

Association for the Assessment
of Learning in Higher Education

AALHE

PROMOTING ASSESSMENT FOR LEARNING

INTERSECTION

A journal at the intersection of assessment and learning

Winter
2020

Contents

President’s Letter.....	2
Note from the Editor-in-Chief	3
Reliability and Validity 101: A Primer for Student Affairs Assessment.....	5
Sustain the Gain: Maintaining an Effective Learning Outcomes and Assessment System.....	11
Assessing Academic Support Centers: A Framework for Comprehensive Assessment.....	17
Defining Early Indicators of Licensure Exam Success.....	22
What Does It Mean to Commit to Excellence? Reflections of an Assessment Professional Serving as a Baldrige	29
Implementing an Effective Meta-Assessment System.....	34

President's Letter

By Jane Marie Souza



Dear Reader, your input is requested.

In 2014, founding editor of *Intersection*, David Eubanks, included in the Fall Edition a question and answer session with then Senior Vice President/Chief of Staff of the Southern Association of College and Schools Commission on Colleges. The interview marked the first time a conversation with an accreditor was included in *Intersection*. Since I have been a long-time advocate for promoting understanding of the accrediting agency perspective, when I took over as editor-in-chief I continued the interviews. The series was sustained until we had published a series of *Conversations with Accreditors* from across all the regional accrediting agencies as well as the Distance Education Accrediting Commission.

In 2016, I started presenting a panel session at the annual AALHE conference during which vice presidents from various agencies would respond to questions about the accreditation process. These live sessions were designed to provide an additional way of promoting the view of accreditors as partners with institutions in the quest for quality in education. This year at the 10th annual AALHE conference, we will mark the fifth annual *Conversations with Accreditors* panel session. I am reaching out to you now to request your input.

In late March, fellow board member Steven Hawks and I will be meeting with the 2020 accreditor panelists. We will be planning the prompts for the initial one-hour formal panel session that will be followed by an informal open question/answer session. **I would love to hear from you, prior to March 26th if possible, with questions you would like to be included in the questions posed to the accrediting agency representatives.** While we cannot promise all questions will be covered, we will do our best to have submissions inform the conversation. Of course, it would be great to have you join us and ask your question in person at the conference in New Orleans. However, if you would like to help us prepare in advance with specific prompts, that would be wonderful. **Please email your questions to me at JSOUZA@AALHE.ORG.**

Steven Hawks and I will plan to capture the questions and responses from the Q&A session and publish that information in the 2020 Conference Proceedings.

Thank you for considering this request. I hope to see you in June.

Jane Marie

Jane Marie Souza, PhD

AALHE President, 2019-2020

Associate Provost for Academic Administration, University of Rochester

Note from The Editor-In-Chief

By *Kathleen Gorski*

Welcome to the 2020 winter edition of AALHE's quarterly publication.

This year, we decided to devote two editions to an open call for articles on current topics related to the assessment of student learning. We were especially interested in submissions aimed at educating and supporting faculty, assessment practitioners, faculty development professionals, and other stakeholders in higher education.

We received many quality submissions on varying topics including maintaining effective systems, using frameworks for assessment, defining licensure exam success, and a reliability and validity for student affairs assessment. We hope you enjoy our open topic issue as much as we have enjoyed putting it together.

We would like to thank all our authors in this edition. Their work continues to inform and inspire. We hope you will consider submitting an article for a future edition to contribute to scholarship in the assessment of student learning. Please visit <https://www.aalhe.org/page/Intersection> for dates and guidelines.

Editorial Board

Giovanna Badia, McGill University

Jeff Barbee, Indiana University

Lisa Bonneau, University of South Dakota

Rebecca Gibbons, City Colleges of Chicago

Kathleen Gorski, Waubensee Community College

Jihee Hwang, University of Oklahoma

George Klemic, Lewis University

Shannon Milligan, University of California San Diego

Karla Pérez-Vélez, Colorado State University

Kate Pinder, Western Governors University

Amy Topper, University of Rhode Island

Shauna Wilton, University of Alberta

Alison Witherspoon, American College of Education

Sarah Wu, Georgia Institute of Technology



Intersection Update

We are in the middle of an exciting academic year for our publication. The Intersection committee has actively been working towards becoming a peer reviewed journal. This summer we applied for an International Standard Serial Number (ISSN) to officially identify Intersection as a journal. We received official approval this fall. We also spent time this fall and winter on guideline revisions to align our publication with requirements of Education Resources Information Center (ERIC) in order to begin submitting articles to get indexed in the online library sponsored by the Institution of Education Sciences (IES) of the U. S. Department of Education. Although we currently practice blind review, we are in the process of setting up structures for peer review by content specialists. We will be reaching out to our members to apply. Our goal is to officially launch the new guidelines at the 10th Annual Conference in June and begin officially using the guidelines and submitting publications to ERIC in fall 2021.

Emerging Dialogues Update

AALHE's Emerging Dialogues publication is intended to be a platform for original or responsive articles focused on learning outcomes assessment that are relevant and thought-provoking while also being well reasoned and supported by research. Emerging Dialogues articles are 1200 words or less and designed to share and discuss. Last year, the Emerging Dialogues committee revised guidelines and restructured the website to provide alternative channels of communication through email, and social media platforms. Each article can easily be shared on LinkedIn, Twitter, Facebook and Outlook. Readers have reached out to let us know that Emerging Dialogues articles are also great to include in monthly assessment newsletters. We encourage you to start or join an Emerging Dialogue Please visit <https://www.aalhe.org/emerging-dialogues>.

AALHE Conference Session

To celebrate AALHE publications, including Intersection and Emerging Dialogues, we will be sharing our new guidelines to all members and offering an AALHE Board session titled “Getting Published: Writing about Your Experience in Assessment in Higher Education” at our 10th annual conference in New Orleans this June. In this session for beginners, presenters will share benefits of writing about assessment for professional publication. Strategies for engaging in professional discourse and getting published will be discussed. Time will be provided to generate writing topics with a partner using guided questions. The ultimate goal of the session is for all participants to leave with a potential article topic.

Dr. Kathleen Gorski is Dean for Learning Outcomes, Curriculum and Program Development at Waubonsee Community College in Sugar Grove, Illinois. She can be reached at kgorski@waubonsee.edu.

Disclaimer: *The views and opinions expressed in the articles in this publication reflect those of the authors and not necessarily those of the Association for the Assessment of Learning in Higher Education.*

Reliability and Validity 101: A Primer for Student Affairs Assessment

By Madison Holzman, Andrea M. Pope and Jeanne Horst

Abstract: Increasingly, student affairs professionals are asked to conduct assessment as part of their jobs. That is, student affairs professionals are expected to collect assessment data to support that their programming contributes to student development and learning. To make claims about students' development or learning, we must ensure that assessment results are meaningful and reflect what we sought to measure. This question about what assessment results mean is a question of validity and in effect means we must collect validity evidence to support the interpretations we want to make from assessment results. Reliability, or consistency of student scores, is often considered a necessary, but insufficient, piece of validity evidence. As such, we often begin addressing questions of validity by evaluating reliability. Although there are numerous excellent assessment resources for student affairs professionals (e.g., Henning & Roberts, 2016; Wise & Davenport, 2019), didactic pieces that explain the importance of reliability and validity in a student affairs assessment context are sparse. In this article, the concepts of reliability and validity are explained and illustrated using examples relevant to student affairs. After reading this article, it is our hope that student affairs professionals have a basic understanding of reliability and validity and their importance, as well as are introduced to additional resources if interested in learning more about these key assessment concepts.

Introduction

Imagine you are responsible for assessing a major program in your campus' conduct office. Over the past several years, you have heard reports of increased use of alcohol by students on campus. Before modifying the existing alcohol intervention program, the program directors want to gather evidence of students' alcohol use. To this end, you find a promising measure, the *hypothetical* Alcohol Consumption Scale (ACS), and administer it to all students on campus. Surprisingly, despite significant increases in alcohol-related hospitalizations and sanctions, you find that students report low levels of alcohol consumption. How should you interpret the ACS results, given this contradictory evidence?

Before interpreting the results, we need to know that the ACS accurately measures students' alcohol consumption. That is, we need to know if the results and our interpretations of those results are reliable and valid. Reliable and valid assessment data help us determine whether students learn or develop in the ways we hope they will. With trustworthy data, we are equipped to use results for programmatic changes and effectively communicate the value of our programs to external stakeholders. However, to use assessment results for these purposes, we need evidence that the data accurately reflect students' knowledge, attitudes, and behaviors (i.e. evidence of reliability and validity).

There are excellent assessment resources for student affairs professionals that explain the purpose of assessment and provide guidance on implementing and conducting successful assessment processes (e.g. Henning & Roberts, 2016; Wise & Davenport, 2019). However, didactic pieces that explain the importance of reliability and validity in a student affairs assessment context are sparse. The purpose of this article is to briefly present the concepts of reliability and validity in the context of student affairs assessment. Although the above example related to the assessment of alcohol use is carried throughout, the concepts apply generally.

The Big Picture

When we consider whether the ACS results—or any assessment results—represent what we claim they represent, we are posing a validity question. Validity relates to the notion of “truth” in the sense that we want to know how closely assessment results reflect reality and the extent to which the conclusions drawn from assessment results are accurate. In the case of our hypothetical conduct office, we want to know “Do low average scores on the ACS indicate low levels of drinking?”

Answering validity questions is not easy. It takes substantial evidence to confidently say that conclusions drawn from assessment results are trustworthy. We often begin, however, by providing evidence of reliability. In the section that follows, reliability is defined and discussed as a starting point for gathering validity evidence.

Reliability

In short, reliability is an index of consistency (Traub & Rowley, 1991). Reliability can be consistency of raters' scores to student essays, referred to as inter-rater reliability. Reliability can be consistency of students' responses on the same measure across time, referred to as test-retest reliability. Or, reliability can be consistency of students' responses across assessment questions on the same measure, referred to as internal consistency reliability. For this paper, we primarily consider the third form of reliability—internal consistency reliability.

To illustrate the general concept of consistency, consider a non-student affairs example. Imagine you are shopping for a bathroom scale. At your favorite department store, you test several scales. The first scale reads 175 pounds. The second scale reads 174 pounds. The third scale reads 175 pounds. The fourth scale reads 159 pounds. Given your weight did not change while testing the scales, and the first three scales report similar weights, you might say the measurements from the first three scales are consistent. You might also conclude that the fourth scale does not function properly, as it does not yield measurements consistent with the other scales. This concept of consistency may be applied in assessment.

Suppose the ACS has 15 questions to which students respond on a scale of 1 (*strongly disagree*) to 5 (*strongly agree*), and those questions are thought to measure students' alcohol use. Given that students' alcohol use will not change while taking the assessment, we expect individual students' responses to be similar across the questions. That is, if students have high levels of alcohol use, they should primarily respond on the upper end of the response scale for each question. If students have low levels of alcohol use, they should primarily respond on the lower end of the response scale for each question.

Another way to illustrate internal consistency reliability is to consider the rank-ordering of students across questions on an instrument. On the ACS, we expect students with high levels of alcohol use to consistently rank higher than students with moderate or low levels of alcohol use. Consider an example with three students with different levels of alcohol use—high, moderate, and low. Their responses to three ACS assessment questions are provided in Figure 1.













Question 1 "I miss class due to drinking or hangover."	Question 2 "I prioritize alcohol over other responsibilities"	Question 3 "I use marijuana regularly."	Scale
			5 Strongly Agree
			4 Agree
			3 Neither agree nor disagree
			2 Disagree
			1 Strongly Disagree
Key  = student with high alcohol use  = student with moderate alcohol use  = student with low alcohol use			

Figure 1. Student responses to three ACS questions

Students are rank-ordered in the same way for questions one and two. However, they are not rank-ordered the same for question three. Given all three items are purported to measure alcohol use, the next logical question is *why* the students' rank-ordering differs for question three compared to questions one and two. When we examine the questions, we see that question three may measure drug use, not alcohol use. Although drug and alcohol use may be related, drug use is a different topic, resulting in different responses for question three.

Internal consistency reliability as demonstrated above is often reported as "Cronbach's Alpha." Reliability as measured by Cronbach's Alpha ranges from 0.0 to 1.0, with higher values indicating higher consistency in student responses across assessment questions compared to lower values (Bandalos, 2018; Traub & Rowley, 1991). Generally, if individual student responses are consistent and rank-ordered similarly across assessment questions on the same instrument, reliability will be high. The further reliability deviates from 1.0, the less precisely students' alcohol use has been measured. If the questions do in fact measure alcohol use, Cronbach's Alpha may be interpreted as the variation in scores due to differences in students' alcohol use. For example, consider you observed a Cronbach's Alpha estimate of 0.75 for students' ACS responses. The value of 0.75 indicates that 75% of the variability in students' scores is due to differences in students' alcohol use, whereas 25% of variability in students' scores is due to error, or factors other than students' alcohol use. A Cronbach's Alpha reliability index of 1.0 is elusive, as there will always be error present in our assessment scores. However, higher values are generally preferred over lower values. For practitioners interested in reading more about the technical definitions and calculations of Cronbach's Alpha reliability, the original paper by Cronbach (1951) and an article by Cortina (1993) are useful resources. For information about reliability concepts more broadly, the didactic module by Traub and Rowley (1991) is a useful resource.

As discussed, reliability is often the first piece of evidence collected to support the meaning of assessment results. However, reliability is insufficient for claiming that ACS scores represent students' alcohol use. For example, consider if all ACS questions were about drug use. Students may be rank-ordered similarly across assessment questions, yielding a high Cronbach's alpha, but scores would represent drug use, not alcohol use. A host of evidence is necessary to confidently say that ACS scores represent students' alcohol use, which brings the discussion back to validity.

Validity

Thus far we have defined validity in simple terms, relating it to the concept of "truth." To deepen our understanding of validity, more precision is needed. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), validity is defined as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). From this definition, three important conclusions can be drawn. First, **validity is a process of accumulating evidence**. Specifically, there are five types of evidence needed to make strong claims about validity: evidence related to content, response processes, internal structure, relation to other variables, and consequences. Although it is beyond the scope of this article to describe these types of evidence in detail, Table 1 summarizes information from the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) and provides examples of research questions and analyses that align with each type of evidence. With respect to the hypothetical ACS example, imagine the test developers provided adequate validity evidence related to content internal structure, and consequences. You are concerned, however, that underage students may be hesitant to report their true alcohol consumption given the illegal nature of their actions. In other words, you believe the validity of ACS scores as indicators of alcohol use may be compromised due to social desirability bias—an individual's motivation "to present herself or himself in a way that society regards as positive" (DeVellis, 2017, p. 136). One way to investigate this concern is to use a think-aloud procedure (or cognitive interview) to collect *evidence based on response processes* (Padilla & Benítez, 2014). This procedure would involve administering the ACS to a small group of students and asking them to verbalize their thoughts while completing the instrument. If students express hesitation or discomfort while responding to the items, it might

suggest that social desirability bias is an issue. Social desirability bias could also be investigated by administering a measure of social desirability (e.g., the Marlowe-Crowne Social Desirability Scale) and correlating scores on the scale with scores from the ACS (Fischer & Fick, 1993; DeVellis, 2017). If a relationship is found between the two variables, it would suggest ACS scores may be partially explained by socially desirable responding. Depending on the strength of the relationship, this *evidence based on relations to other variables* may provide an argument against interpreting ACS scores as indicators of alcohol consumption.

The fact that validity is a process of accumulating evidence leads to an important second conclusion: **validity is a matter of degree**. One common mistake practitioners make is to trust an instrument because its makers claim it is “valid and reliable.” Although statements like this are common, they are problematic because they imply that validity is something that is either present or absent. In reality, validity is a *process*, not an outcome – a journey, not a destination. There are no numerical indices or benchmarks to determine whether validity has been “reached.” Thus, decisions about validity are never clear-cut; they require critical review of a body of evidence that often yields mixed conclusions.

Third, and finally, based on the *Standards* definition, **validity relates to how accurately researchers and practitioners interpret and use test scores**. Thus, it is inaccurate to talk about the validity of an instrument. Instead, conversations about validity should focus on interpretations and uses of *the results obtained from an instrument*. This is an important distinction because it means validity is not an inherent property of an instrument. Consider a college entrance exam like the SAT. There is validity evidence to support interpreting SAT scores as a measure of general aptitude for academic success in college. But how confident should a student affairs professional feel using the SAT to select resident advisors? The SAT was not designed for selection of resident advisors, and as such, major questions should be raised about whether SAT scores are indicative of students’ potential to succeed in a resident advisor position. This example illustrates why conversations about validity *must* take into account how scores from an instrument will be interpreted and used. When we hear that scores are “valid,” we must ask “Valid for what purpose?”

In sum, validity is about making an argument for the interpretations and uses of assessment scores. Thus, when considering the ACS results, we must consider whether sufficient evidence supports interpreting ACS scores as an indicator of students’ alcohol use. Without sufficient validity evidence, we cannot be sure that ACS scores reflect students’ alcohol use.

Conclusion

In short, knowledge of reliability and validity is important for the betterment of assessment. We hope this introduction inspires student affairs assessment professionals to embark on the journey of considering reliability and validity whenever collecting, reporting, and interpreting assessment data. Doing so builds trust for our assessment data and promotes the use of results and improvement of students’ learning.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. New York, NY: Guilford Publications.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.
- DeVellis, R. F. (2017). *Scale development: Theory and applications*. Los Angeles, CA: Sage Publications.
- Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurement, 53*(2), 417-424.

- Henning, G. W., & Roberts, D. (2016). *Student affairs assessment: Theory to practice*. Sterling, VA: Stylus Publishing, LLC.
- Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136-144. <https://doi.org/10.7334/psicothema2013.259>
- Traub, R. E., & Rowley, G. L. (1991). Understanding Reliability: An NCME Instructional Module on. *Educational measurement: Issues and practice*, 10(1), 37-45.
- Wise, V. L., & Davenport, Z. R. (2019). *Student affairs assessment, evaluation, and research: A guidebook for graduate students and new professionals*. Springfield, IL: Charles C. Thomas Publisher, LTD.

Dr. Madison Holzman is a Research Scientist at Curriculum Associates. She can be reached at mholzman@cainc.com.

Andrea M. Pope is a Doctoral Graduate Assistant for Student Affairs Support Services at James Madison University. She can be reached at Popeam@jmu.edu.

Dr. Jeanne Horst is an Associate Assessment Specialist and Associate Professor of Graduate Psychology at James Madison University. She can be reached at horstsj@jmu.edu.

Table 1.

Types of Validity Evidence Applying the ACS Example (AERA, APA, NCME, 2014; Bandalos, 2018)

Types of evidence	Validity-related questions	Analyses that may be used to gain validity evidence
Evidence based on test content: Evidence that the instrument covers the full breadth of the construct without capturing irrelevant information	<ul style="list-style-type: none"> Do the ACS questions reflect our theory-based definition of alcohol use? Do the questions adequately cover all aspects of alcohol use? Is there content irrelevant to alcohol use on the ACS? 	Have subject matter experts review test content, using theory and a table of specifications.
Evidence based on response processes: Evidence that students reason through assessment questions in appropriate ways	<ul style="list-style-type: none"> Do students understand what the ACS questions are asking? Are student responses affected by low motivation? Are student responses affected by social desirability bias? 	Use think-aloud protocols, in which students talk through their thought process while completing the test. Examples of complex analyses include diagnostic classification models or cognitive item response theory models
Evidence based on internal structure: Evidence that the instrument functions appropriately from a measurement perspective	<ul style="list-style-type: none"> How many dimensions underlie the ACS scores? Would two subscales better represent the scores than one overall scale? Do students from different populations respond to the ACS in similar ways? For example, do males and females with the same level of alcohol use answer the ACS questions similarly? 	Examples of complex analyses include <ul style="list-style-type: none"> Reliability estimates Exploratory or confirmatory factor analysis Item response theory Differential item functioning studies
Evidence based on relations to other variables: Evidence that scores from the instrument relate to other instruments and/or observable behaviors in predicted ways	<ul style="list-style-type: none"> Do students who abuse alcohol score higher on the ACS than students who do not abuse alcohol? Do ACS scores relate to scores on similar tests of alcohol use? Do ACS scores predict whether or not someone will abuse alcohol? 	Analyses may include correlations of scores with other variables, examination of group differences, classification analyses, or the examination of a multitrait-multimethod matrix.
Evidence for consequences of testing: Evidence that positive, intended consequences are realized while negative consequences are minimized	<ul style="list-style-type: none"> What consequences (positive/negative and intended/unintended) result from our interpretations and uses of the ACS? Do the benefits using the ACS outweigh any negative consequences? 	Carefully consider the consequences of proposed interpretations and uses of test scores (especially negative consequences that may stem from other validity issues).

Sustain the Gain: Maintaining an Effective Learning Outcome and Assessment System

By Kathy Telban

Abstract: Outcome-based education and assessment have been a focus of higher education for over two decades, yet it is still being cited in accreditation reports as a major concern that requires interim progress reporting. Even when a college is acknowledged for their efforts and seen as a model for others, a change in leadership can put the future of assessment and all the effort at risk. How do we sustain the ground we have gained with learning outcomes and assessment until it is a system that continually improves and evolves and becomes part of the fabric of the institution? After decades of working with many different types of colleges, common themes and issues emerged that are critical to both establishing and sustaining the gain in order to maintain an effective learning outcomes and assessment system. Sustaining the gain involves four integrated processes, four essential support functions, and four major factors. This 4 x 4 x 4 systems approach serves as a new framework to use to develop, assess, improve, and sustain an integrated outcomes and assessment system. Colleges and universities can assess their own system by examining each process, support function, and major factor to determine improvements to the system that will ensure evolution and sustainability.

Introduction

At the Learning Organization, a consulting firm that helps college and workplace educators create and improve learning outcome and assessment systems that are sustainable, the outcome strategists are consistently asked the following question by the colleges they work with: How do we sustain the ground we have gained with learning outcomes and assessment to embed learning assessment in the fabric of the institution? Based on two decades of working for and with colleges and universities, outcome strategists have developed a 4 x 4 x 4 systems approach: four integrated processes, four essential support functions, and four major factors.

Institutions have struggled with designing, developing, and implementing a sustainable comprehensive approach to assessing learning that leads to meaningful results (Kuh, Ikenberry, Jankowski, Cain, Ewell, Hutchings, & Kinzie, 2015). A systems approach is being used in other areas effecting learning. A learning systems paradigm, introduced in the book *Degrees That Matter*, is one that entails thinking about the relationships among different parts of the learning environment in a collaborative way that brings together faculty and staff (Jankowski & Marshall, 2017). The way assessment work is organized should focus on integrating, connecting, communicating, and ultimately supporting meaningful structures and processes that enhance learning (Kuh, Ikenberry, Jankowski, Cain, Ewell, Hutchings, & Kinzie, 2015). An outcome and assessment system can also integrate with other college systems such as a human resource system where assessment efforts are a factor in performance evaluations (Suskie, 2009). The 4 x 4 x 4 systems approach described below builds on existing research and professional experience to provide a new framework to examine and improve an outcome and assessment system.

Four Integrated Processes

Four core curricular processes make up an integrated learning outcome and assessment system. These processes include curriculum design and approval, learning assessment, professional development, and program review as shown in the figure below (Telban & Stiehl, 2017). These four processes are constantly cycling and informing each other.

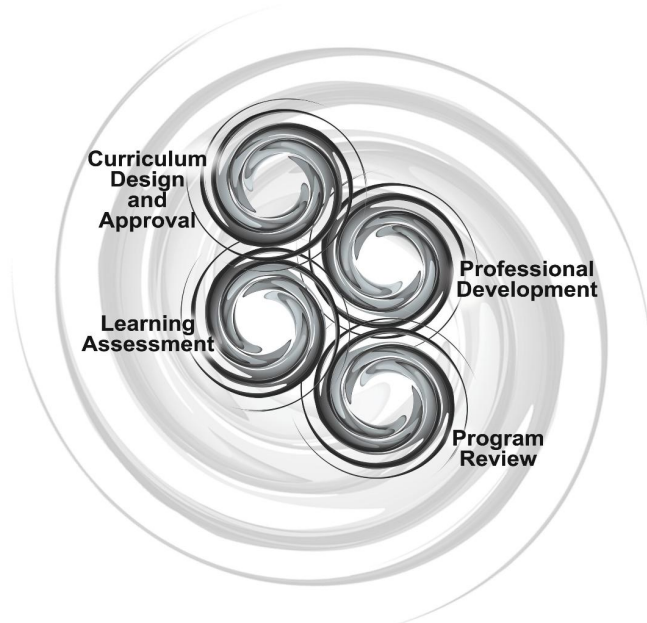


Figure 1: Four Core Curricular Processes in a College Outcomes and Assessment System

The foundation process, *curriculum design and approval*, begins with carefully planned curriculum design that results in establishing robust learning outcomes that are approved by faculty and administrators at the course, program, and institutional levels. According to Suskie (2009), the “curriculum approval process should include complete and feasible assessment plans” (p. 313), thus integrating with learning assessment processes. *Learning assessments* are aligned with learning outcomes to provide the feedback faculty need to make decisions about improvements to help learners meet the intended learning outcomes.

Periodic *program reviews* ensure a quality and relevant program that prepares learners for real-life roles such as a global citizen or related to a specific profession. According to Suskie, (2009) “Program review is a comprehensive evaluation of an academic program that is designed both to foster improvement and demonstrate accountability” and “because student learning is a fundamental goal of any academic program, student learning assessment should be a primary component of the program review process” (pg. 14), thus integrating with learning assessment. Finally, commitment to and a process that provides continual *professional development* of faculty, staff, and administrators is critical to build capacity to do this work.

When these four processes are integrated, they reinforce each other and create a flow of ideas, energy, and information among the interdependent processes. They must work together so the output of one process is the input to the next process. Learning outcomes is an output of curriculum design and approval and the input to the learning assessment processes (course, program and institution).

By having well-defined processes, the overall system is more effective and efficient. When implementing a basic skills initiative at Chaffee College, the college found that once integrated structures were in place, they did not need additional money to sustain them because the integration was more efficient than the previous stand-alone structures (Bailey, Jaggars, & Jenkins, 2015). *Figure 2: A Learning Outcomes and Assessment System* shows these four integrated processes in context:

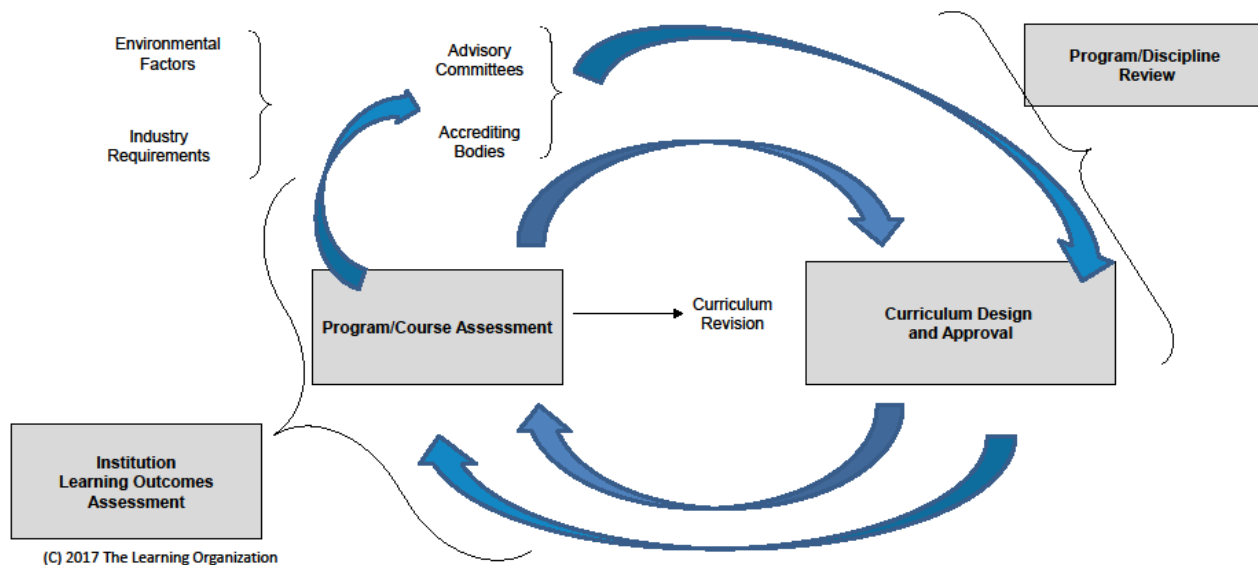


Figure 2: A Learning Outcomes and Assessment System

The system starts with external environmental factors and industry requirements which help justify the development of new degrees and certificates. Advisory committees are established to provide input and involvement that should include determining learning outcomes, assessment criteria, and authentic tasks that graduates should master. Accrediting bodies provide standards and requirements for professional programs and institutions, which serve as inputs to the learning outcomes and assessment system that begins with the curriculum design and approval process. Learning assessment involves both course and program assessment processes. Assessment results drive curriculum revision and should be shared with advisory committees and accrediting bodies. The institutional learning outcome assessment process should be conducted according to a schedule for each institutional outcome. Program review is also conducted on a standard basis for all academic programs and disciplines.

Each of these processes has its own cycle. According to Telban & Stiehl (2017), “A cycle exists whenever the output of a process becomes the input in the next cycle. Without a cycle there, there is no system; there is only an event” (p. 135). Cycles can be defined and repeated so that the outputs of one process are ready when needed by the next process. Since program review should include learning assessment data at the program and institutional level, if the processes for program assessment and institutional assessment have not been completed, there will be missing information to review and discuss during the program review process.

The cyclical movement is a reminder that nothing is static in this work. It is the constant cycling of the four integrated processes sharing information, data, and the feedback loops that energizes and sustains the system.

When mature, an integrated system is regularly monitored and improved through analysis, innovation, and sharing. The Higher Learning Commission has a rubric (see Figure 3) that can be used to evaluate the maturity of the system and determine how it could be improved.

STAGES IN SYSTEM MATURITY: PROCESSES			
Reacting	Systematic	Aligned	Integrated
The institution focuses on activities and initiatives that respond to immediate needs or problems rather than anticipating future requirements, capacities or changes. Goals are implicit and poorly defined. Informal procedures and habits account for all but the most formal aspects of institutional operations.	The institution is beginning to operate via generally understood, repeatable and often documented processes and is prone to make the goal of most activities explicit, measurable and subject to improvement. Institutional silos are eroding and signs of coordination and the implementation of effective practices across units are evident. Institutional goals are generally understood.	The institution operates according to processes that are explicit, repeatable and periodically evaluated for improvement. Processes address key goals and strategies, and lessons learned are shared among institutional units. Coordination and communication among units is emphasized so stakeholders relate what they do to institutional goals and strategies.	Operations are characterized by explicit, predictable processes that are repeatable and regularly evaluated for optimum effectiveness. Efficiencies across units are achieved through analysis, transparency, innovation, and sharing. Processes and measures track progress on key strategic and operational goals. Outsiders request permission to visit and study why the institution is so successful.

Figure 3: Rubric adapted from *Baldrige Excellence Framework: A Systems Approach to Improving Your Organization's Performance (Education)* by The Higher Learning Commission. (April 2015, Source AQIP Pathway Systems Appraisal: A Resource for Peer Reviewers) Note: There is also a rubric for Stages in System Maturity: Results.

Who monitors the system will depend on who is responsible for each process and where in the organization the process resides. In many colleges, learning assessment and program review resides in Institutional Research and Planning while curriculum design and approval resides in Academic Affairs. Having a joint responsibility for the entire system can be a shared responsibility requiring joint planning, or the system could be the responsibility of the Chief Academic Officer to which both areas report. The key point is that these processes should not be in silos and should not be managed as an event with no intent of a systematic method.

Four Essential Support Functions

There are four essential functions that support an integrated outcomes and assessment system that ensure its success. These are design support, documentation support, technology support, and the allocation of time. Typical barriers to assessment that have been identified are limited time and limited resources (Bresciani, 2006).

Design support requires administration leadership's commitment to build the instructional design expertise of faculty so that they develop and deliver effective and engaging learning activities that will build learners capacity to perform the learning outcomes. *Documentation support* means that all the information and resources needed by faculty to design, develop, deliver, and assess learning are readily accessible and current so that faculty know what is expected of them as well as how the four integrated processes work together and what tools to use. *Technology support* for faculty includes software systems to manage workflow, and the storage and reporting of the learning data and all curriculum related documentation. Lastly, the *allocation of time* must be planned for in order to participate in the on-going cycle of planning for learning, assessing learning results, reviewing results, and making changes based on results. This work must involve faculty because they make learning assessment happen. If nothing changes in the classroom, nothing really changes (Telban & Stiehl, 2017).

Integrating processes and support functions are important, but the one challenge that is most critical and often most difficult is getting people to work as one unit with a single purpose. The people who are responsible for the four core curricular processes need to plan, learn, share, and improve the system together as well as celebrate critical milestones and accomplishments. Faculty must also trust the process and have the support of

administration (Telban & Stiehl, 2017). According to Daniels (2007), “the primary benefit of trust, from a leader’s point of view, is that it accelerates learning” (p. 172).

Four Major Factors

There are also four major factors that have the greatest influence on sustaining assessment including consistent leadership over time, faculty involvement, professional development for everyone involved, and system integration to ensure continuity and sustainability. The identification of these four major factors is supported by Gilbert’s (1996) research on behavior engineering and Binder’s (1998) adaptation of his Six Box Model, which are both based on over 60 years of basic behavior science

Leadership is the factor that has the greatest effect on the sustainability of an outcome and assessment system. Leaders are at different levels and positions from the Chief Academic Officer, to the program leader, to the curriculum and assessment committee members, to the faculty who coordinate multiple sections of one course, and finally to the faculty who guide and lead learners. It is a shared and decentralized responsibility. Every time there is a change in leadership, there is a natural disruption. That is why it is important to have both a succession and a transition plan in place at each level of leadership.

Every institution undergoes changes in leadership and the new leader’s vision can either complement initiatives or conflict and hinder momentum. In a case study describing how recent guided pathways have been implemented, researchers found that administrative support and leadership were vital to move from pilot projects to systemic change (Jankowski & Marshall, 2017). If institutional leaders expect that important decisions are based on important goals, assessment becomes a natural part of planning and decision making and becomes part of the fabric of institutional life (Suskie, 2009).

While administrative leadership is important, it’s always the faculty who, in the end, guide learners and collect the evidence of learning. Additionally, *faculty involvement* must be central to improving assessment tools, policies that directly affect learner success, and the four integrated processes. Ideally, faculty help each other to co-create, thereby building trust, which is essential to sustaining an outcome and assessment system.

Professional development is critical for faculty who are subject matter experts in their various disciplines. It is important to not only stay current in the discipline, but also to build the capacity to develop outcome-based curriculum that is assessed and improved so that it stays relevant. Teaching, developing curricula, and assessing learning are different domains that require professional development and support. Many instructors and staff have no formal training and little opportunity to learn (Suskie, 2009). In other reform efforts, faculty, staff, and administrative time as well as professional development were cited as a common need for investment (Bailey, Jagers, & Jenkins 2015).

While a system is a network of interdependent parts that serve a common purpose, *successful system integration* is the efficient flow of ideas, energy and information among the different, interdependent parts of a single system (Telban & Stiehl, 2017). System integration is one factor that is rarely examined. Even the high cost of a college education has only focused on the numbers rather than processes that produce the numbers (Zemsky, 2013). Each process is typically seen as independent with little understanding of how they integrate. In order to function as an effective system, attention has to be given to the inputs, outputs, and feedback loops that connect the four core curricular processes. According to Meadows (2008), “when systems work well, we see a kind of harmony in their functioning” (p. 75).

In Summary

Creating a system that engages everyone in a cycle of assessment resulting in organizational improvement is neither simple or easy. It is a major task and a major accomplishment when it is done well (Stiehl & Null, 2017). Colleges and universities can assess their own system by examining their four integrated processes, four essential support functions, and four major factors to determine what improvements are needed to continually evolve and become part of the fabric of the institution. A sustainability tool using appreciative inquiry was developed to align with this 4 x 4 x 4 systems approach and is available for general use: <http://outcomeprimers.com/wp-content/uploads/Sustainability-Tools.pdf>

References

- Bailey, T., Jaggars, S., & Jenkins, D. (2015). *Redesigning America's community colleges: a clear path to student success*. Massachusetts: Harvard University Press.
- Binder, C. (1998). *The Six Boxes: a descendent of Gilbert's behavior engineering model*. Performance Improvement. http://www.sixboxes.com/_customelements/uploadedResources/SizBoxes.pdf
- Bresciani, M. (2006). *Outcome-based academic and co-curricular program review*. Virginia: Stylus Publishing.
- Daniels, A. & Daniels, J. (2007). *Measure of a leader*. New York: McGraw-Hill.
- Gilbert, T. (1996). *Human competence: Engineering worthy performance*. Maryland: International Society for Performance Improvement.
- The Higher Learning Commission. (2015). *AQIP pathway systems appraisal: A resource for peer reviewers*. Illinois: The Higher Learning Commission.
- Jankowski, N., & Marshall, D. (2017). *Degrees that matter: Moving higher education to a learning systems paradigm*. Virginia: Stylus Publishing.
- Kuh, G., Ikenberry, S., Jankowski, N., Cain, T., Ewell, P., Hutchings, P., & Kinzie, J. (2015). *Using evidence of student learning to improve higher education*. California: Jossey-Bass.
- Meadows, D. (2008). *Thinking in systems*. Vermont: Chelsea Green.
- Stiehl, R. & Null, L. (2017). *The ASSESSMENT Primer: assessing and tracking evidence of learning outcomes*. South Carolina: Create Space.
- Suskie, L. (2009). *Assessing student learning: a common sense guide*. California: Jossey-Bass.
- Telban, K. & Stiehl, R. (2017). *The SUSTAINABILITY primer: sustaining learning outcomes and assessment*. South Carolina: CreateSpace.
- Zemsky, R. (2013). *Checklist for change: making American higher education a sustainable enterprise*. New Jersey: Rutgers University Press.

Kathy Telban is the Chief Executive Manager of The Learning Organization and can be reached at Kathy@outcomeprimers.com.

Assessing Academic Support Centers: A Framework for Comprehensive Assessment

By Paul Mabrey

Abstract: Academic support centers occupy a complex place within the higher education landscape. They play a vital role for student success in that they typically provide supplemental instruction to the learning taking place within the classroom, often guided by peer educators and void of the grading judgement and accompanying power relationships. In these roles, assessment of academic support centers can prove challenging because of institutional placement, familiarity with assessment practices, (inter)disciplinary connections, and complex learning ecosystems. Taking the communication center community and James Madison University Communication Center as an example, this essay provides a framework and model for comprehensive assessment that can help other academic support centers navigate the complexities of assessment and learning improvement within their own institutional contexts. A comprehensive assessment framework should be focused on learning improvement; aligned across outcomes, curriculum, and assessment; collect diverse types of learning improvement evidence; and be comprehensive within and beyond the center.

Introduction

Learning and assessment of learning are not restricted to the college classroom. Assessment calls among Learning Centers have become commonplace but also demonstrate the need to move the assessment discussion forward. Within communication centers specifically, assessment has taken the form of tutor observations, end of appointment surveys, demographic tracking, faculty perceptions, or even supporting institutional communication assessments (Anderson et al, 2015; Clark, 2002; Jones, Hunt, Simonds, Comandena, & Baldwin 2004; Preston, 2011). While good starting points, these assessment practices reflect a compartmentalized approach to communication center assessment (Book and McCoy, 2016; Dyer & Davidson, 2012; Helsel & Hogg, 2006; Magee and Reynolds, 2016). Furthermore, these assessment efforts rarely demonstrate how they connect back to learning improvement. This paper, following the call by Leek, Carpenter, Cuny, & Rao (2015), provides a model for a comprehensive approach to communication center assessment focused on learning improvement. While the model is communication center-based, the framework for and examples from comprehensive assessment can be adapted to other peer education contexts, like learning centers, academic advising, recreation, and more.

A Comprehensive Assessment Framework

A comprehensive assessment framework for communication centers rests on four guiding principles. The four principles are:

1. Assessment should be focused on learning improvement;
2. Comprehensive assessment is aligned across learning outcomes, curriculum, and Assessment;
3. Comprehensive assessment collects diverse evidence of learning;
4. Comprehensive assessment means comprehensive within, across, and beyond the communication center.

First, assessment should be focused on learning improvement. Many examples identified within the literature could be characterized as client satisfaction or program evaluation. Supplementing those measures with a focus on learning assessment can help centers better tell the story of how they impact students and contribute to the institutional mission. According to Fulcher, Good, Coleman, & Smith (2014), we must first assess, then intervene, and then assess again to be able to properly assess learning improvement. Assessing or asking questions at the end does not provide sufficient evidence of learning improvement. One must measure before and after a learning intervention, whether that intervention is a tutoring

appointment, training session, or public speaking workshop, to properly make the case that any observable improvement in learning may be attributed to the learning intervention. Comprehensive learning assessment must be intentionally designed across the range of learning experiences.

Second, comprehensive means consistency and alignment across outcomes, curriculum, and assessment. One cannot focus on assessment alone and not have any sense of a learning outcome. Assessing whether or not a client becomes more proficient with oral citation practices does not mean anything if learning oral citation practices is not one of your learning outcomes. Moreover, you can have a learning outcome and assessment strategy, but still fall short if you do not have an appropriate intervention.

Third, comprehensive should mean including different forms and types of evidence for learning. A center should not focus only on client surveys or observable evidence of learning improvement. The former is easier to collect, but has limitations of self-reporting and not being directly observable. The latter is more difficult to collect, but may provide better learning insights. A comprehensive assessment framework should collect, analyze, and use multiple forms of evidence across a continuum of indirect/direct and immediate/long-term. This approach can balance some of the concerns about assessment: timeframe, expertise, assessment for assessment sake, resources needed, and accountability pressures.

Finally, comprehensive should be comprehensive across the communication center. Learning within a center is not contained to only the student client or even the student tutor. Faculty associates, staff, or faculty leading the center also learn through center administration. Learning is an inter-connected and inter-implicated process, transcending individuals who interact at the center. Comprehensive assessment should theorize, plan for, and implement assessment plans that take all learning into consideration.

James Madison University Communication Center's Assessment Model

Within a comprehensive assessment framework, the assessment plan should be aligned with learning outcomes and the curricular or co-curricular intervention. Furthermore, all of this should be consistent the center's mission. At James Madison University, the Communication Center's mission reflects a shift from public speaking instruction to situating communication peer education around public advocacy (JMU Communication Center, No Date).

Based on our mission, we developed tutor learning outcomes. The learning outcomes are simple and drive tutor training, programming, and assessment. The four learning outcomes are:

1. Apply tutoring best practices
2. Explain communication concepts and practices
3. Recognize oneself as an academic professional
4. Advocate on behalf of the study of communication, the Communication Center, your clients, and yourself.

One can look at the learning outcomes and understand how they align with and are developed from our mission. The mission statement and learning outcomes for tutors are integrated into tutor training for each academic year.

While the tutor learning outcomes came rather easily, the client learning outcomes took more work. In consultation with assessment professionals from the Center for Assessment & Research Studies, we developed client learning outcomes from appointment types we offer. There are ten client learning outcomes, for example: "Students who participate in outlining consultations will have a better understanding of how to complete an outline and progress toward an appropriate, complete outline."

As the learning outcomes were being developed for both tutors and clients, we were simultaneously engaging in a process of identifying where or how we supported that learning. For clients, it was clearer on the surface. If a student needed help on outlines, they signed up for an outlining appointment. We would help them make progress

toward learning about outlining and developing their specific outline. However, we also learned through this process that there was little standardization for how we approached a given client learning experience.

To address the lack of consistency in client learning, we turned to streamline tutor training. One benefit to this alignment process was the development of curriculum standardization. While a work in progress, we developed best practices for how each type of tutoring appointment might operate. We already had a best practices document for how a session should go generally, e.g. introductions, disclaimers, goal-setting, and tutoring. But now we were taking each appointment type and modeling what that session might look like. Take for example, an outlining appointment. The document for an outlining appointment provides an overview of outlining appointments and suggests specific adaptations of different aspects of session management. Different questions might be asked in the diagnosis phase and alternate tutoring techniques, like reverse outlining, are suggested and explained. This process shows how interconnected assessment planning really can be within a comprehensive assessment framework. Student client learning and tutor learning are linked together through training practices and processes to ensure implementation is consistent.

For tutor learning, aligning learning outcomes with curricular support seemed daunting, but was generally already present. For example, applying tutoring best practices was covered through initial tutor training, observations, weekly staff meetings, and professional development. Even a seemingly ambiguous outcome like “recognize oneself as an academic professional” was supported through training on role expectations, formal goal-setting, and research mentoring.

Now that learning outcomes and curricular interventions are drafted and aligned, what are the assessment practices within a comprehensive assessment framework? To start, we drafted a rubric for assessing tutor learning that includes different elements of tutoring best practices, communication concepts, professionalism, and advocacy. The rubric emerged out of the best practices document for session management and is covered during training, used to provide feedback during initial roleplaying simulations, and during later tutor observations. The rubric is a way to capture more direct evidence of learning for both formative and summative assessment.

We also utilize other measures to collect data for assessing tutor learning. For example, end of semester and year surveys collect self-reported perceptions of whether the tutor agrees that the coordinator supported their progression toward each tutor learning outcome. While self-reported data is more indirect, the use of formal goal-setting before and after the year provides more direct evidence of reflection and learning. Tutors are asked to reflect on tutoring practices specifically and non-tutoring related learning that occurred. This is helpful for individual tutor reflection and goal-setting, but also helps provide evidence to assess the tutor learning outcomes.

Different data are collected and reported for client learning. The traditional end of session survey is used successfully, and we ask a variety of questions to get at different aspects of client learning. First, we ask what kind of appointment they had and what their goals were. This helps us understand which outcome(s) they might be working toward, even if their appointment type does not necessarily align. For example, they may have signed up for a delivery or visual aid appointment, but learned more about outlining and brainstorming during that session. Second, we ask questions associated with satisfaction or confidence. While not always associated with learning outcomes, these measures are especially helpful within the context of communication learning outcomes because so much of communication learning is intimately tied up with self-perception, anxiety, and confidence. Third, we ask if they accomplished what they had hoped to accomplish and if not, we ask why they did not. These questions not only help us assess client learning, but also can provide insight into whether or not tutor learning is progressing as desired. Finally, we ask an open-ended question about transferable learning that can provide more direct evidence of learning. Clients are asked if they learned something they can apply to other communication situations and if so, what did they learn. The data helps to provide more comprehensive insights into what is learned that we did not anticipate and to identify gaps through thematic absences.

Using more direct measures of learning is one area where we are experimenting. This is particularly tough for communication centers because of the performative aspect of this learning, differences between what happens at the center versus what happens in class, and needing to include others to collect data. One way we started is with outlines, as the presentation outlines are an important piece of the assignment and student learning process. Through an IRB approved research protocol, we started collecting outline drafts from students when they come to the center. We then either collect the finished outline at the appointment's end or follow-up afterward for the outline version they submitted to their professor. This allows us to compare the before and after outlines to determine what kind of changes they have made based on the work we did together at the center. While not perfect, we are hopeful for the insight this type of evidence may provide into student learning and our assessment practices.

Using the results of our assessment efforts to make changes is the last way we operate within this comprehensive assessment framework. We do this through using client learning assessment evidence to evaluate and improve tutor training, making data-based programming and resource allocation decisions, integrating training changes based off of both tutor and client learning assessment to improve client learning, and utilizing all of the data to advocate for and on behalf of the communication center.

Implications

Taken separately, the elements of the comprehensive assessment framework for communication centers is not necessarily new, but when arranged to emphasize the comprehensive focus on learning improvement it provides an innovative way to think about assessment and evaluation within academic support centers like communication centers, learning centers, advising, peer health educators, and more. First, the framework shifts the assessment discussion from whether or not we should assess to how and why we assess. Second, the framework focuses assessment practices on actual learning rather than on measures more traditionally associated with program evaluation or usage. Third, the framework develops a model for how we can think more closely about comprehensive learning assessment, emphasizing the ways in which tutor, client, and other modes of learning happen within, throughout, and beyond the centers

While developed within the communication center context, the guiding principles for comprehensive assessment and model examples can be applied to other academic support areas. For example, writing centers might be interested in the outline assessment example as evidence of more direct learning. Academic advising or library research peers might benefit from the comprehensive alignment of learning outcomes, intervention, and assessment. Learning centers might be interested in the intervention standardization and training procedure to ensure some curricular consistency by peer educators, whether in café style or appointment-based tutoring.

The JMU Communication Center's assessment strategies gesture toward good practices for assessing learning through publicly sharing different ways to think about learning and measures for evaluating learning. By sharing the comprehensive assessment framework and specific model, we can help establish an empirical research agenda on tutor training, tutoring interventions, and center assessment. Finally, and perhaps most importantly, navigating this assessment journey and developing the assessment framework has demonstrated that the process is always more important than the product. Even if the learning outcomes are significantly revised, data collected not perfect, or analysis flawed—the process of thinking holistically about learning improvement already places us on the path to positively impacting learning.

References

- Anderson, L. B., Hearit, L. B., Morgan, M., & Natt, J. (2015). Using a Mixed-Methodological Approach to Assess the Communication Lab: Gaining Insights and Making Improvements. *Communication Center Journal*, 1, 9-36.
- Clark, R.A. (2002). Learning outcomes: The bottom line. *Communication Education*, 51(4), 396-404.
- Book, C.A. & McCoy, M. (2016). Tutor Observations as a Tool for Creating a Supportive and Productive Tutoring Environment. In W. Atkins-Sayre and E. L. Yook (Eds.), *Communicating Advice: Peer Tutoring and Communication Practice* (301-305). New York: Peter Lang Publishing, Inc.

- Dyer, K.K. & Davidson, M.M. (2012). Speech Center Support Services, the Basic Course, and Oral Communication Assessment. *Basic Communication Course Annual* 24, 122-150.
- Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014). A simple model for learning improvement: Weigh pig, feed pig, weigh pig. *Occasional Paper*, (23).
- Helsel, C. R., & Hogg, M. C. (2006). Assessing communication proficiency in higher education: Speaking labs offer possibilities. *International Journal of Listening*, 20(1), 29-54.
- James Madison University Communication Center (No Date). Communication Center Mission. Retrieved from <https://www.jmu.edu/commcenter/about/index.shtml>.
- Jones, A.C., Hunt, S.K., Simonds, C.J., Comandena, M.E., & Baldwin, J.R. (2004). Speech Laboratories: An Exploratory Examination of Potential Pedagogical Effects on Students. *Basic Communication Course Annual* 16, 105-138.
- Leek, D., Carpenter, R., Cuny, K.M., & Rao, P.A. (2015). Strategies for Assessment in Communication Centers: Perspectives from Across the Field. *Communication Center Journal*, 1, 49-59.
- Magee, N. & Reynolds, C. (2016). Peer Consultant Evaluation. In W. Atkins-Sayre and E. L. Yook (Eds.), *Communicating Advice: Peer Tutoring and Communication Practice* (207-209). New York, NY: Peter Lang Publishing, Inc.
- Preston, M. M. (2011). Assessment of Your Communication Center: Anathema or Asset? *Communicating Advice: Peer Tutoring and Communication Practice* (177-191). New York: Peter Lang Publishing, Inc.

Dr. Paul Mabrey III is the Communication Coordinator and Assistant Professor in the School of Communication Studies at James Madison University. He can be reached at mabreype@jmu.edu.

Predicting Success in a Licensure Program

By *Fiorella Peñaloza*

Abstract: Predictability of licensure exam success as part of the Doctor of Chiropractic at Cleveland University-Kansas City was evaluated using known academic performance metrics. The examined measures included undergraduate GPA at entry, first-term GPA, and performance in specific basic science courses. One basic science course was identified as a predictor of the National Board of Chiropractic Examiners (NBCE) Part I Licensure performance using logistic (logit) regression analysis. The logistic regression equation and effectiveness of the model used to predict licensure exam success are provided in this study. Similar relationships between performance in basic science courses and licensure exams have been reported in other health professions. The logistic regression analysis for determining licensure success provides flexibility in determining success, which may vary by licensure program, and thus this study provides broad applicability of predicting success of licensed professions.

Introduction

The National Board of Chiropractic Examiners (NBCE) written exams represent a distinctive programmatic measure of accountability and success. The Council of Chiropractic Education, programmatic accrediting body, requires all accredited programs to publish pass rates of each board exam. Cleveland University-Kansas City currently has the highest NBCE licensure completion rates in the nation (Cleveland University-Kansas City, 2018). In an effort to sustain this distinction, the Doctor of Chiropractic (DC) program review conducted a comprehensive review of academic determinants on NBCE board exam performance, specifically the NBCE Part I, which was the impetus of this study. The DC Curriculum Committee defined the need to develop a model to predict student success on the NBCE Part I exam by examining specific courses mapped to NBCE Part I competencies, undergraduate GPA, and first-term GPA as possible predictors. An abbreviated curriculum map is provided in Table 1. These courses represent the basic science component of the Doctor of Chiropractic (DC) program curricula, which are introduced to students within their first five trimesters of matriculation. Successful completion of these courses and other program requirements provide students with eligibility for NBCE Part I testing in their seventh term.

Methods

Deemed exempt by the institutional review board of Cleveland University-Kansas City, this study examines academic performance between one dependent dichotomous variable, formed by students that passed or failed the NBCE Part I exam. The difference in academic performance for various metrics between the two groups was investigated as part of the predictive modeling used in this study. Academic performance data (independent variables, see Table 2) was gathered from the Student Information System (SIS). The program's curricular map was used for the identification of courses whose outcomes mapped to NBCE Part I competencies (see Table 1). Concurrent with changes in admissions policy set by the Council on Chiropractic Education, entry-level GPA of applicants was calculated using the courses adding up to 90 undergraduate credit hours with a minimum 3.0 cumulative grade point average (GPA) on a 4.0 scale. The 90 credit hours must include a minimum of 24 semester hours in the life and physical sciences, of which half of the courses must have a laboratory component. First-term GPA and grades of all required basic science courses mapped to NBCE Part I competencies in the Doctor of Chiropractic program (DCP) were also gathered from the SIS. Academic performance data was combined with NBCE Part I performance data (provided by NBCE) given the presence of student identifiers that allowed this type of aggregation.

Student data were initially transcribed to a Microsoft Excel spreadsheet and then converted for analysis in IBM SPSS Statistics for Windows, Version 26.0 released 2018 (IBM Corp, Armonk, NY). Logistic regression was used for analysis involving a dichotomous (binary) dependent variable. Dummy variables were created for each dichotomous variable, where “0” represented students that failed the NBCE Part I and “1” represented students that passed the NBCE Part I exam. For all analysis, the *a priori* level of significance utilized was 0.05 ($p < 0.05$).

When utilizing licensure exam data, it is important to identify a passing score. The threshold for passing scores may vary from state to state and should be considered given the potential for mobility of students. For this study, a score of 375 or higher was used as the passing threshold to account for state-to-state variances of NBCE exam pass scores. This approach was used in the determination of dichotomous outcome variable (see Table 2). Furthermore, given that passing scores are subject to change over time, the development of a dichotomous variable allows for future iterations of this study to be comparable over time.

The initial data set contained 256 students, which represented the total number of first-time test-takers that took the NBCE Part I exam between March 2016 to August 2018. After cleaning the data of missing information, the sample size was reduced to 235 students. The data was cleansed of outliers to increase the quality of the predictive model, which reduced the population to 187 students and are included in this study. Independent variables were plotted against a normal distribution and those that did not follow a normal distribution were excluded, which reduced the number of examined courses from 25 to 18 (see Table 1). Using normalized independent variables ensures the data utilized from our sample population is representative and ensures the predictive model generated is robust.

Results

Given the presence of a dichotomous dependent variable, a logistic regression analysis was the preferred approach for examining predictability (Cabrera, 1994). In addition to the courses mapped to NBCE Part I competencies (see Table 1), the entry-level GPA and first trimester GPA were also tested in the logistic regression model. Of these academic measures, only the Thorax, Abdomen, and Pelvis Anatomy (TAPA) had a measurable predictability on licensure performance (see Table 3). The ability for TAPA to predict NBCE Part I success is significant (see Table 3). Descriptive statistics of the predictor variable suggest that a grade of a “B” or above are indicative of passing the NBCE Part I licensure exam (see Table 4).

The overall model was evaluated using the Likelihood Ratio, Chi-Square, Hosmer-Lemeshow test, and the percentage correct between the observed and predicted values. A significant Likelihood ratio and Chi-Square indicates significant improvement of the tested model over the null model and was demonstrated in the tested model ($p < 0.001$ and $p < 0.01$, respectively; see Table 5). Hosmer-Lemeshow test is commonly used as a goodness-of-fit for logistic regression and indicates a poor fit is the significant level is less than 0.05. Given the Hosmer-Lemeshow test statistic is 0.182 and is not significant for the tested model, goodness-of-fit is demonstrated (see Table 5).

The degree of alignment between the predicted probabilities and actual observations is dependent on the association between these values. While the Cox & Snell (pseudo $R^2 = 0.231$) and Nagelkerke R^2 (max rescaled $R^2 = 0.457$) values provide some insight about this relationship, the Kendall’s Tau and Goodman-Kruskal is the preferred approach given the design of this study. The Kendall’s Tau-b and Goodman-Kruskal values of 0.435 ($p < 0.01$) and 0.946 ($p < 0.001$) respectively, indicate that a positive and significant relationship exists between TAPA performance and performance in the NBCE Part I licensure exam. The accuracy of a logistic regression model was examined by using a receiving operating curve (ROC Curve); where the closer the curve is to one (1), the more accurate the model is. Using TAPA performance as a predictor of licensure performance resulted has an accuracy of

0.88 or 88% (Std. Error = 0.037). In addition to testing the accuracy, the validity of the predicted probabilities is 89.3% (see Table 6).

Analysis and Discussion

Of the 20 indicators tested, Cleveland University-Kansas City found that the best predictor for NBCE Part I success was student performance in the first attempted basic science course Thorax, Abdomen, and Pelvis Anatomy (TAPA). These results suggest that student success in the NBCE Part I exam is dependent on their education, specifically in TAPA. Call and Harvey (2014) tested GPA at entry, in-program GPA, and performance in basic science courses among other indicators as potential predictors of NBCE Part I and II and concluded that the basic science courses ultimately determined student success. Similarly, Kenya, Kenya and Hart (2013) identified basic sciences courses as predicting NBCE Part I licensure success. Basic science courses were also identified as predictors of the Canadian Chiropractic Examining Board (2006). Other health professions have also examined the significant influence of basic science courses on licensure outcomes (De Ball, Sullivan, Horine, Duncan, & Replogle, 2002; McCall, Allen, & Fike, 2006).

While findings from other studies support the results of this study, it is important to note the differences in analytical approaches utilized. Similar studies within chiropractic education have utilized linear regression or variations of this analytical approach (Kenya, Kenya, & Hart, 2013; Lawson & Till, 2006; McCall & Harvey, 2014), whereas this study utilized a logistic regression to examine predictability. The logistic regression analysis required the outcome variable (or NBCE Part I exam score) to be converted to a pass or fail category, which was done by incorporating state-to-state variances of passing thresholds. This method ensures a comparative approach of examining NBCE Part I pass scores, where the scores are not only examined as raw values but are comparatively evaluated across all state thresholds. Furthermore, this approach supports the student profile at Cleveland University- Kansas City, which have sizeable rate of mobility (i.e. leaving Kansas) after graduation that is also a consideration of the program. As a result of this approach, the predictive model generated from is captured in the equation listed below (derived from the logistic regression equation; Tai & Machin, 2014), where β_{Constant} is -6.149, the $\beta_{\text{TAPA Grade}}$ is 3.032, and the TAPA Grade is the numeric value based on its alphanumeric grade:

$$P = \frac{\text{Exp}(\beta_{\text{Constant}} + (\text{TAPA Grade}) (\beta_{\text{TAPA Grade}}))}{1 + \text{Exp}(\beta_{\text{Constant}} + (\text{TAPA Grade}) (\beta_{\text{TAPA Grade}}))}$$

The significance of TAPA performance suggests the need for further distribution of course objectives across the curriculum. For this reason, additional analysis is needed to examine the objectives within the TAPA course and how they are aligned with pre-requisite courses (i.e. Systemic Anatomy and Embryology) and subsequent coursework. Continuous evaluation of TAPA performance can assist with the evaluation of the curriculum itself and where the integration of TAPA course objectives throughout the curriculum would be anticipated to influence student success in the NBCE licensure exam. TAPA performance can also help inform additional competencies that may need to be reviewed in the licensure preparation course (currently optional to students). TAPA performance could also be used as an indicator of potentially at-risk students, and help with forecasting future resources and support the department's planning cycle. Utilizing TAPA performance as an indicator may also have a downstream effect on influencing program retention and completion.

This study provides a quantitative framework for measuring the influence of academic measures on licensure performance that can be used for early identification of at-risk students in licensure programs. Licensed programs interested in designing this study for their programs would have to modify the data collection to include academic performance metrics believed to have an influence on licensure outcomes. Utilizing program requirements, including those used for admission and programmatic academic requirements typically required for graduation, can identify a comprehensive listing of academic variables and provide a starting point for faculty to map requirements to licensure outcomes. For accredited programs, integrating program effectiveness measures as identified by the programmatic accrediting agency can further increase the relevance and application of this effort. In addition to

assembling a program effectiveness map, academic measures must also be defined (see Table 3). Ideally, the process of mapping measures to outcomes and defining academic measures is faculty led. In this study, the Doctor of Chiropractic Curriculum Committee was the driver of inquiries and final governing body of programmatic decisions, including the program map and effectiveness measures used by the program. Furthermore, it is important to note that student outcomes are influenced by additional factors (outside the scope of this study) such as behavior, socioeconomic status, health, and environment among others that could also be incorporated into predictive modeling.

Limitations

Limitations of this study should be considered. Since the research design was retrospective, a purely randomized sample could not be established nor can causality be inferred. The results should be interpreted with caution as factors beyond the academic measures measured for this study may also have an influence on licensure outcomes. Furthermore, this study was conducted at a single institution and thus, the findings may not be generalizable to other chiropractic-offering institutions.

Conclusion

TAPA performance is a relatively strong indicator of student performance on licensure board outcomes. This study contributed to the growing body of chiropractic education literature and affirmed previous studies in the chiropractic field. Based on the outcomes of this study, TAPA performance can be used to identify at-risk populations to target and provide additional academic support. Given that entry-level did not have a significant influence on predicting NBCE outcomes, this suggest that independent of varying levels of academic preparedness, student success in the NBCE Part I exam is significantly dependent on TAPA performance. Future analysis of the TAPA curriculum, including course pre- and post-requisites to examine the structural design of courses within this course sequence, can inform structural changes to program design.

Acknowledgements

We would like to thank the Doctor of Chiropractic Curriculum Committee Members for their inquiries that resulted in the development of the research questions in this study, the NBCE for providing data on licensure outcomes, and Ms. Zarina Hussain for serving as the research assistant for this study. We would also like to extend our thanks to Dr. Carlos Peñaloza and Dr. Jeff Barbee for editorial assistance.

References

- Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. In John C. Smart (ed.), *Higher Education: Handbook of Theory and Research* (Vol. 10, pp. 225–256). New York, NY: Agathon Press. Retrieved from: <https://education.umd.edu/file/chapteronlogisticregressionpdf>
- Cleveland University-Kansas City. (2018). Cleveland University-Kansas City Leads Nation in Board Completion Rates. Retrieved from <https://www.cleveland.edu/news-post/~post/cleveland-university-kansas-city-leads-nation-in-board-completion-rates-20181213/>
- De Ball, S., Sullivan, K., Horine, J., Duncan, W.K., & Replogle, W. (2002). The relationship of performance on the dental admission test and performance on Part 1 of the National Board Dental Examination. *J Dent Educ.* 66:478–484. <http://www.jdentaled.org/content/66/4/478.long>
- Kenya, A.M., Kenya, H.M., Hart, J. (2013). Correlation between academic performance and NBCE part I scores at a chiropractic college. *Journal of Chiropractic Education.* 27(1): 27– 32. DOI: 10.7899/JCE-12-010

- Lawson, D.M. & Till, H. (2006). Predictors of performance of students from the Canadian Memorial Chiropractic College on the licensure examinations of the Canadian Chiropractic Examining Board. *J Manipulative Physiol Ther.* 29(7):566–569. DOI: 10.1016/j.jmpt.2006.06.021
- McCall, K.L., Allen, D.D., & Fike, D.S. (2006). Predictors of academic success in a doctor of pharmacy program. *Am J Pharm Educ.* 70 (5):106. doi:10.5688/aj7005106
- McCall, A.R. & Harvey, R.D. (2014). Predictors of performance on the National Board of Chiropractic Examiners Parts I and II. *Journal of Chiropractic Education:* 28(1): 9-15. DOI: 10.7899/JCE-13-7
- Tai, B. C. & Machin, D. (2014). *Regression methods for medical research*. Retrieved from <https://ebookcentral.proquest.com>

Dr. Fiorella Penalozza was previously the Director of Institutional Effectiveness and Assessment and Assistant Professor at Cleveland University in Kansas City. She is currently a Senior Academic Research Officer for the University of Hawai'i System. She can be reached at fpenaloz@hawaii.edu.

Table 1: List of Basic Science Courses Mapped to NBCE Part I Exam Competencies

Course Name	Term*	Course Included in the Study
Embryology	1	Yes
Systemic Anatomy	1	Yes
Spinal Anatomy	1	No
Histology	1	No
Biochemistry I: Structure and Function of Macromolecules	1	Yes
Cell Physiology	1	Yes
Thorax / Abdomen / Pelvis Anatomy (TAPA)	2	Yes
Extremity Anatomy	2	Yes
Biochemistry II: Digestion / Intermediary Metabolism	2	Yes
Public Health I: The Health Care System	2	Yes
Immunobiology	2	Yes
Cardiovascular / Pulmonary Physiology	2	No
Endocrine / Reproductive Physiology	2	Yes
Head / Neck Anatomy	3	Yes
Basic Nutrition	3	Yes
Microbiology I: Bacteriology	3	Yes
Public Health II: Epidemiology	3	Yes
General Pathology	3	No
Renal / Digestion Physiology	3	No
Neuroanatomy	4	Yes
Microbiology II: Virology / Parasitology / Mycology	4	Yes
Cardiovascular / Pulmonary / Gastrointestinal Pathology	4	Yes
Neurophysiology	4	No
Neuromusculoskeletal / Genitourinary Pathology	5	Yes
Physiology Laboratory	5	No
Students become eligible to take the NBCE Part I licensure exam and begin clinical coursework in the DC program.	7	

Note. Academic terms are in trimesters and the course sequence provided here follows the 10-trimester progression model.

Table 2: Definition of Variables

Variable	Variable Type	Metric Time frame	Definition
NBCE Part I	Dependent Variable	This measure is typically determined during the 7th trimester for students enrolled in the 10-trimester progression, where program completion occurs in the 10th trimester of enrollment in the Doctor of Chiropractic program.	The NBCE Part I variable is made up of two groups of first-time test-takers; the pass group and fail group. The NBCE Part I score of 375 was used to identify the pass and fail groups, a score of 375 or higher was included in the pass group and a score below 375 was included in the fail group. The data was coded as "0" and "1" for fail and pass groups, respectively.
Entry-level GPA	Independent Variable	This measure is determined prior to students beginning the Doctor of Chiropractic program.	Admission to the Doctor of Chiropractic program requires students to have earned 90 credit hours with a minimum 3.0 cumulative grade-point average on a 4.0 grade scale. The 90 hours will include a minimum of 24 semester hours in the life and physical sciences of which half the courses must have a substantive laboratory component.
First-Term GPA	Independent Variable	This measure is determined during the first-term of matriculation in the Doctor of Chiropractic program.	First-Term GPA is the term GPA calculated at the end of the first trimester of matriculation in the Doctor of Chiropractic program at CUKC.
Course Grade	Independent Variable	This measure is determined at the end of each semester of matriculation in the Doctor of Chiropractic program.	Final course grades were converted to numeric values (as established by the program and published in the catalog) analysis. For repeated courses, the initial final grade was utilized for analysis. The passing grade scale includes grades of A, B, and C only.

Table 3: Logistic Regression Analysis for Variables Predicting NBCE Part I Licensure Success

Predictor	β	S.E. β	Wald's χ^2	df	p	e^β (odds ratio)
Thorax / Abdomen / Pelvis Anatomy (TAPA)	3.032	0.556	29.77	1	0.000	20.741

Table 3 provides a summary of the logistic regression analysis performed, which examined the relationship between performance in the Thorax / Abdomen / Pelvis Anatomy (TAPA) course and passing the NBCE Part I

Licensure Exam. The beta coefficient (β) indicates that there is a positive relationship between these measures which has relatively low variance (S.E. β) across the examined cases. The Wald's X^2 value is significant ($p < 0.001$) indicating TAPA as a predictor of NBCE Part I Licensure success. The odds ratio indicates a relatively strong association between these two measures where higher TAPA performance is associated with higher odds (20 to 1) of passing the NBCE Part I exam.

Table 4: Descriptive Statistics of the Model Provide Insight on Grade Threshold

NBCE Part I Outcome	Total Sample	Mean	SD
Failed NBCE Part I	21	2	0.436
Passed NBCE Part I	166	3	0.621
Total	187	3	0.687

Table 4 provides the descriptive statistics included in the model including the sample size, average (mean) and standard deviation (SD). The average grade in the Thorax / Abdomen / Pelvis Anatomy (TAPA) course is included for students that passed and failed the NBCE Part I indicating the threshold of TAPA performance associated with licensure success.

Table 5: Evaluation of the Effectiveness of the Model

Test	X^2	df	p
Overall Model Evaluation			
Chi-Square	49.036	1	0.000
Likelihood Ratio Test	82.349	1	0.001
Goodness-of-fit Test			
Hosmer & Lemeshow	0.182	1	0.670

Table 5 provides a summary of the various tests of model effectiveness performed. The chi-square values indicate the model's overall goodness-of-fit ($p < 0.001$) and the likelihood ratio indicate the likelihood of the model's predictive capacity is significant ($p < 0.001$). The Hosmer-Lemeshow Test is a goodness of fit test specific to logistic regression analysis where the non-significant chi-square value ($p = 0.670$) indicates goodness of fit of the model.

Table 6: Evaluation of the Percentage Correct Predicted by the Model

Observed	Predicted		% Correct
	Yes	No	
No	5	16	76.2
Yes	151	15	91
Overall % Correct			89.3

Note. The cut value is 0.500.

Table 6 provides a comparison of percent correct between the observed values and predicted values, which include the overall percentage correct predicted values (89.3%) as well as the rate of false positive (5/21 or 24%) and false negative (15/166 or 9%) generated by the model.

What does it Mean to Commit to Excellence? Reflections of an Assessment Professional Serving as a Baldrige Examiner

By *Jeremy Penn*

Abstract: Assessment is often framed as a process to use to support institutions' pursuit of excellence. Yet achieving excellence clearly requires much more than completing a few cycles of assessment. What does it mean for an organization to truly commit to excellence? And what role do assessment leaders have in organizations that seek to live out this commitment? This essay attempts to answer these questions by reflecting on the author's experience serving as a Baldrige examiner for two organizations with a deep commitment to excellence. Organizations that are committed to excellence know the areas in which they need to improve and proactively work on getting better, they narrow the number of things they seek to achieve, and operate out of a deeply held internal desire to improve. Assessment leaders in organizations that are committed to assessment should pay less attention to compliance and spend more of their time creating the conditions that support improvement and collaborating on a broad range of projects on student learning and the student experience. The essay concludes that while becoming an excellent organization is tremendously hard work that requires discipline and persistence, it is the only way we will achieve our shared mission of preparing future leaders and citizens of the world.

Introduction

"Excellence" has been a key objective for every one of the four higher education institutions where I have implemented assessment. In their strategic plans, these institutions talked about a "vision of excellence," "educational excellence," or supporting "faculty, staff, and students to pursue excellence." Excellence is a laudable strategic goal. These institutions prepare future teachers, future health care professionals, future engineers, future farmers, and future leaders. All of these individuals will be called upon in the coming decades to solve our planet's pressing problems, such as the climate crisis, food instability and hunger, access to clean water, and equity for all people around the world. Solving these problems will require nothing short of excellence.

Unfortunately, institutions' calls for excellence ring hollow when these plans lack a description of the specific actions, they will take to achieve excellence. In this sense the word "excellence" becomes a cliché, nothing more than a word that institutions use to vaguely describe their goals and to ask their stakeholders to do more of what they are already doing. Reading these strategic plans, I am left wondering: what would it look like if institutions truly committed to excellence? What role would assessment leaders have in institutions that were truly committed to excellence?

In this article I will attempt to answer these questions by reflecting on my experiences as an examiner with the Baldrige Excellence framework with two different organizations. I will start the article by describing the Baldrige Excellence Framework and the review process, then will attempt to answer the two above questions.

Brief Introduction to the Baldrige Excellence Framework

The Baldrige Performance Excellence Program was first developed in 1987 and is currently managed by the National Institute of Standards and Technology in the U.S. Department of Commerce. Much like an accreditation review, participating organizations begin the process by writing a self-study based on seven categories. The seven categories are 1) Leadership, 2) Strategy, 3) Customers, 4) Measurement, analysis, and knowledge management, 5) Workforce, 6) Operations, and 7) Results (Baldrige Performance Excellence Program, 2019). A team of trained examiners review the written response to the categories and then perform a site visit to confirm and clarify their findings and provide written feedback to the organization. The organization's performance is scored, and the

organization is recognized through a state program or, for organizations recognized with top awards in state programs, through a national program.

For readers who have been involved with accreditation, so far this is nothing new. In fact, the Higher Learning Commission (HLC) piloted the use of the Baldrige framework in accreditation and many aspects HLC's AQIP accreditation pathway (which is now being phased-out) ended up drawing strongly from the Baldrige framework (Wolff, 2005).

Examiner training for the Baldrige framework was organized around two distinct areas. For process-oriented categories (categories 1-6), focusing on processes the organization uses to carry out its mission, examiners are trained to use the "ALDI" initialism:

A: Approach. Approach asks if the organization has an approach for carrying out its processes.

D: Deployment. Deployment asks about the extent to which the approach is applied across the organization.

L: Learning. Learning asks if the organization is developing knowledge, skills, or creating innovative practices in response to its own study of the approach.

I: Integration. Integration asks if plans and processes are harmonized across the organization.

For the results category (category 7), examiners are trained to use the "LeTCl" initialism:

L: Levels. Levels ask for the results to be reported on a meaningful measurement scale that includes data on past performance, and achievement goals.

T: Trends. Trends ask for results to be reported over time, with at least three data points, so that it is possible to determine if results are improving or declining.

C: Comparisons. Comparisons asks for results to be put in context by providing comparative data from similar organizations.

I: Integration. As before, integration asks about the alignment and connection of the results to the things that are important to the organization.

Again, in general this is really nothing new to readers who have spent time working with accreditation (if you want more information about the Baldrige process, I suggest looking for opportunities in your state as most states have active Baldrige examination programs). However, the Baldrige examination process does differ from accreditation in two important ways. First, participation in the Baldrige examination process is voluntary. Second, the Baldrige examination process is about identifying an organization's strengths and opportunities for improvement, not determining the level of compliance with pre-set standards.

With this brief overview in mind, I would like to now turn to attempting to answer the two questions I posed at the beginning of this article.

What does it Mean to Commit to Excellence?

Both organizations I examined through the Baldrige examination process demonstrated a significant commitment to performance excellence. This is not to say these organizations were perfect in all they did. They were not. But the first thing that made them stand apart was *they knew the specific areas in which they needed to improve and were actively working on getting better*. Examples of efforts implemented by these organizations included:

- Dedicated offices with significant staffing levels that focused on engaging the entire organization in continuous improvement efforts. Rather than sharing data with an impartial or dispassionate stance, as may be common for many institutional research or assessment offices, these offices proactively propelled continuous

improvement efforts by pointing to key metrics and engaging the organization in activities—such as ‘Rapid Improvement Events’ or ‘Quick Win’ processes—designed to address those issues.

- All staff in these organizations understood the basic principles of continuous improvement and participated by identifying issues and contributing to solutions throughout their everyday work.
- Keeping a scorecard that was updated regularly that showed if they were winning or losing. One organization kept public ‘huddle boards’ in each department that showed how the department was performing relative to its goals. Departments met at least weekly, sometimes more, together at these boards to discuss their progress and identify strategies for continuing to improve.

Improvement is not accidental. Musicians and athletes know that simply doing an action repeatedly does not lead to improvement unless they are receiving feedback on their performance and actively working at improving over time. These two organizations know this and behave as if they believe it.

The second thing that stood out about these organizations was that they dedicated significant resources to the work of improvement. Had these resources been used elsewhere, they certainly would have allowed these organizations to do more things. But committing to excellence requires *narrowing the number of things the organization seeks to achieve* to focus intense attention on those things that are most important for the organization’s success. Higher education institutions, particularly large research institutions, are not good at narrowing the number of things they seek to achieve. Consider a typical faculty appointment that expects the person to be an expert researcher, an expert teacher, and highly engaged in community projects. To become an expert in just one of these areas requires years, if not decades, of practice and training, and to develop true expertise in all three is very rare. As another interesting activity, take a look at the number of degree programs offered by your institution, your college, or your department. Can your institution really provide an extraordinary experience to students in 250 different degree programs? Can your department, with its limited resources and small number of faculty, really provide transformative learning experiences across six different doctoral degrees, eight master’s degrees, and three bachelor’s degree areas? Gary Keller and Jay Papasan support this notion of limiting the number of things an organization seeks to achieve, writing “extraordinary results are directly determined by how narrow you can make your focus” (2012, p. 10). One key benefit of narrowing your focus is that it opens up resources that can be made available to support extraordinary results. Committing to excellence means investing significant resources and creating a climate where improving the organization is a systematic, intentional way of being.

The final thing that stood out about these organizations was that they engaged in the review process out of *a deeply held desire to improve, not out of an external demand that they comply*. It should be noted here that both organizations face significant compliance demands (one organization in the health care field, the other in financial services). This points to an advantage of separating improvement and compliance activities from each other. Over the last few decades, accreditation has greatly increased its attention on processes and activities—like assessment—that are primarily improvement activities. Institutions are expected to have certain improvement process and practices in place as a standard of practice. This coupling of improvement and compliance can have unintended consequences. For instance, at one prior institution where I implemented assessment, the institution quickly threw together a strategic plan 18 months before a site visit because the regional accreditor expected to see a strategic plan in place to guide the institution’s actions and improvement efforts. The institution did not take this planning very seriously, and the Provost stated, in an open, public meeting, that we just needed to “make it look like we have a plan,” rather than worrying about developing something that would actually be used to guide the organization for the next five years.

In this way, the close linking of improvement and compliance made them both a sham. Please do not misunderstand here. Compliance with performance standards, reviewed and verified through an accreditation process, is extremely important and valuable. Would you have your appendix removed in a non-accredited clinic? Would you take your grandmother to dinner at a restaurant that does not comply with health code standards for cleanliness and safe storage of food? Of course not. But forcing institutions to pretend to be engaged in improvement for accreditation purposes is a waste of resources and increases the skepticism many faculty and staff have about organizational improvement efforts. Yet, both organizations I examined found value in engaging in a voluntary, improvement-focused, accreditation-like review to help them achieve their goal of excellence. To me this suggests value in continuing to increase the space between compliance and improvement. Imagine what could be accomplished if accreditors didn't have to spend resources on ensuring compliance with a host of government-selected metrics, instead spending these resources partnering with institutions on improvement efforts!

To summarize, what does it mean to commit to excellence? It means to honestly evaluate performance and identify the areas where improvement is needed, it means narrowing focus on a few areas where extraordinary performance is desired rather than attempting to do everything, and it means developing an internally supported drive for excellence. What are the implications of this for assessment? I turn to that question now.

What is the Role of Assessment Leaders in Institutions that are Committed to Excellence?

One aspect of the work of assessment that continues to astonish me is the diversity of backgrounds that assessment leaders have. A recent thread on the ASSESS listserv revealed assessment leaders with backgrounds in psychometrics, in the humanities, in the sciences, and in education. Equally astonishing is the broad variation in role expectations for assessment positions. Some institutions want an assessment cop who will enforce compliance with assessment requirements, other institutions want an assessment servant who will implement whatever initiative that the assessment committee develops, while other institutions want a Jack- or Jill-of-all-trades who can implement assessment along with a range of additional duties. Regardless of the person's background or the current role expectations, there are big differences for the work of assessment in institutions that are committed to excellence.

First, assessment leaders in institutions that are committed to excellence *will pay much less attention to compliance and accreditation*. This is a big change because assessment and accreditation are deeply intertwined with a shared history. The increase in the number of assessment professionals over the last three decades parallels closely with the increasing demand for assessment in accreditation. Raise your hand if you work at an institution that created your position because it was required for successful compliance with accreditation. As much as we work to promote the idea of assessment as improvement, not compliance, we cannot deny the fact that the rise in our profession has historically had much more to do with institutions' need to comply with accreditors' increasing demands for assessment than with institutions' authentic desire for improvement in learning. This is our history, but it will not be – and cannot be – our future because the environment in which we operate is rapidly changing. Accreditors are changing the way they include assessment in their accreditation standards, the renewal of the Higher Education Act will likely bring changes to the role of accreditors, and higher education assessment software providers are offering tools that are changing assessment practices, with some institutions viewing these systems as a cheaper replacement for a highly-trained assessment leader. Assessment leaders must still do whatever it takes to maintain compliance with accreditation standards, but the profession must take action to separate the work of pursuing excellence with compliance with external standards.

Second, assessment leaders in institutions that are committed to excellence will spend much more of their time *creating the environmental conditions that support improvement*. Assessment alone is insufficient to produce the

kinds of fundamental changes that are required for achieving excellence. The writers of the Association for the Assessment of Learning in Higher Education *Principles of Good Practice for Assessing Student Learning* understood this when they wrote “assessment is most likely to lead to improvement when it is part of a larger set of conditions that promote change” (Astin, et al., 1992). Faculty and staff engaged in the work of assessment at institutions that lack these larger set of conditions rightfully become extremely frustrated when their efforts fail to produce the meaningful changes needed to truly pursue excellence. We are lying to ourselves if we believe we can turn a poorly performing academic department into a star performer by asking it to annually submit an “assessment report” that asks it to describe some discussions it had about “assessment data.” Instead, assessment leaders must seek to understand the full range of factors that are required to produce excellence, including leadership, resources dedicated to improvement, attention to processes, and, yes, collecting, analyzing, and using evidence of performance, and then work to create the environment in which needed transformational changes will occur.

Finally, assessment leaders in institutions that are committed to excellence will spend much more of their time *collaborating with faculty, staff and students, and working on a broad range of projects* that relate to student learning and the student experience. Such efforts will take lots of work over a long period of time, because becoming an excellent organization is tremendously hard work. Assessment leaders in institutions that are committed to excellence must help their organization find the discipline needed to persist in efforts to improve for a long time. Just like obesity is not defeated by skipping one doughnut, institutions are improved through the accumulation of numerous, ongoing, systematic efforts to improve.

Committing to Excellence

Committing to excellence should not be taken lightly. Seeking excellence is hard work - it requires seeking out high-quality data and hearing difficult feedback, making changes repeatedly until excellence is achieved, and taking bold, innovative action. But only by committing to excellence will our institutions be meeting our shared mission of preparing the future leaders and citizens of the world. The work of assessment, although it is quite different in these types of institutions, will continue to be at the heart of this important work.

References

- Astin, A. W., Banta, T. W., Cross, K. P., El-Khawas, E., Ewell, P. T., Hutchings, P., Marchese, T. J., McClenney, K. M., Mentowski, M., Miller, M. A., Moran, E. T., & Wright, B. D. (1992). Principles of good practice for assessing student learning. *AAHE Bulletin* 45(4), p. 1-3.
- Baldrige Performance Excellence Program. (2019). 2019-2020 Baldrige Excellence Framework: Proven leadership and management practices for high performance. Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology.
- Keller, G., & Papasan, J. (2012). *The one thing: The surprisingly simple truth behind extraordinary results*. Austin, TX: Bard Press.
- Wolff, R. A. (2005). Accountability and accreditation: Can reforms match increasing demands? In J. D. Burke & Associates (Eds.), *Achieving accountability in higher education: Balancing public, academic, and market demands* (pp. 78-103). San Francisco: Jossey-Bass.

Jeremy Penn is the Assessment Coordinator for the College of Education at University of Iowa. He can be reached at jeremy-penn@uiowa.edu.

Implementing an Effective Meta-Assessment System

By Jeffrey W. Freels and Ximena Canelo-Pino

Abstract: Meta-assessment is widely accepted as a beneficial practice in student learning outcomes assessment, but many higher education institutions face significant challenges in implementing effective meta-assessment systems. The current literature in this area provides helpful tips in certain aspects of meta-assessment, but comprehensive blueprints on meta-assessment implementation are lacking. This paper outlines in detail the steps that one institution took in order to successfully create such a system. This case study describes how the institution conceptualized, implemented, acted upon, and refined its meta-assessment system over a two-year period to advance the institution's assessment practice.

Introduction

Meta-assessment, or “the process of evaluating the quality of assessment practices” (Fulcher, Coleman, & Sundre, 2016, p. 1), facilitates meaningful institutional advancement in student learning outcomes assessment. Whether administered by committees or individuals, meta-assessment can help institutions move from an assessment culture emphasizing compliance to one emphasizing improvement (Fulcher & Orem, 2010). In recent years, consensus has emerged in the higher education learning assessment community that meta-assessment is a worthwhile endeavor. A body of literature informs the development of various practices associated with meta-assessment, and many institutions have posted meta-assessment rubrics online, yet literature on the specific procedures needed to develop a working meta-assessment system is sparse. Published case studies describing the creation of meta-assessment systems have featured processes based in peer review (Bendikas, 2015; Fong Bloom, 2010; Heinerichs, Bernotsky, & Danner, 2015), which is considered a best practice in meta-assessment (Fulcher et al., 2016). Peer review, however, demands considerable resources and may not be replicable at every institution. This paper describes a comprehensive meta-analysis option that does not rely on peer review.

Implementing an effective meta-assessment system can be challenging and time-consuming (Fulcher, Swain, & Orem, 2012). Those challenges are magnified at many higher education institutions by a lack of financial and/or human resources that could help to mitigate them. Many institutions with robust meta-assessment procedures have committees of faculty and/or staff, who sometimes receive stipends or service credit for their participation, to administer the system. In contrast, the meta-assessment system described in this paper was developed by one full-time assessment director and one part-time doctoral student (the “assessment office”). Although it may have been preferable to create a meta-assessment system based on peer review, various constraints made that impractical in this case.

This paper focuses on the first year of what will be an ongoing meta-assessment process. After describing the case study context, the paper proceeds chronologically through the development, implementation, and refinement of that system. It concludes with a brief analysis of the factors that contributed to the success of the project and a discussion of recent changes to it. By the end, it will be clear how this meta-assessment process contributed to significant improvements in one institution's assessment culture and practice, primarily by clarifying institutional assessment requirements and distributing customized actionable feedback to all academic programs.

Meta-Assessment Literature

Meta-assessment, broadly understood as the practice of assessing the assessment process (Ory, 1992), involves a thoughtful examination and systematic evaluation of the assessment process itself. Whereas the purpose of assessment is to evaluate and assess a particular program or service, Peters (2000) suggests that meta-assessment is appropriate for evaluating the “preconditions, limitations, and assumptions” (p. 337) in which the program operates or the service is offered. In doing this, meta-assessment occupies the “conceptual space” between the philosophy of the program and the assessment of it (Peters, 2000, p. 334). Meta-assessment allows for the analysis of the possible relationships between a system and its environment by promoting reflection and examination of

the context and value of the assessment activity itself (McDonald, 2010). By focusing on the context in which assessment data were generated rather than the data itself, meta-assessment provides a framework for guided reflection and practical improvement within programs.

An effective meta-assessment system advances institutional learning outcomes assessment processes as well. In describing the characteristics of good higher education assessment programs, Shavelson (2009) suggests that assessment directors play “an important role in setting the tone and engaging and supporting faculty” (p. 98) and that “aggregated data are pertinent to program improvement and tracking campus-wide progress” (p. 98). Furthermore, meta-assessment processes promote several of Banta and Blaich’s (2011) 17 characteristics of effective learning outcomes assessment programs. Banta and Blaich argue for the assessment of both process and outcomes, the cultivation of an assessment environment that is “receptive, supportive, and enabling” (2011, p. 23) and the ongoing evaluation of the assessment process itself, all of which are consistent with meta-assessment.

Assessment scholars have differentiated between meta-assessment at the program and the institutional levels (Bresciani, Gardner, & Hickmott, 2010; Fulcher et al., 2012; Suskie, 2009). While program level meta-assessment allows practitioners to evaluate the process by which academic programs create and use assessment results, institutional level meta-assessment aims to appraise the degree to which assessment is methodically conducted within an entire institution (Orem, 2012). Whereas program meta-assessment can guide programs towards more effective practice, institutional meta-assessment can provide data to inform senior administrators, satisfy external demands for accountability, and highlight strengths and weaknesses in institutional assessment processes (Fulcher et al., 2012; Fulcher & Good, 2013).

The meta-assessment rubric is arguably the most important component of a meta-assessment system. Where rubrics in general can succinctly communicate performance expectations (Goodrich, 2007), the meta-assessment rubric models best practices and suggests improvements to learning assessments (Fulcher & Good, 2013). For programs, the meta-assessment rubric provides a formative opportunity for positive feedback that can lead to continuous improvement. For institutions, data provided by meta-assessment rubrics offer knowledge about the state and quality of institutional assessment practices (Fulcher et al., 2012; Rodgers et al., 2013). A recent study of meta-assessment rubrics in higher education suggests that virtually all of them measure “objectives/student learning outcomes, methods, results, and use of results” (Fulcher et al., 2012, p. 14), but that they otherwise diverge significantly in terms of the level of detail they contain and how they are applied. Regardless of their particular characteristics, meta-assessment rubrics must be appropriate to the unique context in which they will be applied.

Case Study Context

The institution at which the events described in this paper took place is a large public research university in the southern United States offering more than 400 baccalaureate, master’s, professional, and doctoral degrees. For a variety of reasons, student learning outcomes in those programs are assessed through approximately 280 assessment plans (the exact number fluctuates from year to year), as the various degree options in a single discipline (e.g., disciplines that offer both a bachelor of arts and a bachelor of science) have been consolidated into one assessment plan. A central assessment office under the university’s chief academic officer oversees and coordinates campus-wide assessment with the assistance of faculty members, staff, and administrators throughout the university. The institution uses a web-based assessment management system (AMS) to manage reporting and documentation. All academic programs are required to use the AMS to report annually on their student learning outcome assessment efforts.

Significantly, the institution at which this case study took place underwent reaffirmation of regional accreditation in 2017-2018. The meta-assessment system described here featured prominently in the institution’s self-study report to the accreditation agency. The off-site reaffirmation committee reviewed that report, noted the meta-assessment system, and deemed the institution to be in compliance with the student learning outcomes assessment standard. Moreover, the on-site committee issued no recommendations to the institution at the conclusion of the site visit.

The current university assessment process requires that academic programs formulate local mission statements that align with the institution's mission. Program faculty then refine their local mission statements by identifying goals that articulate what they are trying to achieve. Program goals inform the development of program student learning outcomes, or specific, realistic, and measurable statements of the knowledge or skills students will be able to demonstrate as a result of their studies. In formulating program-level student learning outcomes, academic program faculty members are encouraged to identify those proficiencies that most represent their discipline and serve as hallmarks of the program.

Each fall semester, academic programs are expected to update five components of their assessment reports in the AMS: the (1) *goals* and (2) *outcomes* they will assess that year, along with the (3) *assessment methods* through which they will assess them; at the same time, they must report (4) *findings*, or the results of those assessments, from the previous academic year and (5) *next steps*, or how they used or plan to use those results to improve the program. Thus, in the fall of 2016, academic programs submitted their 2015-2016 assessment findings and next steps and their 2016-2017 goals, outcomes, and methods.

Building a Meta-Assessment System

In order to promote high-quality assessment practices and provide feedback to faculty and staff on their assessment plans and reports, the university assessment office initiated a meta-assessment of all 280 academic program assessment reports in the fall of 2016. The focus of the meta-assessment reviews were the 2015-2016 assessment reports from academic programs (findings and next steps) and the 2016-2017 assessment plans (goals, outcomes, and assessment methods), collectively referred to here as "assessment reports." The two necessary precursors to establishing an effective meta-assessment system in this case were: refining the institution's requirements for student learning outcomes assessment and creating an appropriate meta-assessment rubric. The review process also needed to be efficient since all reviews were to be completed by one full-time staff member and one part-time graduate student assistant.

The assessment office reviewed documents from an extensive 2015 reform of the university assessment process in order to identify minimum reporting requirements. Crafted by a university-wide institutional effectiveness committee, those documents include statements outlining the principles on which the university's assessment process is based and guidelines for recommended assessment practices (see Appendix A). After consulting with other university personnel involved in assessment, elements from the principles and guidelines were operationalized into a list of 10 clear and brief institutional expectations for student learning outcomes assessment (see Appendix B). The goal of the assessment office in establishing this list was threefold: (1) promote increased awareness of institutional student learning outcomes assessment expectations, (2) set clear minimum baselines for meta-assessment evaluation of program assessment reports, and (3) provide space and flexibility for academic programs to create assessment plans appropriate to their unique needs and contexts. With minimum reporting requirements established, the next step was to formulate an appropriate meta-assessment rubric.

In this case, a simple and straightforward rubric was essential to the success of the meta-assessment project. Given the large numbers of academic programs at the university, having a succinct, one-page rubric would make it possible to complete and distribute quickly and widely. Furthermore, it needed to be easily comprehensible to personnel without detailed knowledge of typical assessment terminology, reflect institutional expectations for assessment, and promote recommended assessment practices. Although the university had a comprehensive assessment report rubric prior to 2016, the level of detail it contained was not conducive to the timely review of a large volume of program assessment reports by the two people (1.5 FTE) completing the reviews.

In order to adapt the previous institutional rubric and learn from existing examples, the assessment office scanned the internet for publicly available meta-assessment rubrics. Rubrics from Boston University, James Madison University, Sam Houston State University, and West Texas A&M University were particularly helpful in the development process. A draft of an academic program assessment report rubric appropriate to the university's needs was then created with those examples in mind. That draft was shared with select personnel at the university whose work encompasses assessment of academic programs or student support units. Their feedback informed the

final version of the rubric on the academic program assessment report evaluation form (see Appendix C), which became the basis for reviews of all 280 academic program assessment reports at the university.

Process Implementation

Using the meta-assessment rubric, the five components of each program assessment report (i.e., goals, outcomes, and assessment methods for 2016-2017, findings and next steps from 2015-2016) were evaluated from October 2016 to January 2017 on two broad criteria: (1) whether the report met the 10 institutional expectations for program assessment, and (2) the degree to which the report was consistent with recommended assessment practices. Depending on the degree to which it met those broad criteria, each component of the report was rated as *Exemplary*, *Acceptable*, *Emergent*, or *Absent*. Components rated as *Exemplary* exceeded institutional expectations and were thoroughly consistent with recommended assessment practices. Those rated *Acceptable* met institutional expectations and were minimally consistent with recommended assessment practices. *Emergent* components met institutional expectations but were not consistent with recommended practices. Components that did not meet institutional expectations or were not included in the report were rated *Absent*. Programs received ratings of *Acceptable* or *Absent* for program mission depending on whether their mission statement was loaded in the AMS.

Based on the cumulative ratings it received on all five assessment report components, each program received an overall evaluation that indicated the degree to which its assessment report met institutional expectations and/or had components that could be improved. Reports that received ratings of *Emergent* or higher on all five components were deemed to be in compliance with institutional expectations. Reports on which at least one component was rated as *Absent* were deemed out of compliance with institutional expectations. Rubrics were clearly marked to indicate whether the assessment office suggested minor changes (in cases where the program report was compliant with institutional expectations, but could be improved) or requested major changes (where reports were not compliant). The result in each case was a two-page "feedback form," with a completed rubric on page one and detailed comments and suggestions for improvement on page two. Comments and suggestions identified specific report components in need of revision and included targeted recommendations for improvement.

In order to facilitate the efficient review of academic program assessment reports, the assessment office designed a reusable online form using a web-based survey builder. Construction of the online form proceeded using a Cartesian reductionist framework. Described in Part Two of *Discourse on the Method*, Descartes (2000) explains how complex ideas can be deconstructed into their simplest and most easily known objects in order to facilitate greater understanding of a larger object or idea. Through this method of analysis, he argues that all the things that human beings can know are inter-deducible. Mulaik (2004) describes Cartesian reductionism as a "method of analysis and synthesis" (p. 2). The various elements that make up a complex idea are separately identified and defined according to their most fundamental truths. Once established, those truths become the basis of a reconstruction process in which the larger, complex idea may be better understood. Through this method, complex or nebulous ideas become discernible as the sum of various well-defined truths.

In this case, the larger object under examination was the academic program assessment report. Of necessity, reports at the university are already broken down into five essential components. In turn, each of those components is made up of more discrete characteristics, which were identified on the previously-developed assessment report rubric. For example, the rubric defines an exemplary outcome as one that is specific, measurable, aligned with the unit's mission, and likely to yield meaningful information about the program. But what does it mean for an outcome to be specific or measurable? Using a Cartesian reductionist framework, the assessment staff explored the fundamental truths of the characteristics on the rubric, discussing the meaning and reaching a mutually agreed-upon definition of each. In this case, the reviewers agreed that specific outcomes articulate precise actions (usually denoted by active verbs in outcome statements) that students will be able to undertake or complete in order to demonstrate attainment of a clearly identifiable learning object (e.g., a skill, competency, or knowledge domain). They further defined measurable outcomes as those that are feasible to

assess, in a specific period of time or under a specific set of circumstances, and evaluated using clearly defined standards.

Subsequently, reviewers conducted preliminary test reviews of approximately 20 academic program assessment reports for inter-rater reliability. Once complete, they analyzed the results of the test reviews and met to harmonize their application of the ideas imbued in the rubric. The language and structure of the reusable online review form were finalized during that norming session to ensure consistent understanding and application. With that mutual understanding in place, reviews of all 280 program assessment reports commenced using the reusable online form.

Altogether, it took the two reviewers (1.5 FTE) approximately 169 hours over four months to complete and deliver all 280 academic program assessment report reviews, or about 36 minutes per report review. During that time, the reviewers met at least weekly to calibrate their application of the rubric and interpretation of institutional expectations. Filling out feedback forms for each academic program, which included completing a rubric and writing comments and suggestions for improvement, was among the most time-consuming components of the process. To streamline that work, the assessment office collected comments on commonly recurring problems, saved and standardized them in a plain text editor, and copied-and-pasted into the feedback forms as appropriate. Some of the most commonly recurring problems included assessment reports that had not been updated to reflect plans for assessment in the current academic year, programs that failed to identify any direct methods of assessing student learning, and absent, insufficiently articulated, or weak next steps.

Despite that standardization, almost all programs received unique feedback specific to their assessment reports. Out of 855 total comments and suggestions on all feedback forms (about three comments per program, although some comments were more detailed than others; the median comment length was 46 words), 302 were unique, 220 were standardized duplicates, and 333 were hybrids. Comments were carefully designed in order to maximize their utility and promote positive engagement with the university's assessment process. The strategy for writing feedback was grounded in guidelines recommended by Quality Matters (QM) and Penn (2012)¹.

Of the five feedback delivery strategies recommended by Penn (2012), three were particularly germane to the case described here. Specifically, Penn suggests that feedback should: focus on the work of assessment rather than the people doing the assessment, "provide specific comments on errors and suggestions for improvement" (2012, p. 9), encourage incremental change over time, and clearly communicate expectations. The synthesis of the QM guidelines with Penn's recommendations determined the feedback approach in this case. The list of 10 institutional expectations for outcomes assessment set clear minimum baselines for all components of the assessment process. As often as possible, the feedback drew from the language of those expectations (e.g., "Institutional expectations require that academic programs assess at least one student learning related goal and its corresponding outcomes annually, but this program has not reported any findings for 2015-2016."). Furthermore, the reviewers avoided commenting on minor errors in the assessment reports, instead focusing on the most critical problems (e.g., the program did not assess student learning or did not assess learning with a direct measure). Comments directly referenced problematic sections of the assessment reports, explained why those sections did not meet institutional expectations, included focused recommendations for revisions, and highlighted positive aspects of the reports. Assessment staff designed the feedback to be clear, concise, and actionable.

Completed feedback forms for all programs in their areas were emailed to administrators and/or staff (typically associate deans and their assistants) who manage assessment in each of the university's colleges in January 2017.

¹ QM is an organization that facilitates quality assurance in online education by administering peer reviews of online college courses. Prior to reviewing courses, QM peer reviewers are trained to provide feedback that is constructive, specific, measurable, sensitive, and balanced. Whenever possible, QM reviewers are instructed to align their recommendations to the language of the QM standards, cite specific examples from the course to contextualize the recommendation, and avoid the use of personal pronouns in framing their comments. Penn (2012) echoes the QM guidelines, although the strategies he outlines are specific to the field of meta-assessment.

Those emails requested that administrators distribute the feedback forms to their academic programs and convey a June 1 deadline for final revisions to academic program assessment reports. Most of the administrators confirmed that they had forwarded the feedback forms to the appropriate personnel by mid-February 2017.

Benefits and Impact of the Process

When assessment staff completed the academic program assessment report reviews in January 2017, they found that 45% of programs had submitted assessment reports that met minimum institutional expectations and 55% had submitted reports that did not meet expectations. From January to June 2017, assessment staff communicated with hundreds of academic program faculty and staff in a variety of ways: email, phone, and in-person consultations were frequent, along with a series of professional development events. In numerous instances, those faculty and staff expressed appreciation for the assessment report feedback and confirmed that the feedback from comments was useful and actionable. By mid-June 2017, 88% of assessment reports met expectations and 12% did not. In addition to the feedback forms themselves, several activities fueled by the meta-assessment system implementation helped to bring about that improvement.

Prior to the meta-assessment system implementation, program faculty and staff often asked to see examples of good assessment practice. Therefore, during the meta-assessment review of academic program assessment reports, the assessment office selected five examples of exemplary practice in each of the five assessment report component areas, obtained permission from program faculty and staff to publish those examples, and made them available to the campus community. In two weeks after the exemplars were published on the assessment office's learning management system website, traffic to the website tripled over typical levels. The assessment exemplars continue to be downloaded regularly, indicating that they remain a critical resource for university personnel.

In addition to permitting the publication of exemplars, data collected during the assessment report review process have been used in multiple ways. Quantitative data were useful in revealing the components of the assessment process in most need of overall improvement. For instance, 84% of academic program assessment reports had *Exemplary* or *Acceptable* outcomes, suggesting that the university's programs identify high-quality program outcomes. By contrast, only 23% of reports had *Exemplary* or *Acceptable* next steps. In response to those data, the assessment office organized a series of professional development events in April 2017 on using assessment results to formulate next steps. A panel discussion featuring faculty and staff from programs with high-quality assessment findings and next steps drew close to 150 viewers either in-person or via a live stream of the event. Soon after, the assessment office offered two workshops on formulating next steps, both filled to capacity.

Acting on the Meta-Assessment Data

The assessment office obtained qualitative data about the university's academic program assessment efforts by collecting and analyzing all of the comments and suggestions written by the assessment office reviewers on the assessment report feedback forms. The qualitative data analysis proceeded according to a grounded theory framework, in that empirical data were reviewed in context using a constant comparative method in order to develop and verify hypotheses about the data as they emerged (Charmaz, 2006). This approach permitted of several benefits.

Significantly, the grounded theory approach highlighted four categories of errors in assessment reports.

- *Clerical* errors occurred when the report contained simple mistakes, such as when an indirect assessment method was mistakenly labeled as a direct assessment method or when a finding was incorrectly placed in the AMS.
- *Articulation* errors occurred when components of the assessment report were improperly framed, such as when programs wrote outcomes that read like assessment methods (e.g., "Students will be able to score 70% or above on an exam.") or when outcomes encompassed numerous discrete skills or competencies.

- *Incompletion* errors occurred when a program failed to report on an assessment report component or reported insufficient information (e.g., they identified an assessment method, but did not describe the method in any meaningful way).
- *Non-compliance* errors occurred when components of the assessment report directly violated one of the 10 institutional expectations for outcomes assessment, such as when a program assessed student learning exclusively through indirect methods or reported students' overall course grades as an assessment measure.

Reporting error category data of this kind can inform professional development planning. For example, personnel from programs with numerous clerical reporting errors could benefit from additional training in the use of the university's AMS, while personnel from programs with numerous non-compliance errors might need to be reminded of institutional expectations for outcomes assessment. This kind of targeted professional development could help to optimize the allocation of assessment resources to areas in most need of improvement.

The collection and analysis of feedback form comments also allowed assessment staff to develop standardized comments as the reviewing process progressed, as mentioned above. Saved and categorized by the assessment report component at which they were directed, standardized comments were composed according to the QM and Penn (2012) feedback guidelines and copied-and-pasted into feedback forms as appropriate. Where necessary, language was added to standardized comments to highlight specific assessment report components that did not meet expectations. In all cases, comments pointed directly to problematic portions of the report, explained the problem in the context of institutional expectations, and suggested possible remedies for addressing them. For example, one standardized comment on program report next steps reads:

Institutional expectations require that every program identify at least one strong improvement action per year in its Next Steps. It's not clear that any of this program's 2015-2016 next steps qualify as strong improvement actions. Next Steps should outline a clear course of action, preferably with a timeline and an explanation of the person or group that is responsible for taking the action. It may also be appropriate to describe how the proposed next steps will improve the program. Please review it.

Standardization of comments saved time in that commonly occurring assessment report issues could be quickly and efficiently commented upon in the feedback forms. The use of standardized comments further ensured that programs received consistent messages about their assessment efforts.

Finally, the feedback form comments from year one of the meta-assessment system informed the evolution of that system in year two in two ways. First, the reusable online review form underwent an extensive revision in year two based on the data generated in year one. The year two review form was more detailed than the year one form and allowed for the collection of more data on program assessment reports, while decreasing the amount of time it took to complete the reviews. Second, the pool of standardized comments was improved and expanded in year two based on analysis of year one reviews (see Appendix D for examples of year two standardized comments). Some comments were revised to convey a more positive tone, some were lengthened to provide more detailed suggestions for improvement, and new comments were written to address less frequent, but still regularly occurring report deficiencies. This further simplified the process of creating feedback forms. These changes further streamlined the assessment report review process in year two of its implementation.

Analysis

It is not possible at this time to determine whether the efforts described above led to improvements in student learning, but somewhat clearer is the impact the process had on faculty and staff involved in academic program assessment. Many of them reported that the list of 10 institutional expectations for outcomes assessment better clarified the basic requirements of the university's assessment process. They also told assessment staff that the feedback forms greatly assisted their work by outlining in detail the steps that programs needed to take to improve their assessment efforts and reports. Finally, ongoing support from the assessment office, through professional

development and consultations with program faculty and staff, has contributed to an improved culture of assessment at the university.

The meta-assessment project described here is the sum of many tasks that went into creating it and each was essential to its success. The list of 10 institutional expectations set baseline expectations for each of the five components of the assessment reports. The rubric organized those expectations into a readable format and modeled recommended assessment practices for programs seeking improvement. The assessment office's rubric norming ensured consistent application of it through hundreds of assessment report reviews. The web-based survey builder facilitated the efficient review of those reports, tracked the amount of time it took to review them, and allowed for the collection of voluminous amounts of data on the university's program assessment efforts. Review data were organized and analyzed in a spreadsheet exported from the survey builder. That spreadsheet facilitated the efficient completion of feedback forms and served as a repository for all feedback form comments and suggestions. Frequently recurring comments and suggestions were standardized in a plain text editor and copied-and-pasted into feedback forms when appropriate in order to reduce the amount of time spent writing comments and suggestions. Data obtained throughout this process fueled the design of a professional development curriculum and informed changes to the meta-assessment system in the second year of its implementation. Better use of technology in year two led to the collection of more data on the university's assessment efforts in a shorter amount of time.

Process Evolution

The meta-assessment system described here continues to evolve. Significantly, the assessment office revised the meta-assessment rubric for year two by adding an *Inadequate* performance level in between *Absent* and *Emergent*. This allowed reviewers to more precisely distinguish between programs that did not report any information for particular assessment report components and those that reported information that did not meet minimum institutional expectations.

Another major change involved efforts to decrease the amount of time it took to complete and distribute assessment report reviews in year two. The available data suggest that was successful. In year one, it took one full-time staff person and one part-time graduate student approximately 169 hours to review all of the program assessment reports, fill out the feedback forms, and distribute them to program faculty and staff. The same two people implemented a revised meta-assessment system in year two (fall of 2017) and completed it in 126 hours, or 43 fewer hours than year one. More precisely, it took 13 fewer hours to review the assessment reports through the reusable online form and 30 fewer hours to fill out the feedback forms and distribute them.

Several factors impacted the decreased time on task. First, in year two, the reusable online form was modified to be easier to use but more comprehensive in terms of the data it collected. Second, feedback form comments and suggestions from year one was revised for increased precision and clarity and more of them were used in year two than in year one. Third, the exported spreadsheet from the survey builder in year two was programmed to automatically generate standardized comments where appropriate so they could be copied-and-pasted into the feedback forms with minimal editing. Fourth, the personnel completing the reviews had more experience in implementing the system, thus could work more quickly. Finally, increased awareness among university faculty and staff of institutional expectations and recommended assessment practices likely led to better reporting in year two. While it is impossible to determine which of the process variables contributed most significantly to the improvement, it is clear that the process was more efficient and less time-consuming in year two.

Year two also saw an improvement in terms of programs meeting institutional expectations. At the conclusion of the meta-assessment review process in year one, 45% of academic programs had submitted assessment reports that met institutional expectations. At the same point in year two, 50% of academic programs had submitted

assessment reports that met institutional expectations². Whereas only 53% of programs had next steps that met institutional expectations in year one, 73% of programs had next steps that did so in year two. Additionally, programs were permitted to revise their assessment reports after receiving the assessment office's feedback in both years. By the conclusion of the annual assessment cycle in June 2017, 88% of academic programs had assessment reports that aligned with institutional expectations; by June 2018, 93% of programs had assessment reports that aligned with expectations. Between the heightened efficiency of the meta-assessment process, targeted professional development from the assessment office, and clear improvements in initial academic program assessment reporting, year two of the meta-assessment system could be credibly viewed as a success.

Moving forward, the assessment office will address a number of challenges in the years to come. Perhaps most importantly, longitudinal exploration of connections between the university assessment process and the student experience will help to better inform the university's continuous improvement efforts. Also, the university's current assessment cycle culminates in the 2019-2020 academic year and the assessment office has begun planning for institutional assessment beyond 2020. As part of that planning, an institution-wide faculty/staff committee on assessment has convened to consider future changes to the institutional assessment process. In addition to considering changes in the overall process, this committee will review the current meta-assessment system and recommend changes for its improvement. The refinement and improvement of this meta-assessment system is expected to continue for the foreseeable future.

Further Discussion and Future Directions

As indicated above, meta-assessment through peer review is considered a best practice in the field. The system described here is not based in peer review, but its implementation has nonetheless been advantageous to the institution at which it was created. This system offers multiple benefits to institutions looking for alternative methods of meta-assessment. The most important of those benefits is related to cost. Where Fulcher et al. (2012) estimate that their peer review-based meta-assessment system takes numerous personnel approximately 600 hours per year to administer, the system described here was created, implemented, and completed by two people (1.5 FTE) in 169 hours in year one and 126 hours in year two.

Another benefit of the system is related to the question of reliability. Two people created and implemented the system described in this paper. In addition to making an effort to ensure inter-rater reliability before the initial meta-assessment reviews took place, the reviewers worked to ensure continued fidelity throughout the process. While the benefits of peer review are evident, the increased number of people it involves also increases the challenges related to meta-assessment consistency and reliability. Those challenges were not non-existent in this case, but they are perhaps lessened by the smaller number of reviewers.

Finally, the data generated by this meta-assessment project has been considerable. The assessment office in this case has detailed data on the components of the assessment process in most need of development, the most common types of reporting errors that programs make at this particular institution, which programs make which types of errors, and, over time, how the institution's various programs have progressed in their assessment practice. The identification and publication of exemplary assessment practices from around the university was also made possible by this initiative. These data have been useful in external accountability reporting, informing future revisions of the institutional assessment process, planning for professional development, and enriching the institution's culture of assessment.

Assessment scholars and practitioners seeking to build on the research presented here should explore the connection between meta-assessment and learning improvements within academic programs. In particular, qualitative research into the connection between meta-assessment, program improvements, and learning

² The percentage of programs that met expectations in year two would have been higher except for the fact that a sizable number of programs (about 14%) made clerical errors in submitting their assessment plans for 2017-2018 (for several possible reasons, clerical errors were more common in year two than year one).

enhancement would be a valuable contribution to the field. Despite the ubiquity of graphics depicting learning assessment processes as systematic and circular, those processes often promote improvement that is reached through less than systematic means. In one such possible scenario, faculty in a program meet to discuss the meta-assessment review of their program assessment report and the results of a learning assessment. In that meeting, they conclude that their curriculum could be better aligned to the program objectives and formulate an action plan to achieve that alignment. In such a scenario, program improvement occurs within the context of the institution's assessment/meta-assessment process, but independent of the actual data that resulted from a student learning assessment. Quantitative research on the connection between assessment and learning improvement might fail to register that manner of change, but qualitative research could illuminate it powerfully. Stories about the qualitative improvement of academic programs through the assessment/meta-assessment process could be persuasive in helping institutions move further towards an assessment culture emphasizing improvements as opposed to one emphasizing compliance.

References

- Banta, T. W., and Blach, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning*, 43(1), 22–27. <https://doi.org/10.1080/00091383.2011.538642>
- Bendikas, K. (2015). Implementing a peer review of assessment reports. *Assessment Update*, 27(5), 5–12. <https://doi.org/10.1002/au.30034>
- Bresciani, M. J., Gardner, M. M., & Hickmott, J. (2010). *Demonstrating student success: A practical guide to outcomes-based assessment of learning and development in student affairs*. Stylus Publishing.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis* (1st ed.). SAGE Publications.
- Descartes, R. (2000). *Discourse on method and related writings* (D. Clarke, Trans.). Penguin Classics. (Original work published 1637).
- Fong Bloom, M. (2010). Peer review of program assessment efforts: One strategy, multiple Gains. *Assessment Update*, 22(5): 5–7. <https://doi.org/10.1002/au.225>
- Fulcher, K. H., Coleman, C. M., & Sundre, D. L. (2016). Twelve tips: Building high-quality assessment through peer review. *Assessment Update*, 28(4), 1–2, 4–16. <https://doi.org/10.1002/au.30062>
- Fulcher, K., & Good, M. (2013, November). The surprisingly useful practice of meta-assessment. National Institute for Learning Outcomes Assessment Viewpoint. <https://www.learningoutcomesassessment.org/wp-content/uploads/2019/08/Viewpoint-FulcherGood.pdf>
- Fulcher, K. H., & Orem, C. D. (2010). Evolving from quantity to quality: A new yardstick for assessment. *Research and Practice in Assessment*, 5(Winter), 13–17. <https://www.rpajournal.com/evolving-from-quantity-to-quality-a-new-yardstick-for-assessment/>
- Fulcher, K. H., Swain, M., & Orem, C. (2012). Expectations for assessment reports: A descriptive analysis. *Assessment Update*, 24(1), 1–2, 14–16. <https://doi.org/doi:10.1002/au>
- Heinerichs, S., Bernotsky, R. L., & Rieser Danner, L. (2015). Guiding principles to impact an institution-wide assessment initiative. *Research and Practice in Assessment*, 10(Summer), 60–64. <http://www.rpajournal.com/guiding-principles-to-impact-an-institution-wide-assessment-initiative/>
- McDonald, B. (2010). Improving learning through meta assessment. *Active Learning in Higher Education*, 11(2), 119–29. <https://doi.org/10.1177/1469787410365651>
- Mulaik, S. A. (2004). Objectivity in science and structural equation modeling. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (1st ed., pp. 426–47). SAGE Publications.
- Orem, C. D. (2012). Demonstrating validity evidence of meta-assessment scores using generalizability theory [Doctoral dissertation, James Madison University]. ProQuest Dissertations & Theses Global.
- Ory, J. C. (1992). Meta-assessment: Evaluating assessment activities. *Research in Higher Education*, 33(4), 467–81. <https://doi.org/https://doi.org/10.1007/BF00973767>
- Penn, J. D. (2012). Assessing assessment: Strategies for providing feedback on assessment practices. *Assessment Update*, 24(6), 8–9, 13.

- Peters, T. A. (2000). Current opportunities for the effective meta-assessment of online reference services. *Library Trends*, 49(2), 16.
- Rodgers, M., Grays, M. P., Fulcher, K. H., & Jurich, D. P. (2013). Improving academic program assessment: A mixed methods study. *Innovative Higher Education*, 38(5), 383–95. <https://doi.org/10.1007/s10755-012-9245-9>
- Shavelson, R. J. (2009). *Measuring college learning responsibly: Accountability in a new era*. Stanford University Press.
- Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.) Jossey-Bass.

Dr. Jeffrey W. Freels is the Director of Academic Policy & Compliance at the University of Texas at Austin. He can be reached at jeff.freels@austin.utexas.edu.

Ximena Canelo-Pino is a Doctoral candidate in Higher Education Leadership at the University of Texas at Austin. She can be reached at ximena.canelo@austin.utexas.edu.

Appendices

- Appendix A: Principles and Guidelines
- Appendix B: Institutional Expectations for Assessment
- Appendix C: 2017-2018 Academic Program Assessment Report
- Appendix D: Examples of a Year Two Standardized Comments

Appendix A: Principles and Guidelines

ASSESSMENT PRINCIPLES

The university's assessment principles are the fundamental tenets that guide our university's actions, expectations, conduct, and practice around assessment.

The university values assessment and its use across all levels of the institution.

The institution undertakes assessment in order to inquire into and improve student learning and organizational effectiveness. Assessment helps the university determine how well it is achieving its mission; it both drives and conveys institutional and program priorities and goals.

Driven by utility and practicality, assessment is applied research that involves gathering information for decision making, action, and ongoing improvement. The institution supports all findings, even those that show lower than expected performance.

Assessment tells a story about educational experience and organizational performance. It cultivates a sense of achievement across campus and complements other ongoing academic review and strategic planning processes.

The institution's assessment processes are well-planned, collaborative, transparent, embedded, adaptive, sustainable, context-sensitive, and evolving.

Engaging in assessment helps the university demonstrate its commitment to excellence while fulfilling accreditation and other accountability requirements.

Student learning assessment is integral to the teaching-learning cycle. It is driven by curiosity about student learning: what works and why, what doesn't work and why, and what can enrich the educational experience and facilitate student learning and development over time. Similarly, organizational assessment is deeply ingrained in the operational cycle of a unit. It focuses on personnel and organizational development and learning and seeks to improve services, efficiency, and effectiveness.

Student understanding of their learning outcomes fosters their comprehension and ownership of their education. Similarly, faculty, administrators, and staff, with a shared understanding of their unit and program goals, strengthen assessment's role in advancing institutional excellence.

Assessment is distinct from performance evaluation or faculty review; it is not used to judge individuals, but serves to guide program level improvement.

ASSESSMENT GUIDELINES

The university's guidelines are suggested courses of action to consider as we engage in assessment.

Institutional Level

1. The institution has a system in place to support program and general education assessment.
2. The institution establishes clear expectations and processes and offers responsive and periodic training to support program assessment.

Program Level

1. Faculty/staff who have primary responsibility for the program are engaged in assessment: planning, implementing, analyzing, and decision making. They receive periodic training in the theory and practice of assessment.

2. Faculty/staff have shared understanding of goals for student learning and institutional effectiveness. They regularly discuss departmental mission, goals, and outcomes and make them public.
3. Programs encourage ownership and collaboration among primary stakeholders, including faculty, staff, administrators, students, and, where appropriate, employers.
4. Faculty and staff construct curriculum maps, logic models, and/or strategic plans to identify significant elements of their programs and the relationships among goals, outcomes, and activities (strategies).
5. Programs use simple yet meaningful methods to generate useful and usable information for decision making. They identify and embed iterative assessment practices.
6. Programs have a process in place to ensure identified changes are made, disseminated, and studied, i.e., assessment efforts result in action on curricula, pedagogy, operations, processes, or services.
7. Programs continually reflect on and revise assessment processes, including the transparency, use and dissemination of results and the methods of collaborative decision-making.
8. Programs actively communicate their outcomes, practices, and findings with colleagues and across campus.

Appendix B: Institutional Expectations for Assessment

Institutional Expectations for Assessment

1. Academic programs and non-academic units must upload yearly assessment plans (goals, outcomes, and methods) into TracDat by October 1.
2. Programs and units must upload yearly assessment results (findings and next steps) into TracDat by the following October 1.
 - Thus, 2017-2018 assessment plans should be loaded into TracDat by October 1, 2017.
 - 2017-2018 assessment results should be loaded into TracDat by October 1, 2018.
3. Every program and unit must identify at least one goal and one outcome.
4. Outcomes must be measurable. Goals do not have to be.
5. Academic programs must assess at least one student learning related goal and its corresponding outcomes annually.
6. Academic programs must assess at least one outcome per year using at least one direct method of assessment.
7. Assessment methods must relate directly to the goal/outcome being assessed.
8. A results-oriented target must be identified for every assessment method.
9. A strategy for achieving the outcome must be identified for every assessment method.
10. Every program/unit must identify at least one strong improvement action annually. Improvement actions are identified in "Next Steps" in TracDat.

Assessment Unit:

Date:

Plan/Report Overall Evaluation
<input type="checkbox"/> No changes needed (all fields are "Acceptable" or "Exemplary")
<input type="checkbox"/> Minor revisions suggested (one or more fields is "Emergent," none are "Inadequate" or "Absent")
<input type="checkbox"/> Major revisions requested (one or more fields is "Inadequate" or "Absent")

	ABSENT	INADEQUATE	EMERGENT	ACCEPTABLE	EXEMPLARY
		Information is present, but does not meet minimum institutional reporting requirements	Meets minimum institutional reporting requirements, but is not consistent with effective assessment practices	Meets institutional reporting requirements; is minimally consistent with effective assessment practices	Exceeds institutional reporting requirements; is robust and thoroughly consistent with effective assessment practices
Program Mission	<input type="checkbox"/> No mission statement.	-	-	<input type="checkbox"/> Program purpose is clearly described.	-
Goals	<input type="checkbox"/> No goals provided.	<input type="checkbox"/> Plan includes at least one goal, but it is not focused on student learning and/or it is insufficiently or improperly articulated.	<input type="checkbox"/> Plan includes at least one goal, but it may not clearly align with the program mission.	<input type="checkbox"/> There is at least one goal and it is clearly aligned with the program mission.	<input type="checkbox"/> A well-organized list of goals (3-5 is ideal) clearly aligns with program mission. Goals encompass the scope of the mission but are manageable to evaluate and assess.
Outcomes	<input type="checkbox"/> No outcomes provided.	<input type="checkbox"/> Plan includes at least one outcome, but it is not focused on student learning and/or it is insufficiently or improperly articulated.	<input type="checkbox"/> Plan includes at least one outcome, but it may not be measurable and/or it may not align with a goal.	<input type="checkbox"/> There is at least one measurable outcome. It is clearly aligned with and mapped to a goal.	<input type="checkbox"/> There is at least one specific and measurable outcome for each goal. Each clearly aligns with the mission and a goal. Assessment of the outcomes is likely to yield meaningful information about the program.
Assessment Methods	<input type="checkbox"/> No assessment methods provided.	<input type="checkbox"/> Plan identifies at least one assessment method, but may not: (1) assess student learning, (2) include a target and/or a strategy to achieve outcome, and/or (3) be appropriate for outcome.	<input type="checkbox"/> Plan identifies at least one direct assessment method. Plan may lack a well-defined target. Method may not fit the outcome it is designed to assess.	<input type="checkbox"/> At least one direct method of assessment, with a specific target, is identified and described. The method fits the outcome it is designed to assess.	<input type="checkbox"/> Plan identifies multiple direct and indirect methods of assessment, each with a specific, reasonable target. Methods are feasible, purposeful, described with sufficient detail, and fit the outcome they are designed to assess.
Findings	<input type="checkbox"/> No findings provided.	<input type="checkbox"/> Findings are inadequately described. No analysis was provided.	<input type="checkbox"/> Report presents some findings, though they may lack sufficient detail. Analysis of findings may be incomplete. Report may or may not provide supporting documentation.	<input type="checkbox"/> Report adequately describes findings. Analysis of findings is logical and stems from evidence produced by the method. Justification for target conclusion (met, not met, etc.) is provided. Supporting documentation (rubrics, surveys, etc.) is provided.	<input type="checkbox"/> Report thoroughly describes findings. Analysis of findings is logical, stems directly from the evidence produced by the methods, and aligns with language of the corresponding target. Justification for target conclusion (met, not met, etc.) is logical. Supporting documentation (rubrics, surveys, etc.) is provided.
Next Steps	<input type="checkbox"/> No next steps provided.	<input type="checkbox"/> Next steps do not specify any changes that will take place as a result of the program's assessments.	<input type="checkbox"/> Report provides next steps, but they may be vague or fail to specify a clear course of action. It may not be clear how the next steps will improve the program.	<input type="checkbox"/> Next steps specify a concrete course of action that is informed by the assessment data OR designed to improve the program in some way.	<input type="checkbox"/> Next steps specify a realistic and concrete course of action that is informed by the assessment data and designed to improve the program in a specific way. Timeframe for implementation is stated and personnel responsible for implementation are identified.

Concerns to Address:

(Items in **boldface** need to be addressed in order to bring the report in line with institutional expectations.)

Overall Evaluation:

Appendix D: Examples of a Year Two Standardized Comments

PROGRAM MISSION

- Please upload the program's mission statement to the assessment management system. All program assessment flows from the mission. Without it, it's impossible to determine if the program is truly assessing that which it values most.

OUTCOMES

- The program assessment plan has one goal and one outcome, which meets minimum institutional expectations for outcomes assessment but is not a recommended practice. You might consider adding some goals and outcomes in the future in order to get a wider glimpse of student achievement in the program.
- There are a lot of outcomes in this plan. That's fine, but we are concerned that the assessment of so many outcomes will be burdensome for the program faculty. Please keep in mind that you don't necessarily have to assess all of these outcomes every year. Institutional guidelines merely require that each outcome be assessed twice in the period from 2015-2019.
- This assessment plan/report does not indicate what is going to be assessed in [current academic year]. Please revise it as soon as possible.
- The outcomes in this assessment plan are not framed in accordance with recommended assessment practice. Outcomes should be framed in terms of the skills, knowledge, or competencies that students will develop in the course of the program. The outcomes as currently written are framed in terms of specific products that students will produce; the work products are assessment methods, not outcomes.

ASSESSMENT METHODS

- Institutional expectations require that academic programs assess at least one student learning-related goal and its corresponding outcomes annually using at least one direct method of assessment.
- Direct methods of assessment are those in which students' work is evaluated directly by an expert (i.e., faculty).
- The assessment methods should be described more fully so that anyone outside the program will have a clear understanding of what is being assessed and how.

DOCUMENTATION

- The assessment plan for this program mentions that a rubric was used to evaluate students' written work. Would it be possible for you to upload a blank version of that rubric to the assessment management system? Supporting documentation of that kind is very useful in providing context for program assessment efforts.

FINDINGS

- Institutional expectations require that academic programs assess at least one student learning related goal and its corresponding outcomes annually, but this program has not reported any findings for [reporting year].
- The findings themselves are minimally adequate, but they are described with very little detail and no supporting documentation was provided. That could make it challenging for people outside the program to understand and interpret the findings.

- Institutional standards for outcomes assessment and accreditation guidelines prohibit the use of overall course grades for program assessment. You can may use a single assignment, or even a set of assignments, but the holistic course grade cannot be used.

Examples of Year Two Standardized Comments

- The findings themselves are minimally adequate, but they would be improved by the inclusion of contextual information about the assessment. For example, it would be helpful to include the total number of students that were assessed, a breakdown of their performance on the assessment (e.g., the percent that earned a grade of A, B, C, or less on the assignment used in the assessment), and the criteria that were used in the assessment (i.e., the rubric in the case of written assignment).
- Whenever possible, the program should provide the data that was produced by an assessment method, a description of how it was obtained, the criteria that was used to analyze it, and a conclusion regarding its validity.

NEXT STEPS

- Institutional expectations require that at least one goal have a strong improvement action tied to it in the "Next Steps". It's debatable whether or not the next steps identified for this program would qualify as strong improvement actions. Please review it.
- The information provided in the Next Steps does not meet minimum institutional standards for program assessment. Next Steps are the improvement actions that the program will carry out in order to improve itself and increase the likelihood that the outcomes are attained.
- Next Steps need to lay out a clear course of action, preferably with a timeline and an explanation of the person or group that is responsible for taking the action. It may also be appropriate to describe how the proposed next steps will improve the program.

OVERALLEVALUATION

- This program assessment plan/report does not meet institutional expectations for outcomes assessment and needs to be revised as soon as possible. See above for details about what needs to be changed or added.
- There are some strong components to this assessment plan/report, but ...
- This is a solid assessment plan/report overall. No changes need to be made at this time, but please take note of the suggestions above for future reporting.
- This is a solid assessment plan/report overall, but the lack of a strong improvement action in the Next Steps means that the plan/report does not meet institutional expectations for outcomes assessment. As a result, it needs to be revised.
- The assessment plan/report for this program meets minimum institutional standards for program assessment, but its design is not uniformly consistent with recommended assessment practices. In order to increase the likelihood that this plan produces authentic and meaningful data, we recommend that revisions be made along the lines detailed above.

Registration for **AALHE 2020** at the Sheraton in New Orleans is open. Please visit www.aalhe.org.

AALHE 2020
TENTH ANNUAL CONFERENCE
NEW ORLEANS, LA
JUNE 8 - JUNE 11, 2020



Association for the Assessment
of Learning in Higher Education

AALHE

PROMOTING ASSESSMENT FOR LEARNING

INTERSECTION

Editorial Board

Kathleen Gorski, Waubensee Community College, Editor in Chief
 Giovana Badia, McGill University
 Jeff Barbee, Indiana University
 Lisa Bonneau, University of South Dakota
 Rebecca Gibbons, City Colleges of Chicago
 Jihee Hwang, University of Oklahoma
 George Klemic, Lewis University
 Shannon Milligan, University of California San Diego
 Karla Pérez-Vélez, Colorado State University
 Kate Pinder, Western Governors University
 Amy Topper, University of Rhode Island
 Shauna Wilton, University of Alberta
 Alison Witherspoon, American College of Education
 Sarah Wu, Georgia Institute of Technology

Intersection is the quarterly publication of the

Association for Assessment of Learning in Higher Education

60 Terra Cotta Ave., Suite B #307, Crystal Lake, IL 60014 | 859-388-0855 | info@aalhe.org

This work is licensed under a

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

ISSN: 2688-7207