). Accreditation-centered approaches require the institution to demonstrate, with evidence, conformity to a standard established outside of the institution (Ewell, 2009). The incentive for the institution is to look as effective as possible--deficiencies must be

Figure 1
Comparison of Accreditation-Centered Versus Learner-Centered View of the Assessment Cycle



Note: Adapted from Jankowski and Marshall (2017).

avoided lest the institution receive a warning from their accrediting body. In contrast, within a learner-centered approach, assessment is driven by the mission defined by the program and institution. A learner-centered approach incentivizes discovering what is and is not working in the educational process so that it can be improved-deficiencies in performance must be detected and reported so they can be acted upon.

The capability for assessment systems to facilitate institutional change was addressed recently by Singer-Freeman and Robinson through the lens of grand challenges (2020a; 2020b). Using characteristics such as "is extremely hard to do, yet doable" and "would produce positive outcomes potentially affecting large numbers of people" from previous work on grand challenges (Gould, 2010; Stephan et al., 2015) Singer-Freeman and Robinson examined publications, websites, discussion boards and blogs from 2015–2019 and identified four grand challenges in assessment. Specifically, the four grand challenges are: (a) use assessment findings to increase equity, (b) use assessment findings to direct immediate pedagogical improvements, (c) produce visible and actionable assessment findings that drive innovation, and (d) examine changes in institutional effectiveness (including student learning) over time (Singer-Freeman & Robinson, 2020b). If programs (and institutions) develop new learner-centered assessment systems that address these challenges, the result will be a shift from compliance-driven assessments toward assessment as a tool to improve courses, programs, and institutions.

This paper describes a learner-centered program assessment system that works within a larger institutional assessment system and engages faculty members with Singer-Freeman and Robinson's grand challenges. We describe how the system was developed, how it is used to collect longitudinal data about student learning, and how those data have been, and will continue to be, used to identify, implement, and assess the effectiveness of curricular reform efforts at the program level. The intended audience for this paper is faculty within a department who aim to develop a meaningful assessment system or an assessment coordinator who is directing development of an assessment system.

A learner-centered approach incentivizes discovering what is and is not working in the educational process so that it can be improved-deficiencies in performance must be detected and reported so they can be acted upon.

Method

To meet Singer-Freeman and Robinson's grand challenges, faculty should adhere to seven principles. The National Institute for Learning Outcomes Assessment (NILOA) describes the first five principles that, if used in a mission-relevant manner, result in effective, useful, and meaningful learning-centered assessment systems (NILOA, 2016).

1. *Assessment systems should be based on specific, actionable learning outcomes statements*. Learning outcomes should be written to convey curricular priorities and use active, operational verbs, which specify observable and measurable behaviors. Additionally, learning outcomes should be aligned across courses and scaffolded

throughout the curriculum (Biggs, 1996; Gaston, 2015). This requires a shift in faculty mindset away from "my class" to one of "our curriculum" (Jankowski & Marshall, 2017).

2. *Assessment systems should connect learning goals with actual student assignments and work*. Using faculty-designed, course-embedded assignments [authentic artifacts (NILOA, 2016)] as evidence of learning is preferable to "add-on" assessments such as exit exams or alumni surveys. Authentic artifacts work best for assessment because they are already integrated into courses and help students develop and demonstrate their learning as part of their normal classroom practices (Jankowski & Marshall, 2017). One way in which programs can address this principle is with the use of "signature assignments," defined by the Association of American Colleges and Universities as those assignments that require students to demonstrate and apply their proficiency in one or more learning outcomes (Roach & Alvey, 2021). They are often assignments that a program uses in multiple courses, sections, and/or across time (Roach & Alvey, 2021; Salt Lake City Community College, 2022; Stitt-Bergh, 2015). Thus, signature assignments are particularly appealing for use in assessment systems because they: a) can be collected across time, b) are embedded in courses, and c) have the potential to be collected and aggregated as evidence of student learning at the program level.

3. *Assessment systems should collaborate with relevant stakeholders, beginning with the faculty*. Faculty involvement is a critical component in creating an effective assessment system (NILOA, 2016; Welsh & Metcalf, 2003a; Welsh & Metcalf, 2003b). When faculty are deeply involved in the development of these processes, they will be more likely to be invested in using the assessment data to improve student learning. While assessment development must begin with faculty, additional voices should be included in the conversation through a process called "tuning," the process by which essential learning elements within a discipline are first defined by faculty and then refined through a collaborative and iterative process including other stakeholders (Jankowski & Marshall, 2017; Marshall, 2017; Marshall et al., 2017).

4. *Assessment systems should generate actionable evidence about student learning that key stakeholders can understand and use to improve student and institutional performance*. The connection between student learning data and the mechanisms for stakeholders to review and make sense of these findings and determine necessary actions should be clearly defined. They should also include standard processes for presenting results in transparent and understandable forms.

5. *Assessment systems should focus on improvement (and compliance will take care of itself)*. Ultimately, if an assessment system is designed to align with the previous four principles, and stakeholders are engaged and invested, the system is likely to meet the requirements of accrediting bodies.

Jankowski and Marshall (2017) add two more considerations to NILOA's five principles: (a) assessment systems should focus on *individual learners*, ensuring every student is learning (vs. an institutional focus where a sample of students are assessed); and (b) assessment systems should utilize *equitable practices* by considering the agency, positionality, and power of those creating the curriculum and assessment. Combined, these seven principles provide the foundation on which a meaningful and useful assessment program can be built. However, to attain the overarching goal of addressing Singer-Freeman and Robinson's grand-challenges, institutions must be attentive not only to the nuts-and-bolts found within these principles but also to the very complex ways in which the various parts of the assessment system work together as a whole. Institutions have unique stakeholders, stakeholder goals, levels of assessment (course, program, division, institution, etc.), and mechanisms by which those levels interact with one another. Even if a programmatic assessment system follows the principles outlined above, the system may be ineffective if it does not work within the larger institutional context.

A systems thinking approach is one way to reconcile this tension and is the approach used by the programs described in this paper (Bowers et al., 2022; Kim, 1999; Orgill et al., 2019; Stavrianeas et al., 2022). An assessment system built using a systems thinking approach will: (a) identify interactions between the assessment system and its institutional environment, (b) recognize the assessment system as one of many pieces within the institutional system, (c) examine the relationships between the assessment system and other systems within the institution, (d) examine how the assessment system changes over time, and (e) identify variables that cause changes to the assessment system. Assessment systems built with not only the seven principles of good assessment, but also with a systems thinking focus on how that assessment system will fit within the larger institutional context, are better positioned to address "big picture" questions such as Singer-Freeman and Robinson's grand-challenges, ultimately leading to meaningful curricular change that increases student success.

A backward design approach was used by faculty members to develop the program assessment system.

Designing the Program Assessment System

While programmatic assessment systems are a requirement at the institutional level, the development of this integrated and longitudinal system was a grassroots effort within the department largely influenced by our participation in a Council for Undergraduate Research Curriculum Transformations Grant (NSF 16-25354). Department faculty members discussed the commitment and efforts required to reform the curriculum and design a new learner-centered assessment system and all agreed to participate in weekly meetings, short homework assignments, and annual retreats with external consultants. Collegiate level buy-in was secured when the provost wrote a letter indicating that participation in this project would be considered high-level college service, providing additional incentive. Many department members also recognized the potential to publish curricular innovations that resulted from the project; several have capitalized on this opportunity (Bowers, 2020; Bowers et al., 2021; Chase et al., 2020; Mertz & Neiles, 2020; Mertz, et al., 2023; Neiles et al., 2019; Neiles & Bowers, 2020; Sherrer, 2020).

With buy-in secured, a backward design approach was used by faculty members to develop the program assessment system (Wiggins & McTighe, 2011). The first step was to identify what faculty stakeholders wanted our students to know and be able to do when they complete the programs. A skills curricular inventory was prepared in which all faculty were asked to identify skills beyond content knowledge that they wanted to see in our graduates. The department then worked as a team to discuss and ultimately group similar skills and values into several overarching skill categories. The skill categories were used as guides to fine-tune previous program learning outcomes (PLOs) and create new ones when necessary.

The specific skills and values associated with each skill category were incorporated into each PLO as sub-categories. For example, one PLO initially stated, "Upon completing the Chemistry/Biochemistry program, students will be able to effectively communicate and disseminate the results of the scientific process to a diverse audience." This outcome includes four types of communication skills as sub-categories that are assessed at different points in the curriculum: lab notebooks, oral communication, written communication, and poster presentations.

An important part of our process was allowing the PLOs to be dynamic rather than static, consistent with the tuning process described in the theoretical framework section. As the assessment system was designed, department faculty revisited and discussed the PLOs to create a shared understanding of their meaning. As a result of this fine tuning, some PLOs were removed while others were consolidated. The department also used survey input on students' understanding and perceived usefulness of the PLOs. In the Spring 2021 semester, a survey was administered to students enrolled in all department courses in accordance with approved St. Mary's College of Maryland (SMCM) Institutional Review Board (IRB) protocol (IRB SP21_36). The survey was used to evaluate students': (a) awareness of the six PLOs, (b) perceptions as to whether the PLOs were understandable, and (c) beliefs whether the PLOs were relevant to their field of study. The survey received 151 responses representing a wide range of chemistry/biochemistry courses and students' year in school. Student responses were reviewed to identify common themes to guide faculty in making the PLOs more easily understood by students. The PLOs were then edited based on these themes to result in a final set of six PLOs for both our Biochemistry and Chemistry programs found in Table 1.

Table 1
SMCM Chemistry and Biochemistry Programs' Learning Objectives

Program Learning Objective (PLO)
Upon completion of the chemistry/biochemistry major, students will be able to:
1. Correctly use the fundamental concepts of biochemistry/chemistry.
2. Solve problems by understanding and implementing the stages of the research process.
3. Effectively communicate and disseminate the results of the scientific process to a diverse audience.
4. Demonstrate the ability to identify, locate, and evaluate primary literature.
5. Work both individually and collaboratively with other students and faculty.
6. Exhibit the skills of a successful professional.

Scaffolding PLO Subcategories

After reaching consensus regarding a set of PLOs and sub-categories for the PLOs, each PLO was scaffolded throughout the curriculum using the principles of backward design (Mertz & Neiles, 2020) and assembled in a master scaffolding document. The department agreed on three learning levels: foundational, developing, and capstone. The foundational level introduces a skill or knowledge area for the first time whereas the developing level enhances or strengthens the opportunities for achieving a program learning outcome previously introduced. The capstone learning level is a masterly level and designates the expectation for the end of students' time in their major program. Individual faculty identified the PLO sub-categories they currently addressed in their course(s) and the learning level at which they taught the sub-category. The department then looked over the PLO coverage as a whole, discussing gaps in coverage, redundancy, or inconsistencies in learning level. For example, when scaffolding the communication PLO, the department identified a hole in the curriculum regarding scientific poster presentations but a redundancy in the number of oral presentations. Subsequently, in the Biochemistry I course (CHEM 420), an assignment presentation was changed from an oral to a poster format to address this finding (Sherrer, 2020). An example for the communication PLO is found in Table 2.

By developing the assessment system around signature assignments, the same artifacts are collected and archived each year from all sections of a course, regardless of instructor.

Signature Assignments

For the skill-based PLOs, signature assignments are used as evidence of learning in the assessment system. The signature assignments identified in each course feed into multiple PLO sub-categories and in almost all cases, into more than one skill-based PLO. The signature assignments within each course were generally pre-existing assignments (either as-is or with minor modifications). Armed with this information, the department identified the courses that would be used to assess each subcategory, the learning level for that subcategory in each course, and the signature assignment and scoring mechanism used to assess each sub-category within each course.

By developing the assessment system around signature assignments, the same artifacts are collected and archived each year from all sections of a course, regardless of instructor. A small cultural shift toward collecting and archiving the results for these assignments ensures our department has a rich database of assessment data for all skill-based PLOs in every academic year. While the content-based PLO (PLO 1) went through the same scaffolding process as the skill-based PLOs, signature assignments were not used to assess the content sub-categories. Course-based exams, either standardized or instructor created, are used to assess content learning objectives.

Table 2
Scaffolding Table for PLO 3 - Communication

Course	Lab Notebook	Oral	Written	Poster
General Chemistry II (CHEM 106)	***Signature Lab***		***Signature Lab***	
Organic Chemistry I (CHEM 311)	***Signature Lab***		***Signature Lab***	
Organic Chemistry II (CHEM 312)	***Signature Lab***		***Signature Lab***	
Inorganic Chemistry (CHEM 405)		***Signature Lab***	***Signature Lab***	
Biochemistry I (CHEM 420)	***Signature Lab***	***Bioinformatics Project***	***Bioinformatics Project*** **(Mertz and Streu 2015)**	***Bioinformatics Project*** **(Sherrer 2020)**
Biochemistry II (CHEM 425)		***Grant Project***	***Grant project***	
Physical Chemistry I (CHEM 451)	***Signature Lab***	***Presentation***		
Senior Capstone (CHEM 494)		***Presentation***	***Paper***	

Note: Learning levels depicted as light grey (foundational), medium grey (developing), and dark grey (capstone).

Table 3
Proficiency Scores in the Program Level Megalo-Rubric and Their Meaning.

Proficiency Score	0	1	2	3	4
Proficiency Level	Pre-foundational	Foundational	Developing	Capstone	Capstone+
Description	work is missing or does not meet a basic foundational standard	student exhibits basic understanding of a knowledge, skill, or competency	student exhibits strengthened or enhanced understanding of a knowledge, skill, or competency	student performs at the level of mastery expected of a graduating student	student's performance on a knowledge, skill, or competency exceeds the expectations of a graduating student

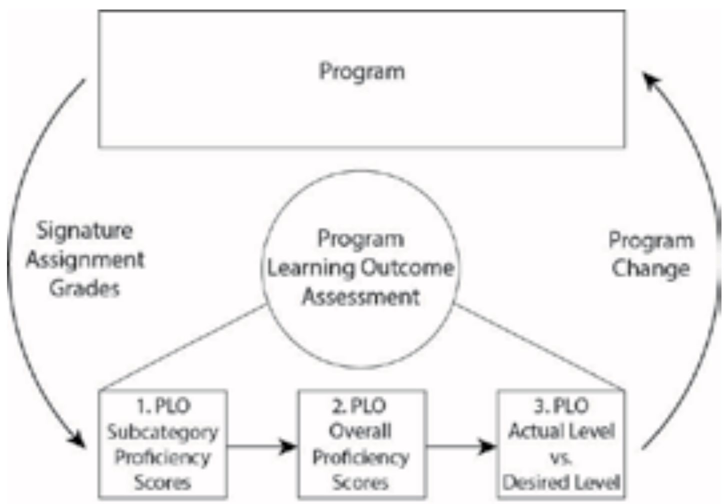
Developing the Megalo-Rubric

The megalo-rubric (named for both its size and as a reference to the local abundance of megalodon fossil teeth) was developed to map skill-based assessment data from each course's signature assignments directly onto the programmatic assessment system. The rubric has rows for each sub-category of the skill-based PLOs and columns that specify proficiency scores ranging from 0 (pre-foundational) to 4 (capstone+) (Table S1 in supplemental material). By mapping all of the signature assignment data onto one common set of proficiency scores within the megalo-rubric (Table 3), students or cohorts can be tracked longitudinally at both the PLO and sub-category levels, proficiency between specific groups of students can be compared, and easy-to-interpret information can be created that allows us to make data-driven choices about curricular reform.

Course Megalo-Rubric Mapping

The task of mapping each course onto the megalo-rubric was assigned to the faculty member(s) with primary responsibility for each course and is represented in Figure 2.

Figure 2
Flowchart for Mapping Course Assessment onto Program Assessment Using the Megalo-Rubric



Note: A separate feedback loop occurs for each PLO a course is assessing.

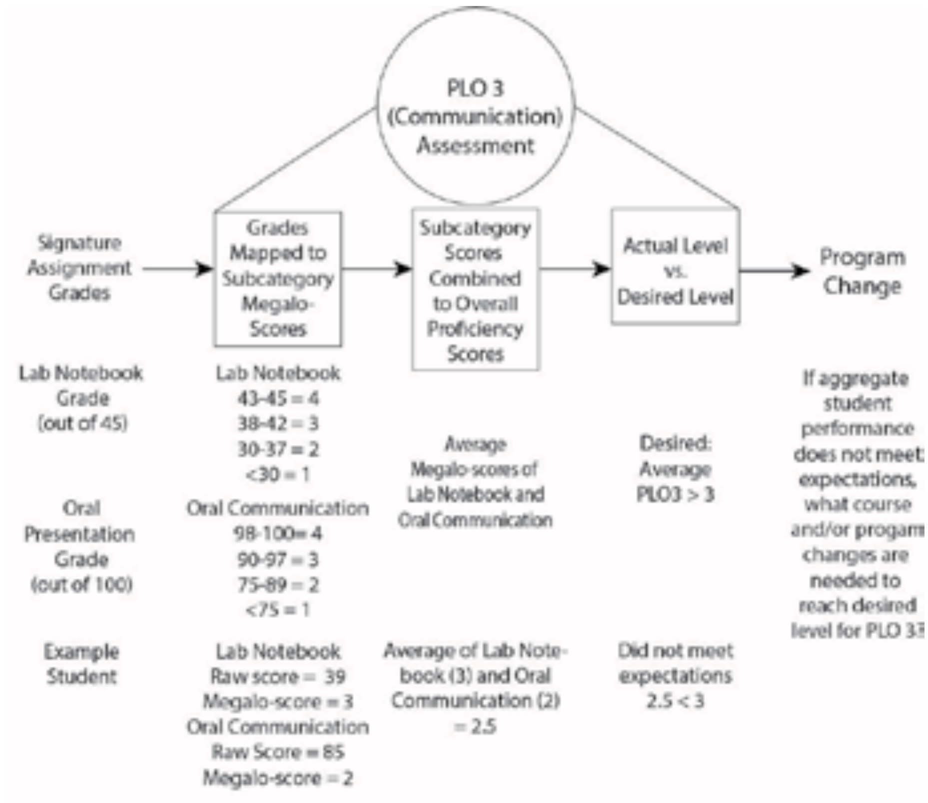
The megalo-rubric was developed to map skill-based assessment data from each course's signature assignments directly onto the programmatic assessment system.

After consulting the scaffolding summary documents and signature assignments, faculty specified how scoring structures in place within the course (e.g., grading process) would be used to assign a proficiency score for each PLO subcategory they are tasked to assess within the context of the learning level for their course (Box #1 in Figure 2). Faculty also determined how the individual subcategory proficiency scores would be used to generate an overall proficiency score for the overall PLO for each student (Box #2). The proficiency score necessary to meet the learning level expectations of the course was also identified–students with overall proficiency scores at or above the "meets expectations" value are considered to have met the expectations (Box #3). Often, the "met expectations" proficiency score increased for a specific skill as students moved from first-year to senior-year courses. For example, a proficiency score of 1 could be considered meeting expectations for a general chemistry course where a skill was first introduced (e.g., a foundational learning level) while a proficiency score of 3 for that skill may be expected in an upper-level course (e.g., a developing or capstone learning level). In this way, the program can evaluate longitudinal data (either by individual student or in aggregate) to determine whether students are successfully progressing through the megalo-rubric levels and ultimately, identify where programmatic changes are needed.

Figure 3 provides an illustrative example of how a Physical Chemistry I course uses scores to assess PLO 3 (communication). The figure provides broad information as to how the course assignments feed into the assessment program and an example student to illustrate these processes.

Figure 3

Flowchart Example for Using Signature Assignment Grades to Map Course Assessment onto Program Assessment for Physical Chemistry I, PLO 3.



The signature assignment in this course is a two-week, team-based bomb calorimetry experiment. Student laboratory notebooks and oral presentations are two of the grading artifacts of this signature assignment, which feed into the program assessment system. Both of these artifacts have an associated grading rubric used within the course. The figure illustrates how an example student's scores on each of these artifacts dictates their proficiency scores for lab notebooks and oral communication, and also how these combine to give an overall proficiency score for the communication PLO, which then determines whether the student has met programmatic expectations.

It should be noted that both years [of data collection] were impacted by the COVID-19 pandemic; the pandemic was an environmental factor that contextualized many of the subsequent data discussions.

Megalo-Rubric in Action

Data were collected using the megalo-rubric during the 2019-2020 and 2020-2021 academic years and in accordance with the protocol approved by the SMCM Institutional Review Board (IRB SU21_12, FA19_59). It should be noted that both years were impacted by the COVID-19 pandemic (one semester of the first year and both semesters of the second year); the pandemic was an environmental factor that contextualized many of the subsequent data discussions. Data were collected for all the skill PLOs (2-6) and their sub-categories using signature assignments identified in the scaffolding process (exception: CHEM 312 was not assessed during the Sp20 semester due to the pandemic).

In this section, data are reported for PLO 3, *"Upon completion of the chemistry/biochemistry major, students will be able to effectively communicate within a research team and disseminate the results of research to a diverse audience,"* with a focus on the sub-category Written Communication. The use of a single PLO and subcategory serves as a relatively straightforward example of the data generated and its potential use. The data for PLO 3 can be aggregated and dis-aggregated in many forms depending on the questions faculty and/

or other stakeholders are asking. Data were collected for PLO 3 in nine courses and from 461 students across the two academic years (courses and course numbers are listed in Table 2). Some data points may come from the same student who is being assessed in more than one course and potentially at more than one learning level.

The following section provides a wide range of examples of how megalo-rubric data were used to "close the loop" and identify potential action items by collaboratively interpreting assessment data to inform curricular reform. Curricular actions are dictated by the environment of the institution and its programs; thus, the reader's program may choose very different actions given the same data. Each comparison contains a "making the graph" section that describes what student learning data were compiled to create the figure, a "questions answered" section that describes the type of inquiry that this plot allows, and a "closing the loop conversation" that describes the subsequent discussion.

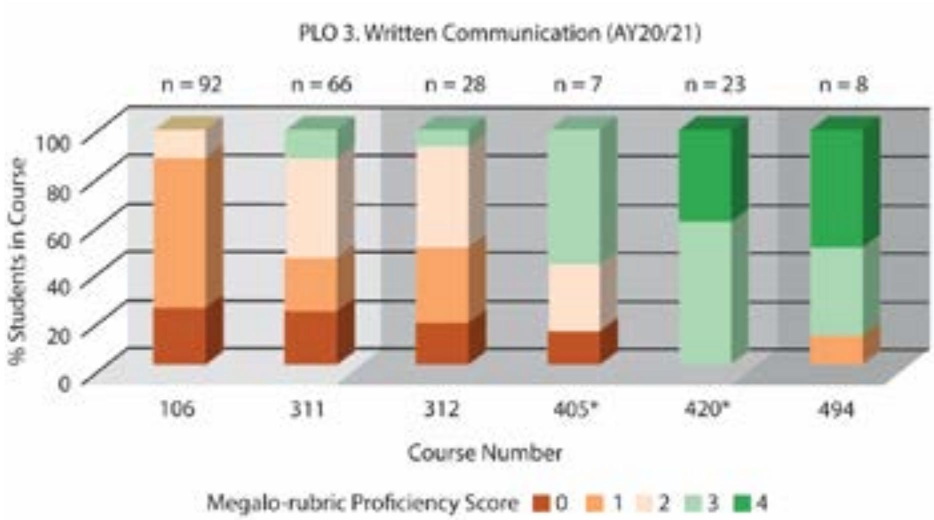
Results

Comparison 1: Between Courses on Written Communication in Academic Year 20/21

Curricular actions are dictated by the environment of the institution and its programs; thus, the reader's program may choose very different actions given the same data.

Making The Graph. Figure 4 compares PLO 3 sub-category Written Communication scores between courses within the chemistry and biochemistry programs for one academic year (AY20/21). Each column represents the Written Communication proficiency score distribution for one course. The highlighted sections behind the columns indicate the learning level of those courses. Not all courses within the programs are used in the analysis of each sub-category; thus, the graph does not include all courses within the chemistry and biochemistry programs.

Figure 4
Comparison of Student Proficiency Scores Between Courses on the PLO 3 Written Communication Sub-Category.



Note: Highlighted boxes behind the columns represent learning levels depicted as light gray (introducing), medium gray (developing), and dark gray(capstone). CHEM 405 (Inorganic) and 420 (Biochemistry) are only required for one major, chemistry and biochemistry, respectively.

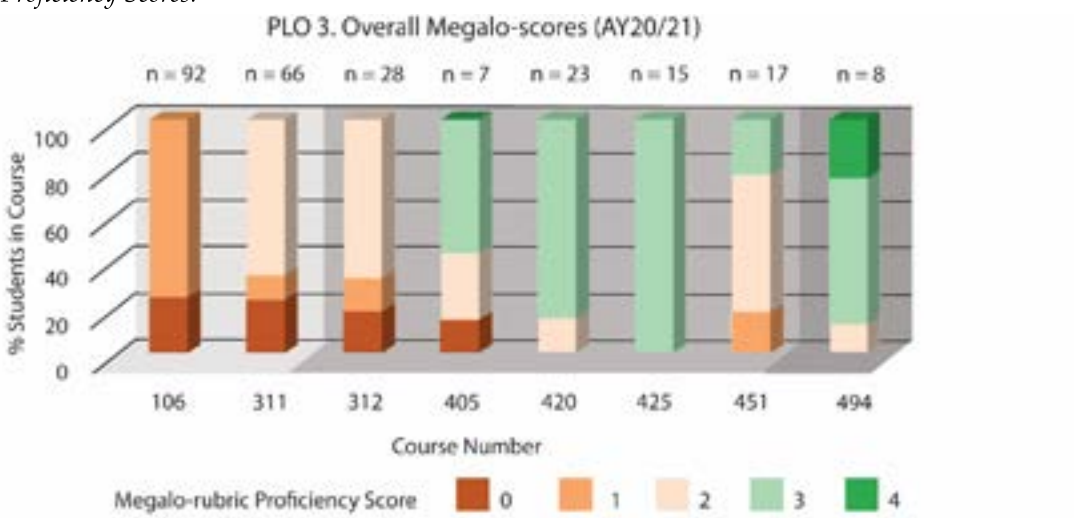
Questions Answered. This graph addresses programmatic questions as to whether courses are generally well-scaffolded with regards to Written Communication skills. For example, it can answer whether our courses are providing students the opportunity to progress from foundational performance (0 or 1) to capstone performance (3 or 4) on these skills as they move through the curriculum.

Closing the Loop Conversation. When considering this graph, the department saw that there was indeed a systematic improvement in Written Communication proficiency scores as one progresses from introductory courses to upper-division courses in both the chemistry and biochemistry programs. The department spoke specifically about the progression from each course to the next and whether there were any areas of concern taking into consideration both the ns of each course as well as contextual information about the course. In this case, no areas of concern were found and it was determined that no curricular changes were necessary.

Comparison 2: Scores Between Courses on Overall Communication PLO 3 in Academic Year 20/21

Making the Graph. Figure 5 compares overall PLO 3 scores between courses within the chemistry and biochemistry programs for one academic year. As described earlier, the overall PLO proficiency score is determined by combining one or more sub-scores, as dictated by the course faculty-of-record. For a detailed example of how sub-categories can be "rolled up" into an overall score, see Figure 3. As with the last plot, these data do not represent a single group of students moving through the courses but instead separate courses (and in most cases separate students) within a single academic year. Each column represents one course and shows the breakdown of megalo-rubric scores for PLO 3.

Figure 5
Comparison of Student Proficiency Scores Between Courses on the Overall PLO 3 Megalo-Rubric Proficiency Scores.



Note: Highlighted boxes behind the columns represent learning levels depicted as light gray (introducing), medium gray (developing), and dark gray (capstone).

Questions Answered. This graph can be used to answer programmatic questions as to whether courses are generally well-scaffolded with regards to overall communication skills. Additionally, this graph can provide information as to how the sub-categories are impacting the overall communication PLO 3 megalo-rubric scores by comparing it to the sub-category graphs (such as Figure 4). For example, when comparing the overall graph to individual sub-category graphs, stakeholders can determine whether any single sub-category is over-represented when "rolled up" into the overall PLO 3 scores.

The department saw that there was indeed a systematic improvement in Written Communication proficiency scores as one progresses from introductory courses to upper-division courses in both the chemistry and biochemistry programs.

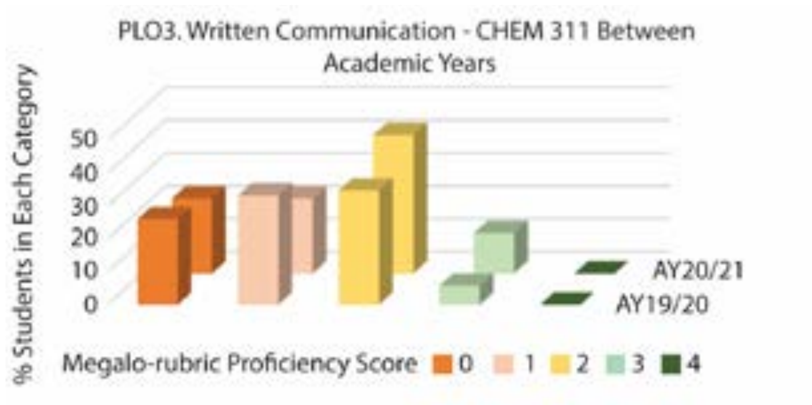
Closing the Loop Conversation. When looking at these data, the department noted a mostly systematic improvement in overall communication proficiency scores as one progresses from introductory courses to upper-division courses with one notable exception, CHEM 451 (Physical Chemistry I). Upon further investigation it was found the students in CHEM 451 were earning scores at foundational (or even pre-foundational) levels on the laboratory notebook sub-category, thus bringing their overall communication scores down. It is possible that additional instruction is needed in the introductory sequence to better prepare students for expectations in later courses or that laboratory notebook expectations need to be standardized across the curriculum. Conversations around these findings led to departmental members working together to identify a common lab notebook format that students now encounter in all laboratory courses (though some discipline-specific differences still exist in how notebooks are assessed).

Comparison 3: Scores Between Two Academic Years (AY19/20 and 20/21) on Written Communication Within a Single Course

Making the Graph. Figure 6 compares Written Communication scores between two academic years (AY19/20 and 20/21) within a single course (CHEM 311 Organic Chemistry I). Each column represents the percentage of students that received each proficiency score in a given academic year. The front row represents AY19/20 and the back row represents AY20/21.

Figure 6

Comparison of Proficiency Scores from the Megalo-Rubric Between two Academic Years (Ay19/20 and 20/21) on the Written Communication Sub-Category.



Questions Answered. This graph can be used to determine whether proficiency scores are consistent from year to year. It can also be useful in determining whether a specific intervention or other variable at the course level, such as a change in instructor, affected this sub-category. Additionally, this type of comparison could be done at the program level to compare scores between years for all program courses.

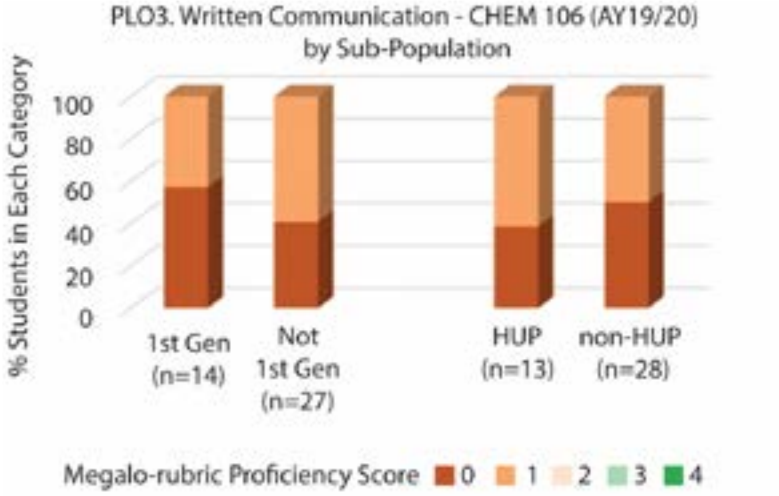
Conversations around these findings led to departmental members working together to identify a common lab notebook format that students now encounter in all laboratory courses.

Closing the Loop Conversation. In these data, we see that students are generally performing between 0 (pre-foundational) and 2 (developing) on their proficiency scores with a few students performing at the capstone level (3). This aligns with the course's learning level for the written communication sub-category, which was set at developing during the scaffolding process. Additionally, a slight shift toward higher scores in the second year is observed. Faculty discussed these data in comparison with similar data for general chemistry and identified a need for better alignment between the two introductory sequences in terms of the writing instruction. Once this new writing instruction has been implemented, these data will be compared to those collected after the curricular reform to determine whether it has been effective.

Comparison 4: Comparison of Sub-Population Performance on Written Communication Within a Course for AY19/20

Making the Graph. One of the four grand challenges of assessment identified by Singer-Freeman and Robinson (2020a, 2020b) was to use assessment findings to increase equity. To do this, assessment data need to be disaggregated by student sub-population so that differences in group performance can be identified. Figure 7 compares Written Communication scores between four different sub-populations [first-generation, non-first-generation, historically underrepresented populations (HUP), and non-HUP] during one semester (Sp20) within a single course (CHEM 106 General Chemistry II). At SMCM, HUP students are defined as Black and Hispanic students while non-HUP students are all other students. Each colored block within the column represents the percentage of students who received each megalo-score.

Figure 7
Comparison of Sub-Populations on the Written Communication Sub-Category within General Chemistry II Course During AY19/20..



Questions Answered. This graph can answer questions as to how different sub-populations are performing within our courses or even across the entirety of the program. This type of disaggregated data analysis allows us to investigate whether our students have equitable opportunities for success (Bensimon & Malcom, 2012).

Closing the Loop Conversation. The graph shows that our non-first-generation students may be outperforming our first-generation students given that more non-first-generation students received a one vs. a zero proficiency score for Written Communication. On the other hand, our HUP students may be outperforming our non-HUP students. These differences may or may not be statistically significant; however, they do warrant further investigation to determine whether teaching practices can be improved to increase student success in the under-performing groups.

Discussion

The program assessment system described in this paper addresses the four grand challenges described by Singer-Freeman and Robinson (2020b): (a) use assessment findings to increase equity, (b) use assessment findings to direct immediate pedagogical improvements, (c) produce visible and actionable assessment findings that drive innovation, and (d) examine changes in institutional effectiveness (including student learning) over time. The system enables longitudinal tracking of individual students as well as dis-aggregation of assessment data to track student groups such as first-generation college students or historically under-represented populations.

Program faculty meet annually to review and discuss assessment results, and based on those results, modify the assessment system, the PLO scaffolding, or individual program courses to help students meet program objectives. Perhaps most importantly, all program faculty understand that the assessment system is a tool for evaluating, reflecting upon, and improving the learning experience for our students.

It is important to note that while content-related subcategories for the content PLO have been identified, program faculty are currently still in the late stages of scaffolding the content-related learning outcome. In the same process described for the previous PLOs above, faculty are currently engaged in conversations regarding where content sub-categories are covered in each program's courses and at what level they are taught (foundational, developing, or capstone). The general system described here for reflecting upon and taking coherent, coordinated action to address student performance for the skill PLOs is anticipated to be just as applicable once the content PLO has been fully scaffolded. Also, while the system presented here does not involve co-curricular activities (such as student clubs, speaker series, etc.), creating PLOs that involve

Disaggregated data analysis allows us to investigate whether our students have equitable opportunities for success.

these activities could easily mirror the process described here. Essentially, these co-curricular activities could be included as scaffolding points just as course-based activities are used in the system described here. The major difference, of course, is that these activities are likely not *required* of students, so it would be difficult to ensure that all students have exposure to the necessary scaffolding.

Data collected as part of the assessment system drive evidence-based innovation when coupled with reflection and meaningful faculty conversations centered around student learning.

In summary, the assessment system described here adheres to both NILOA's principles for effective assessment systems and addresses the grand challenges posed by Singer-Freeman and Robinson. It also serves as a vehicle for faculty to collaborate amongst themselves, across programs, and with students and other stakeholders. The robustness and flexibility of the system ensures that it is meaningful for faculty and will continue to be so as it evolves into the future. Most importantly, this system shifted assessment within the chemistry and biochemistry programs at SMCM from a compliance requirement to a learner-centered, innovation-focused vehicle for reflective collaborations that improve student learning.

Future Directions and Conclusion

Now that this assessment system is in place, the data created by the system allows program faculty to determine how specific interventions or curricular changes are affecting different groups across the curriculum, within specific courses, and over time, allowing faculty to make programmatic or curricular changes that directly address equity. Assessment findings can also be used to direct immediate pedagogical improvements, such as the writing instruction modules being implemented in General Chemistry as a result of Comparison 3.

Data collected as part of the assessment system drive evidence-based innovation when coupled with reflection and meaningful faculty conversations centered around student learning. Faculty within the program have used student learning data to generate and assess pedagogical change in specific courses. For example, a scholarly laboratory model was implemented at St. Mary's College of Maryland to better address several PLOs and Course Learning Outcomes in the participating courses. An assessment of that particular innovation was published in 2021, including a discussion of modifications needed to the scholarly model based on assessment findings (Bowers et al., 2021). Similarly, in 2023, a manuscript was published by multiple members of the department on use the of a rubric to measure students' collaboration skills (one of the program PLOs) (Mertz et al., 2023). Finally, signature assignment data are archived, permitting us to examine student performance with respect to any PLO or PLO sub-category over time within courses or of a particular student or student cohort over time. The evidence-focused nature of this system creates ample opportunities for publishable research on student learning at both the course and program levels.

Acknowledgements

The authors gratefully acknowledge funding from the Council for Undergraduate Research Curriculum Transformation Project (NSF 16-25354) and the helpful input of our two consultants, Kimberley Frederick (Skidmore College) and Bridget Gourley (DePauw University). We also acknowledge support from the Office of the Provost at SMCM to participate in the CUR-CTP and annual participant meetings. This manuscript was written by Geoffrey M. Bowers, Pamela S. Mertz, and Kelly Y. Neiles, who made equal contributions to the content, writing, and data analysis. All authors participated in the assessment system development process and contributed assessment data used in this paper.

Funding details: Council for Undergraduate Research Curriculum Transformation Project (NSF 16-25354)

Data availability statement: The data that support the findings of this study are available from the corresponding author, KYN, upon reasonable request.

Competing interests: There are no relevant financial or non-financial competing interests to report.

References

Bensimon, E. M., & Malcom., L. (2012). *Confronting equity issues on campus: Implementing the equity scorecard in theory and practice*. Stylus Publishing. https://doi.org/10.4324/9781003443582

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364.

Bowers, G. M. (2020). Enhancing career readiness by integrating alumni interactions into upper-level chemistry curricula. In K. Y. Neiles, P. S. Mertz, & J. Fair (Eds.), *Integrating professional skills into undergraduate chemistry curricula* (pp. 179–194). ACS Publications. https://doi.org/doi:10.1021/bk-2020-1365.ch010

Bowers, G. M., Mertz, P. S., & Neiles, K. Y. (2022). *A systems thinking approach toward meaningful program assessment*. ChemRxiv. https://doi.org/doi:10.26434/chemrxiv-2022-wb4mr

Bowers, G. M., Larsen III, R. K., & Neiles, K. Y. (2021). Scholarship-based undergraduate laboratory courses modeled on a graduate school research rotation. *Journal of Chemical Education 98*(4), 1152–1162. https://doi.org/doi:10.1021/acs.jchemed.1c00009

Chase, D. T., Neiles, K. Y., & Koch, A. S. (2020). Intentionally scaffolding career skills in an organic chemistry laboratory course. In K. Y. Neiles, P. S. Mertz, & J. Fair (Eds.), *Integrating professional skills into undergraduate chemistry curricula* (pp. 147–164). ACS Publications. https://doi.org/doi:10.1021/bk-2020-1365.ch008

Ewell, P. T. (2009). *Assessment, accountability, and improvement*. National Institute for Learning Outcomes Assessment. https://www.learningoutcomesassessment.org/wp-content/uploads/2019/02/OccasionalPaper1.pdf

Ewell, P., Paulson, K., & Kinzie, J. (2015). *Down and in: Assessment practices at the program level*. National Institute for Learning Outcomes Assessment. https://files.eric.ed.gov/fulltext/ED535126.pdf

Gaston, P. L. (2015). General education transformed: How we can, why we must. *Association of American Colleges and Universities*.

Gould, M. (2010). GIScience grand challenges: How can research and technology in this field address big-picture problems. *ArcUser 13*(4), 64–65.

Jankowski, N. A., & Marshall, D. W. (2017). *Degrees that matter: Moving higher education to a learning systems paradigm*. Stylus Publishing. https://doi.org/doi:10.4324/9781003444015

Kim, D. H. (1999). *Introduction to systems thinking*. Pegasus Communications, Inc.

Kuh, G. D., & Hutchings, P. (2014). *Assessment and initiative fatigue. In using evidence of student learning to improve higher education*. John Wiley & Sons.

Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Ewell, P. T., Hutchings, P., & Kinzie, J. (2015). *Using evidence of student learning to improve higher education*. John Wiley & Sons. https://doi.org/doi:10.1177/1521025116657835

Marshall, D. W. (2017). *Tuning: A guide for creating discipline-specific frameworks to foster meaningful change*. National Institute for Learning Outcomes Assessment. https://eric.ed.gov/?id=ED574503

Marshall, D. W., Jankowski, N. A., & Vaughan, T. (2017). *Tuning impact study: Developing faculty consensus to strengthen student learning*. National Institute for Learning Outcomes Assessment. https://eric.ed.gov/?id=ED590524

Mertz, P. S., Sherrer, S. M., & Bowers, G. M. (2023). Teaching and assessing undergraduate collaboration skills scaffolded through the biochemistry curriculum using collaboration rubrics and student learning contracts. *Biochemistry and Molecular Biology Education*, *51*(5), 499–507. https://doi.org/10.1002/bmb.21760

Mertz, P. S., & Neiles, K. Y. (2020). Scaffolding career skills into the undergraduate curriculum utilizing a backward design approach. In K. Y. Neiles, P. S. Mertz, & J. Fair (Eds.), *Integrating professional skills into undergraduate chemistry curricula* (pp. 43-55). ACS Publications. https://doi.org/doi:10.1021/bk-2020-1365.ch004

Neiles, K. Y., & Bowers, R. A. (2020). A general chemistry cocurriculum focused on the development of professional and academic skills. In K. Y. Neiles, P. S. Mertz, & J. Fair (Eds.), *Integrating professional skills into undergraduate chemistry curricula* (pp. 105–146). ACS Publications. https://doi.org/doi:10.1021/bk-2020-1365.ch007

Neiles, K. Y., Bowers, G. M., Chase, D. T., VerMeulen, A., Hovland, D. E., Bresslour-Rashap, E., Eller, L., & Koch, A. S. (2019). Teaching collaborations and scientific practices through a vertically scaffolded biodiesel laboratory experience. *Journal of Chemical Education*, *96*(9), 1988–1997. https://doi.org/doi:10.1021/acs.jchemed.9b00008

NILOA. (2016). Higher education quality: Why documenting learning matters [A policy statement]. *National Institute for Learning Outcomes Assessment*. https://files.eric.ed.gov/fulltext/ED567116.pdf

Orgill, M., York, S., & MacKellar, J. (2019). Introduction to systems thinking for the chemistry education community. *Journal of Chemical Education*, *96*(12), 2720–2729. https://doi.org/doi:10.1021/acs.jchemed.9b00169.

Roach, S., & Alvey, J. (2021, February 4). Fostering integrative learning and reflection through "signature assignments". *AAC&U Liberal Education Magazine*. https://www.aacu.org/liberaleducation/articles/fostering-integrative-learning-and-reflection-through-signature-assignments

Salt Lake City Community College. (2022, August 30). *Signature assignments | SLCC*. https://www.slcc.edu/eportfolio/signature-assignments.aspx

Sherrer, S. M. (2020). Using scientific poster presentations to scaffold professional communication skill experiences into biochemistry courses. In K. Y. Neiles, P. S. Mertz, & J. Fair (Eds.), *Integrating professional skills into undergraduate chemistry curricula* (pp. 165–178). https://doi.org/doi:10.1021/bk-2020-1365.ch009

Singer-Freeman, K., & Robinson, C. (2020a). Grand challenges in assessment: Collective issues in need of solutions. *National Institute for Learning Outcomes Assessment*. https://files.eric.ed.gov/fulltext/ED612032.pdf

Singer-Freeman, K. E., & Robinson, C. (2020b). Grand challenges for assessment in higher education. *Journal of Research and Practice in Assessment*, *15*(2), 1–20. https://files.eric.ed.gov/fulltext/EJ1293386.pdf

Stavrianeas, S., Bangera, G., Bronson, C., Byers, S., Davis, W., DeMarais, A., Fitzhugh, G., Linder, N., Liston, C., McFarland, J., Otto, J., Pape-Lindstrom, P., Pollock, C., Reiness, C. G., & Offerdahl, E. G. (2022). Empowering faculty to initiate STEM education transformation: Efficacy of a systems thinking approach. *PLOS ONE 17*(7), e0271123. https://doi.org/10.1371/journal.pone.0271123

Stephan, M. L., Chval, K. B., Wanko, J. J., Civil, M., Fish, M. C., Herbel-Eisenmann, B., Konold, C., & Wilkerson, T. L. (2015). Grand challenges and opportunities in mathematics education research. *Journal for Research in Mathematics Education*, *46*(2), 134–146. https://www.nctm.org/Publications/journal-for-research-in-mathematics-education/2015/Vol46/Issue2/Grand-Challenges-and-Opportunities-in-Mathematics-Education-Research/

Stitt-Bergh, M. (2015). *Collect evidence of student learning using a signature assignment, a workshop on program learning outcomes assessment*. Retrieved March 3, 2024. https://drive.google.com/file/d/1tcsbS5yy7W-AGHwlJ1pbNWV4PjDTy431/view

Welsh, J. F., & Metcalf, J. (2003a). Cultivating faculty support for institutional effectiveness activities: Benchmarking best practices. *Assessment & Evaluation in Higher Education*, *28*(1), 33–45. https://doi.org/doi:10.1080/02602930301682

Welsh, J. F., & Metcalf, J. (2003b). Faculty and administrative support for institutional effectiveness activities: A bridge across the chasm? *The Journal of Higher Education*, *74*(4), 445–468. https://doi.org/doi:10.1353/jhe.2003.0032

Wiggins, G. P., & McTighe, J. (2011). *The understanding by design guide to creating high-quality units*. ASCD.