

ASSESSING LEADERSHIP EDUCATION IN THREE INSTRUCTIONAL MODALITIES: LESSONS LEARNED

Abstract

Student learning outcomes assessment, when done right, is an ongoing and systematic process. Faculty and staff for leadership education programs often struggle to develop a comprehensive approach to assess the effectiveness of their programs, especially when those programs are offered through multiple modes of instruction. One department at a Midwestern public university has developed a thorough, five-stage process which any leadership education program can follow, regardless of size or method of delivery. This department considers three modes of instruction, including face-to-face on a traditional American campus, asynchronous online, and face-to-face at foreign universities via international partnerships. This manuscript describes the comprehensive process and offers readers considerations for implementation at their own institution.

Introduction

In 2009, Goertzen first called upon leadership educators to reflect upon and engage in purposeful conversation about whether our programs are actually achieving the learning goals we establish. In other words, being mindful to ask, are students learning what we are teaching? He later asserted the same call, adding that in order to do such a thing, we must be willing to embrace the process of assessment (Goertzen, 2013). A search of the *Journal of Leadership Education* database returns a multitude of articles focused on assessing leadership behaviors in college students using singular assignments, activities, or courses; however, little is discussed on assessing leadership education programs. Due to the increased attention on student learning assessment in recent years, one leadership studies department at a midwestern public university has taken a holistic approach to developing

an assessment process that aligns with department and institutional culture and accommodates a mix of program offerings.

This manuscript outlines a comprehensive assessment process at a university offering an undergraduate major in organizational leadership, facilitated through multiple modalities—including on-campus, online, and onsite at two international partner universities in mainland China. Developed over the course of three years (pre-COVID-19), this assessment process includes: developing program-level student learning outcomes, mapping outcomes and courses, designing assessment methods, collecting and analyzing data, and establishing plans for continuous improvement, known as “closing the loop.”

Review of Literature

Higher education has heeded multiple calls for accountability concerning student learning assessment for decades (Shavelson, 2007). It began in the late 1970s, following an almost 80-year progression of standardized testing and assessment. This level of external accountability was led primarily by federal policy makers and the general public who questioned the value of a post-secondary education (2007). It soon led regional and specialized accreditors to require both direct and indirect assessments of student learning in university self-studies (Rhodes, 2016). With these requirements, institutions are now challenged to find the appropriate methods of measuring student learning consistently, and as a result, leadership education was also confronted with the long-awaited challenge.

Upcraft and Schuh (1996) define assessment as “any effort to gather, analyze, and interpret evidence which describes institutional, departmental, divisional or agency effectiveness” (p. 18). The literature on program learning assessment in leadership education highlights various reasons programs have been reluctant to take on the task. Riggio, et al. (2003) believed the complex nature of leadership makes the discipline difficult to assess. Sowcik (2012) asserted the context and conceptual framework of each individual program hinders the discipline’s ability to formalize an assessment process of all leadership programs across the field, as each institution or organization’s context influences the framework, pedagogy, and assessment of the program.

Nevertheless, both the initial and subsequent 2020-2025 National Leadership Education Research Agenda suggest leadership programs should have a plan for assessment through which data provides information to identify program improvements and advance student outcomes. (Boyd et al., 2020; Andenoro, Haber-Curran, Jenkins, Sowcik, Dugan, & Osteen, 2013). Leadership education programs across the nation have since embarked on an assessment journey, from defining programmatic learning outcomes, to planning for comprehensive

program review.

According to the International Leadership Association’s *Guiding Questions*, programs are most often designed as a result of the institution’s identity, politics, and/or culture (2009). Guthrie, Teig, and Hu (2018) reported there are over 1,500 academic leadership programs in the United States, which are offered through various departments or disciplines or are considered interdisciplinary. The 2018 study also found the only consistent offerings within the programs were courses focused on communication and experiential learning. With little course offerings in common, it is clear no leadership program is identical, or works to achieve the exact same learning outcomes, as the program’s context influences its conceptual framework. Some leadership programs focus on the development of self-efficacy, others on the development of relationships, and others focus solely on leadership in organizations (ILA, 2009). As an institution’s culture influences its leadership program curriculum, it also influences its approach to assessing its chosen learning outcomes (ILA, 2009). Regardless of the established program learning outcomes, a practical assessment process is required for any degree of assessment to be successful. According to Metzler and Kurz (2018),

A standard process enables instructors to complete assessments efficiently, with the goal of measuring student learning against program goals and student learning outcomes so that the resulting data may be used to improve a program’s curriculum or the student experience. Without a standardized process, systematic assessment becomes very difficult to document or use on a large scale (p. 5).

Department Overview

The Department of Leadership Studies discussed in this manuscript is housed in the institution’s College of Arts, Humanities, and Social Sciences, and offers an undergraduate major, minor, and multiple certificates, as well as a graduate program

in Organizational Leadership. The undergraduate programs are delivered in multiple modalities of instruction, including on-campus face-to-face (F2F), online, and F2F at multiple international partner institutions located in mainland China. Each program is assessed on an annual basis, which contributes to a five-year comprehensive self-study program review process. The assessment framework is straightforward and can be applied to nearly any academic or professional leadership development program.

The assessment of organizational leadership programs takes place in five stages, illustrated in Figure 1: 1) identify and compose measurable program learning outcomes, 2) map program learning outcomes to courses, 3) select learning activities and identify assessment methods, 4) collect and analyze data, and 5) share results of data collected to adjust learning activities in future courses.

Figure 1

Five Stages of Program Assessment



These efforts involve all department faculty and are intentionally driven and purposefully designed within our academic unit. Combined, the three modalities of instruction involve assessing approximately 3,000 students at various developmental and experiential levels, academically and professionally. In other words, it is a complex undertaking and requires a deep understanding of the organizational context within which this practice is embedded.

Institutional and Program-Level Context. To achieve a comprehensive understanding of the scope and size of the assessment initiative, it is necessary to provide organizational context to the F2F and online

learning environments. While the department’s undergraduate programs each prescribe identical curriculum, course objectives and student learning outcomes, the nuances of each modality and type of student, not to mention logistics, create the need for varying resources, materials, pedagogical strategies, and differentiated instruction, which directly and indirectly affect the assessment of program outcomes. Additionally, the number of students enrolled in the various modalities is quite disparate. Figure 2 displays the total number of students pursuing undergraduate degrees across all modalities over a six-year period. This necessitates intensive planning to create an efficient, systematic process cycle-to-cycle.

Figure 2

Historical Undergraduate Headcount

Year	2014	2015	2016	2017	2018	2019
On-Campus	81	65	64	54	47	39
Online	120	139	148	144	141	99
China	1459	1294	1437	1508	1512	1485

Domestic. There are a variety of reasons students enroll in leadership studies courses—not all are seeking degrees. They could be pursuing a certificate, minor, or major. Some students may also be taking courses through the department as electives or for general interest or as part of a concentration area for a Bachelor of General Studies program. Typically, traditional residential students take their classes F2F on-campus, while most non-traditional students take classes online; however, this is not required. Online courses are heterogeneous, accommodating a mix of younger students with little to no professional experience as well as adult learners with substantial professional experience in various types of organizations and industry sectors. While somewhat rare, some of our F2F classes on-campus also include adult, professional learners.

International. The University maintains two collaborative partnerships in China through the Sino-Foreign Cooperation Agreement of the People's Republic of China. Students complete a Bachelor of Arts or Bachelor of Science in Organizational Leadership while studying a concentration area in one of several other fields through the Chinese host university. They enroll in 24 credit hours (eight classes) of required core courses, while completing and transferring general education and elective credits through the host institution. One hundred percent of our international students are pursuing an undergraduate major and are not permitted to take courses for any other reason. There are no stand-alone certificates, minors, or concentration areas. All students are traditional-aged, residential students with no professional experience, are English as a Second Language (ESL) learners, and (until emergency deployment of remote/online classes due to COVID-19) engage exclusively in a traditional F2F modality. In other words, they are homogeneous. Additionally, class sizes are typically much larger in China than the U.S. equivalent, and English is the language of instruction with each class facilitated by expatriate western faculty. Significant language, cultural, developmental, and technological barriers

create distinct challenges in the assessment of student learning outcomes.

Low student English proficiency at these transnational sites presents the need for additional classroom considerations (Bourgeois & Bravo, 2019; Chang et al. 2014). Implementing teaching strategies to accommodate language differences, such as frequent repetition and student recitation, translated materials, increased reliance on experiential learning opportunities, and allowing students to communicate in their native language during small group activities predicated adjustments deemed necessary to leverage cultural differences relative to course concepts and definitions of leadership (Bourgeois & Bravo, 2019), and, ultimately, in the assessment data collection plan.

Cultural tolerances and practices regarding what constitute academic dishonesty/plagiarism render individual take-home, writing assignments as assessment measures all but impractical. Students complete exams and other assignments in designated university computer labs with the functionality to prevent students from searching online websites for translation assistance or answers to exam questions. As China operates with higher societal levels of in-group collectivism (Javidan et al., 2004), students at the international sites readily offer support by sharing what they believe are the “correct” responses with their classmates. For many classes—including those in other disciplines—it is common for students to create secret websites with the questions and appropriate responses to homework assignments, short-answer essays, and exam questions.

As an additional consideration, traditional Chinese education emphasizes rote memorization and mastery of concepts (Salili et al., 2001), a noted contrast to the experiential learning, constructivist philosophy, and critical thinking practices foundational to western practice. Assignments and activities requiring personal reflection, integrative knowledge, and/or subjectively relative analysis is, figuratively and literally, a foreign concept to these students when they begin their leadership studies

classes. The students are not only learning new curriculum, they are learning how to learn differently. Therefore, classroom-embedded assessments for our international program will vary, sometimes significantly, from those for our domestic students.

Finally, government-initiated restrictions and policies regarding technology significantly limits technology-based resources available to students. While the curriculum of the organizational leadership program was based in western traditions, access to outside sources presents a unique set of issues. Many western-based software and internet companies, such as those managed by Google, have refused to agree to the *Public Pledge on Self Discipline for the Chinese Internet Industry*—a commitment to censor content and provide access to user's personal information—choosing organizational ethics over government regulations (Hamilton et al., 2009). With decreased access to research databases and websites, video resources such as YouTube, and other prohibited materials, student engagement with and deeper exploration of course concepts was substantially diminished in comparison to their stateside undergraduate peers. Accordingly, approaches to program level assessment reflected accommodations for the international constraints.

Description of Practice: Five-Stage Assessment Framework

To begin, our department established a five- to seven-person assessment committee, comprised of faculty members representing each modality, to develop the five stages. The committee provided recommendations to all full-time faculty members for discussion and revision. College and institutional support were also present through each major stage of the assessment plan development; however, this process was designed solely at the department level, as no program-specific accreditation or external review currently exists for leadership education programs. A template for each stage is shared with a keen awareness of how our insights in developing

this plan may be adapted in different leadership education contexts.

Stage One: Identify and Compose Program Learning Outcomes. Program-level student learning outcomes are the foundation of any assessment plan seeking to better articulate what students should learn. Identifying program learning outcomes should not be confused with establishing program learning goals. Program learning goals (PLGs) identify broad themes of the educational program, which can be gleaned from both curricular and co-curricular activities. For example, an academic major may connect learning outcomes to all program goals, while a certificate program may capture only a few program goals. Our program learning goals are foundational to everything we do and are depicted in Figure 3.

Figure 3

Leadership Program Learning Goals

Knowledgeable	Self-Reflective	Improvement-Oriented	Engaged Collaborator	Living with Integrity
Leadership is a set of learned capacities, rather than a set of inherited traits. Students possess knowledge of leadership theories and skills and can transfer them to organization, community, and global contexts.	Students have the capacity to be self-aware and identify their own strengths and challenges. Self-reflective students accept and utilize constructive criticism for continual personal development. Students possess the capacity to demonstrate emotional intelligence to impact others and the world in a positive way.	Students take initiative to address the challenges of the organization or community. In an effort to improve effectiveness, they courageously and strategically challenge policies, laws, and practices that are ineffective. Students take the role of change agent by envisioning 'what out to be' and persist throughout the change process, resulting in transformational change for the collective good in any context.	Students possess the ability to create and nurture relationships with various stakeholders to foster a team environment. These influence-based relationships result in creativity, transformational change, and lasting results.	Students accept responsibility for their own decisions and actions and demonstrate concern for how their choices impact the local and global world. Students are champions of principle.

Outcomes can be created at the *course* or *program* levels. Program learning outcomes (PLOs), therefore, are statements that describe the measurable knowledge and/or skills a student can demonstrate upon completion of their academic program. Learning outcomes are not to be confused with objectives, which describe the intentions of the course, as well as the content, activities, and resources students use to engage in the subject (RPI, 2010). For the purposes of assessment, learning outcomes must be established to measure the “product” of the student experience (e.g. academic program). Guided by these learning goals, the department began developing learning outcome statements, defined as statements describing measurable knowledge, skills, or competencies students can demonstrate accurately (2010).

We established multiple learning outcomes that could be measured through demonstrated student behavior and reflective practice. When writing the statements, we referred to Bloom’s Taxonomy (IACBE,

2018), which guided our selection of action verbs. A selection of our program learning outcomes are listed in Figure 4. Each of these outcomes are measured through specified criteria and rubric instruments, which is discussed in stage three.

Figure 4

Selected Leadership Program Learning Outcomes

PLO1	Demonstrate capacity of leadership theories and concepts in multiple contexts.
PLO2	Demonstrate emotional intelligence.
PLO3	Demonstrate cross-cultural competency.
PLO4	Design contextually appropriate plans to overcome leadership challenges.
PLO5	Demonstrate ability to effectively work across factions with multiple stakeholders.
PLO6	Deploy appropriate influence & conflict resolution techniques for collaborative efforts.

Stage Two: Course Mapping. Faculty members unanimously agree the most beneficial component of the assessment planning process is course mapping, which refers to the creation of a matrix that shows the intersection between individual courses and program learning outcomes (Ewell, 2013). Course maps are used by academic programs nationwide, often because specialized accrediting bodies heavily champion them in institutional self-studies (Hutchings, 2016).

Course mapping highlights meaningful connections between individual course objectives and program outcomes, but this exercise also increased the faculty

members' understanding and appreciation for each course in each learning modality and the necessary sequencing required for maximum scaffolding. Other studies and resources have found this phenomenon to be consistent with our experiences (Haworth & Conrad, 1997; Uchiyama & Radin, 2008). To begin this stage, faculty members teaching courses in all modalities shared their syllabi, each of which included consistent course-level learning objectives and outcomes. The assessment committee then mapped all major courses to program learning outcomes. Figure 5 demonstrates how five course-level learning outcomes in a single course maps to each program-level learning outcome.

Figure 5

Example of Course-Level Outcomes Mapped to Program Learning Outcomes

Leadership Class X	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6
Define modern organizational contexts.	X					
Identify organizational strengths and limitations.				X	X	
Analyze organizational strengths and limitations.				X	X	
Design strategies for organizational improvement.				X		
Design leadership development programs.				X	X	

Next, we mapped the required program courses at three levels of student achievement: Introductory (1), Broadening (2), and Fulfilling (3) (see Figure 6). This process provided an opportunity for faculty to think deeply about a student's trajectory through the leadership education program and at what level the outcomes are achieved in each class. For example, does it make sense for a student to have achieved a program learning outcome at the introductory level in a sophomore-level course, and fulfill the same learning outcome in a senior-level course, without a broadening course in between the two experiences? These types of questions guided individual course revision in the later stages of the assessment process and identified gaps in course offerings.

Figure 6

Example of Courses Mapped to Program Student Learning Outcomes by Level of Achievement

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6
Leadership Class 1	1	1	1			
Leadership Class 2	1		2	1	1	1
Leadership Class 3	2			2		2
Leadership Class 4	2	2			2	
Leadership Class 5	3	3	3	3	3	3

KEY: 1 = Introductory; 2 = Broadening; 3 = Fulfilling

Stage Three: Select Learning Activities, Identify Assessment Methods, Set Benchmarks. The most challenging stage of creating the assessment plan was stage three—developing learning activities, associated assessment methods, and setting benchmarks for success. With multiple modalities, courses, and program learning outcomes, it was imperative to carefully select assessment metrics that emphasized *quality* and richness of data without producing an unmanageable *quantity* of data. Faculty members, through departmental discussions, shared learning activities for each course in each modality. A focus of these discussions included the purpose and intent of the learning activity or assignment through which assessment takes place to ensure alignment to learning outcomes, which, according to Hutchings (2016), is key for students when demonstrating intended learning.

As a result of collaborative brainstorming, faculty agreed on multiple methods of program-level assessment to be embedded into intentionally chosen courses. Consistency between course sections is imperative. Student learning is assessed at two intervals- one at the mid-point (sophomore) and one at the end (senior), including a culminating experience. This approach was chosen for two primary reasons. First, while the sophomore-level course is the mid-point of the major and minor programs, it is also the end-point for a popular certificate program. Collecting data at this juncture provides insight into major and minor students' progress to date, as well as certificate students' level

of achievement. Second, student learning is assessed at the senior-level, as they are challenged to apply everything they have learned in the program into a culminating internship experience, which provides an opportunity for them to apply their learning through experiential and reflective practice, considered vital to the overall program goals.

A variety of holistic and analytic rubrics were developed or revised for each assessment. Developing interrater reliability and consensus estimates on each rubric is critical to the accuracy and consistency of scoring amongst the faculty team (Stemler, 2004; Oakleaf, 2009), so faculty have met on a fairly regular basis to ensure the rubrics are measuring the desired knowledge and skills. Most rubrics designed for our department's program-level assessment are analytic, though some holistic rubrics are embedded in courses throughout the program. Using analytic rubrics, instructors score individual parts of the student's product or performance, whereas a holistic rubric requires the instructor to score the product or performance as a whole (Mertler, 2000).

Beyond the practice of designing quality assessment methods, a program should also establish benchmarks for success. These benchmarks, or defining characteristics of success, identify the end goal set for our students. We established two distinct steps. First, faculty set the *standards*, which reflect agreement about the lowest level of acceptable performance, also referred to as the "good enough" threshold, and then articulated its opposite as the

standard. Second, we established performance *targets*, which describe the percentage of student work that will meet the performance standard for a given assessment. An example for one of our program outcomes is depicted in Figure 7.

Figure 7

Benchmarks for PLO 6: Influence & Conflict Resolution

Metric	Standard	Target
Final Effectiveness Rubric in LDRS X	<i>Proficient</i> level or higher on rubric items: Conflict Resolution; Influence	80%
Final Reflection Paper Rubric in LDRS X	<i>Distinguished</i> level on rubric items: Conflict Resolution, Decisions and Consequences	80%
Internship Supervisor Rubric	<i>Distinguished</i> level on rubric items: Conflict Resolution, Decisions and Consequences	80%

Ideally, all educators aspire to have 100% of their students achieve the highest level of success 100% of the time; however, in practice we know this is not possible. So how do we determine a reasonably acceptable benchmark for our assessment plans? Setting benchmarks is somewhat arbitrary, depends upon the objective, and requires experienced faculty judgment. Benchmarks can follow local and/or external standards, reflect value-added dimensions, and/or be based on historical benchmarks. All have pros and cons and should be carefully considered through faculty collaboration.

Stage Four: Collect & Analyze Assessment Data. Collecting data from the course-embedded assessments is perhaps the least complex stage in the process; however, it does not come without challenges and can be time-consuming to set-up. Efficiency was our aspirational goal, especially given the size of our student body. We intentionally built the plan in a way that maximizes testing and rubric tools available through the institution’s learning management system (LMS) and assessment software. In an effort to establish patterns and trends in student learning,

the committee agreed it was best to collect data for all learning outcomes each semester. All students taking courses in our on-campus F2F and domestic online modalities are assessed each semester. Because of the size of our international student body, we collect data from a representative random sample of 25%.

The chair of the assessment committee communicates with all faculty teaching a given course in which an assessment is embedded, and ensures rubrics and pre/post-tests are accurate. Once the assessment has been completed, the committee chair retrieves and analyzes the data for each program learning outcome and prepares a draft report, which is first reviewed in detail by the assessment committee. An exemplar excerpt of data analysis for a program learning outcome related to written and oral communication is depicted in Figure 8.

Figure 8

Excerpt of Data Analysis for PLO

Target, Standards, Benchmark	At least 80% of students will achieve Proficient or higher on the Final Effectiveness Rubric items: Communication (Written, Communication (Oral).	At least 60% of students will achieve Distinguished on the Final Reflection Paper Rubric item: Writing Quality.	At least 60% of students will achieve Distinguished on the Internship Supervisor Rubric items: Communication (Written), Communication (Oral).
Summary Data Results	(Written): 71% Proficient or > (Oral): 82% Proficient or >	60% Distinguished	(Written): 62% Distinguished (Oral): 62% Distinguished
Review / Analysis	Compared to 2018, levels of achievement for written communication decreased in all modalities (from 77%). The benchmark for written communication was not met. Compared to 2018, levels of achievement for oral communication decreased (from 91%), but still met the benchmark.	The level of achievement for this measure decreased from 85% in 2018; nevertheless, the benchmark was met for 2019. It is difficult to discern what may have caused this result to decrease so significantly; it may be an anomaly.	Compared to 2018, levels of achievement increased for both written communication (from 42%) and oral communication (from 52%); therefore, the benchmark was met. This could be the result of multiple factors. For example, to improve oral communication, faculty in the summer and fall 2019 sections of the course incorporated new assignments through Big Interview, a tool used by Career Services to prepare students for job interviews. These assignments allowed students to practice professional oral communication in the classroom, which may have translated to their work with the internship supervisors.

After the annual report is approved by the committee, it is shared with the entire department for discussion, heavily focusing on necessary action steps to “close the loop.”

Stage Five: “Close the Loop” & Share Results for Improvement. The entire purpose of assessment is to identify strengths and weaknesses that lead to direct improvement in student learning. At this stage, decisions need to be made collectively to determine if/whether and how/what/when changes will be made. This critical step is often referred to as “closing the loop,” where faculty review results from assessment, identify areas of improvement, and act. According to Banta and Blaich (2011), this is often where assessment efforts fall short.

Data should systematically inform program revisions in each delivery modality. If the data reveal unmet outcome(s), changes may be necessary; however,

caution is warranted, especially after only one assessment cycle. Unmet benchmarks may be attributable to one or more of many factors. Figure 9 depicts those factors in the first column, with corresponding possible action steps for closing the loop.

Figure 9

Types and Examples of “Closing the Loop” Changes

Types of Change	Examples
Curriculum	Change prerequisites or GE requirements; Add required courses; Replace existing courses with new ones; Change course sequence; Add internships, labs, and other hands-on learning opportunities
Faculty Support	Provide targeted professional development opportunities; Increase number of TAs or peer mentors; Add specialized support to faculty (Library, Academic Technology, etc.); Increase support to promote dialogues and community among faculty
Pedagogy	Change course assignments; Add more active-learning components to course design; Change textbooks; Increase opportunities for formative feedback and peer-assisted learning
Student Support	Increase tutors; Add more online resources; Improve advising to make sure students take the right courses; Provide resources to encourage community building among students and between students and faculty
Resources	Change the course management system; Improve or expand lab spaces; Provide resources to support student independent research
Assessment Plan	Refine PLO statements; Change methods and/or measures; Change where (e.g. courses) the data are collected; Collect additional data; Improve data reporting and dissemination mechanisms; reconsider targets and/or benchmarks

Adapted from Division of Academic Affairs: Assessment and Institutional Effectiveness (2020).

Results After Year Two

At the time this manuscript was written, the department was finalizing its third year of data collection and analysis for the student learning assessment plan, which occurred during the COVID-19 pandemic, when all university classes were moved (rather hastily) to an online and/or remote environment. Student learning may have been significantly impacted by numerous confounding variables, most notably the learning environment for our domestic and international students enrolled in primarily F2F courses, not to mention the physical and emotional toll on both students and faculty. This will almost certainly impact reliability of the assessment data during the third cycle, but we nevertheless learned much during the first two cycles, and remain deeply committed to, and enthusiastic proponents, of this comprehensive assessment process, including seven specific action steps toward closing the loop.

Course Objectives & Learning Outcomes Revisions. During stage two of our process, faculty identified

several courses in which the *course-level* learning outcome statements could not be appropriately mapped to our *program-level* outcomes, thereby affecting assessment measures. As a result, we are engaged in a systematic revision of course objectives and learning outcomes in all core courses to ensure the language is consistent with the subject matter, learning activities, and course-level formative and summative assessments and assignments.

Rubric Creation/Revision. As previously mentioned, many of our assessment measures rely on the use of rubrics. Several data points from the first two assessment cycles revealed the need to create new/ significantly refine existing rubrics for several courses, most extensively for the two courses mapped to all eight program learning outcomes (one at mid-point and the other at capstone). We created a standardized four-item rubric scale, including detailed descriptors for each level of multiple dimensions, as depicted in the example in Figure 10. The exercise of working through the descriptors proved valuable to increasing inter-rater reliability.

Figure 10

Example of Four-item Assessment Scale Rubric

Dimension	Novice	Apprentice	Proficient	Distinguished
<i>Performance elements are in this column.</i>	<i>Descriptors for lowest performance levels will be in this column.</i>			<i>This column represents what is the highest level you expect to see.</i>
Working Across Factions; Collaboration	Shows little evidence of engaging others in the conversation and /or project; Does not listen to others; Never invites others to engage	Shows evidence of engaging others in ways that facilitate their contribution by restating the views of others and/or asking questions for clarification; Listens to others sometimes; Rarely invites others to engage.	Shows evidence of engaging others in ways that facilitate their contributions by constructively building upon or synthesizing the contributions of others; Listens to others consistently; Sometimes invites others to engage.	Shows evidence of engaging others in ways that facilitate their contributions by constructively building upon or synthesizing the contributions of others; Listens attentively to others; Invites others to engage.

Culminating Experiences. Using the exact same assessment measure for all modalities and all students proved to be less informative in some circumstances. For example, while F2F on-campus students found great benefit from and demonstrated growth from their culminating internship experience, our online adult learners struggled to meet the requirement and found it to be less beneficial to their learning, as most of these students have been in the workforce for many years. As a result, the department embarked on a redesign of the culminating experience, allowing online adult learners to engage in reflective practice through the creation of an outcomes-based ePortfolio, a method of assessment used by many at undergraduate and graduate levels in leadership education, as well as in other academic disciplines (Goertzen, McRay & Klaus, 2016; Jenkins, 2020; Olsen, 2008). Now, instead of exclusively requiring an internship as the culminating experience, a new capstone course has been created to provide students with two paths of achievement—an internship (for our traditional, F2F on-campus learners), or an ePortfolio (for our online adult learners).

Quantitative versus Qualitative Data. Another area we are currently evaluating relates to the value of relying on mostly quantitative data points resulting

from heavy use of analytic rubrics. Metzler and Kurz (2018) suggest, “non-numeric, holistic data may be an important supplement that helps us to identify learning gaps and developmental problems that we might otherwise miss when performance is reduced to a simple numerical score” (p. 19). In other words, qualitative data may help us better understand the “why” to the numeric result. This can include focus groups, interviews, open-ended survey items, examples of student work, and observations, among others. We expect to add more holistic measures in future assessment cycles.

Gaps in Online Assessment. When assessing our online students, we also found gaps in our ability to assess certain skills. For example, assessments regarding oral communication skills were significantly lacking in the online environment. Essentially, we were unable to assess oral communication skills because students were not assigned opportunities to demonstrate them. Subsequently, at least two courses were adjusted to include learning activities, assignments, and assessment measures specifically targeting oral communication skills.

Curriculum Gaps. It became apparent after the second assessment cycle that we needed to pay

more attention to the outcome requiring students to demonstrate cross-cultural competency. We discovered the subject was not covered well enough in our courses to produce the desired benchmark, so we established a working group to significantly redesign our introductory course on leadership theory by focusing more explicitly on foundational concepts of diversity and global leadership. This formative change improved students' ability to reflect on their cross-cultural competencies in other sophomore-level courses where the outcome was assessed, subsequently providing a more accurate depiction of our students' abilities.

Assessment of International Students. This process has truly enlightened our faculty on the difficulties of assessing ESL students using the same measures we use for our native English-speaking students. As a result, we changed the wording on our pre- and post-assessments to better account for cultural barriers and language proficiency by employing more familiar vocabulary, translations, and facilitation. Attention to context and language ensured our nonnative English-speaking students could better understand the questions being asked (Bourgeois & Bravo, 2019).

Assessing our students at the international partner institutions also proved difficult at the upper-division, senior-level. Unlike domestic on-campus and online students, the internationally-located students completed a culminating experience (e.g. internship) administered by faculty from the partner institution, not the stateside department. Consequently, it was difficult to measure student achievement on all program learning outcomes in a single course-embedded project. Over two assessment plan cycles, the department attempted two separate endeavors to capture learning at the senior-level of the program. Each attempt garnered significant participation; however, a culturally different acceptance of academic dishonesty was apparent among student submissions. Instead of assessing what students had learned, the data confirmed students' abilities to access shared responses. At the

same time, the amount of time and human resources needed for each attempt was not sustainable. In response to the high frequency of shared responses, the committee—and the department—opted to abandon a strategy of collecting assessment data through the implementation of a long-form writing assignment for a more structured approach that included guided short-answer responses to fictional case examples. This method afforded the assessment team to measure individual competencies and outcome achievement with a higher level of clarity and accuracy. Additionally, we are in the process of creating a separate assessment plan for our international students that includes more realistic benchmarks and a holistic measure of integrative learning appropriate for ESL students.

Lessons Learned

Program learning outcomes assessment is a dynamic process of continuous improvement; seeking new ways to improve student learning through a variety of assessments never ends. Regardless, if a leadership education program is seeking a solid framework, we offer a few recommendations for success based on the lessons we have learned thus far.

First, a culture of assessment from the top is imperative, institutionally and departmentally. Assessment committees and coordinators must have faculty buy-in and participation. This participation must be present from the onset, when the program determines its program-level learning outcomes, and faculty must remain active throughout data analysis in order to advance changes for improvement.

Second, ultimately, faculty know their students best and can provide a variety of perspectives during each stage of data analysis and recommendations for improvement. Further, faculty must be willing to collaborate across course sections to create consistency in the student learning experience. Though faculty have the autonomy on how to teach their courses, they must be willing to measure

their students' performance consistently—across sections and across terms. Without a uniform use of assessment measures, identifying emerging trends in student learning cannot occur.

Third, technology is key. The amount of time and human resources invested in assessment data collection is far too great when done ad-hoc or manually. Administrators should seek the most efficient technological solutions available. Our institution implemented an assessment data collection software during the third year of our assessment plan, which significantly streamlined our ability to efficiently collect data for large numbers of students. All data is housed in a central space, and all faculty involved in data collection have access.

Fourth, assessment facilitators should not be afraid to experiment. Student learning assessment involves much trial and error, which should be embraced. Although trial and error can be frustrating, it is nonetheless a necessary part of the process and reinforces the necessity to consider the whole student experience, not just the numbers resulting from assessment measures. We tried two comprehensive ways to assess upper-division Chinese students that were less than ideal. We are in the process of developing a better, more refined and ultimately more accurate measure this academic year. Learning what doesn't work can often lead to better insights about what does.

Fifth, do not take quantitative data as the sole descriptor of student learning achievement. There is so much more to the student learning experience than a few assessment measures. The process of data analysis, reflection, and recommendations for improvement should not be solely focused on whether students meet an arbitrary benchmark. Consider growth over time and ways experiences outside of the classroom contribute to it. Using additional evaluation metrics of program success, like student surveys, alumni surveys, exit interviews, and/or focus groups, can help better illuminate the student experience holistically. This type of information can highlight the

“why” behind your results and provide support for long-term initiatives or curriculum changes.

Sixth, when a significant challenge is introduced into your program, such as a global pandemic, a continuous process such as this makes it somewhat easier to address deficiencies or gaps in student learning because you can always return to the foundational process. Regardless of the challenging experiences faculty and students faced in the years of COVID-19, the fundamentals established by this assessment process remind us of the importance of adhering to the established learning goals and outcomes, even when the methods of instruction are varied. It is entirely foreseeable that wholly new modalities may emerge post-pandemic, and this systematic process can be easily adapted to accommodate that eventuality.

Conclusion

Student learning assessment is a comprehensive process filled with unexpected challenges, made more complex when operating across multiple modalities of instruction and student types. Establishing a consistent culture of assessment in any department, for any program, takes time, but developing a comprehensive assessment plan using this five-stage framework is a solid place to start. Though simple on its face, this framework provides faculty and administrators for leadership education programs to thoroughly reflect on their work, and whether their programs achieve the student learning goals they have set. This framework also adds to the body of knowledge on assessing leadership education programs.

While there are plenty of learning and assessment activities available to leadership educators in the literature, this framework offers faculty and administrators a way to look at “the big picture,” and brings together what are often isolated assessment

practices into a cohesive whole. Though Sowcik (2012) asserts the individuality of leadership education programs hinders the discipline's ability to formalize assessment processes, we disagree. A formal assessment process can be tailored to any individual leadership program, so long as faculty and administrators work together to identify quality learning outcomes, robust learning activities and assessment methods, combined with the willingness to intentionally reflect on the results and establish plans for improvement. The main goal of this framework is continuous improvement. We must practice what we preach as leadership educators, and ensure our students are gaining the knowledge and skills necessary to succeed as leaders in the future. We cannot do that without systematically and continuously examining our own practices, which is the foundation of our five-stage assessment framework.

References

- Andenoro, A. C., Allen, S. J., Haber-Curran, P., Jenkins, D. M., Sowcik, M., Dugan, J. P., & Osteen, L. (2013). *National Leadership Education research agenda 2013-2018: Providing strategic direction for the field of leadership education*. <http://leadershipeducators.org/ResearchAgenda>.
- Banta, T.W. & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning*, 43(1), 22-27. doi: 10.1080/00091383.2011.538642
- Boyd, B. L., Armstrong-Smith, C., Forbes, A., & Holmes, A. C. (2020). Understanding the leadership learner: Priority 3 of the National Leadership Education Research Agenda 2020-2025. *Journal of Leadership Studies*, 14(3), 50-55. doi: :10.1002/jls.2171
- Bourgeois, J. & Bravo, C. (2019). Engaging students beyond discussion: Leadership education and nonnative English-speaking classrooms. *Journal of Leadership Education*, 18(3), 113-130. doi: 10.12806/V18/I3/R8
- Brungardt, C. L., Greenleaf, J. P., Brungardt, C. J., & Arensdorf, J. (2006). Majoring in leadership: A review of undergraduate leadership degree programs. *Journal of Leadership Education*, 5(1), 4-25.
- Chang, T. S., Bai, Y., & Wang, T. W. (2014). Students' classroom experience in foreign-faculty and local-faculty classes in public and private universities in Taiwan. *Higher Education*, 68(2), 207-226.
- Division of Academic Affairs: Assessment and Institutional Effectiveness. (2020 January 27). "Closing the loop": Types and Examples of Changes. California State University-Fullerton. http://www.fullerton.edu/data/assessment/sla_resources/closetloop.php#:~:text=One%20recommended%20way%20of%20doing,i.e.%20the%20outcome%20is%20met
- Ewell, P.T. (2013). The Lumina degree qualifications profile: Implications for assessment. (Occasional Paper No. 16). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. <http://www.learningoutcomesassessment.org/documents/EwellDQPop2.pdf>
- Gelfand, M. J., Aycan, Z., Erez, M., & Leung, K. (2017). Cross-cultural industrial organizational psychology and organizational behavior: A hundred-year journey. *Journal of Applied Psychology*, 102(3). <https://doi.org/10.1037/apl0000186>
- Goertzen, B. J. (2009). Assessment in academic based leadership education programs. *Journal of Leadership Education*, 8(1), 148-162.
- Goertzen, B. J., McRay, J., & Klaus, K. (2016). Electronic portfolios as capstone experiences in a graduate program in organizational leadership. *Journal of Leadership Education*, 15(3), 42-52. doi: 10.12806/V15/I3/A5
- Guthrie, K. L., Teig, T. S., & Hu, P. (2018). *Academic leadership programs in the United States*. Tallahassee, FL:

Leadership Learning Research Center, Florida State University.

- Hamilton, J. B., Knouse, S. B., & Hill, V. (2009). Google in China: A manager-friendly heuristic model for resolving cross-cultural ethical conflicts. *Journal of Business Ethics, 86*(2), 143-157.
- Haworth, J., & Conrad, C. (1997). *Emblems of quality in higher education*. Boston, MA: Allyn and Bacon.
- Hutchings, P. (2016, January). Aligning educational outcomes and practices. (Occasional Paper No. 26). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- International Accreditation Council for Business Education [IACBE]. (2018). *Bloom's Taxonomy of educational objectives and writing intended learning outcomes statements*. Lenexa, KS: Author. <http://iacbe.org/wp-content/uploads/2017/09/Blooms-Taxonomy-of-Educational-Objectives-partial.pdf>
- International Leadership Association [ILA]. (2009). *Guiding questions: Guidelines for leadership education programs*. College Park, MD: Author. <http://www.ila-net.org/communities/LC/GuidingQuestionsFinal.pdf>
- Jenkins, D. M. (2020). What the best leadership educators do: A sequential explanatory mixed methods study of instructional and assessment strategy use in leadership education. *Journal of Leadership Education, 19*(4), 37-55. doi: 10.12806/V19/I4/R4
- Metzler, E. T., & Kurz, L. (2018, November). Assessment 2.0: An organic supplement to standard assessment procedure. (Occasional Paper 36). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Mertler, C. A. (2000). Designing scoring rubrics for your classroom. *Practical Assessment, Research, and Evaluation, 7*(25). doi: <https://doi.org/10.7275/gcy8-0w24>
- Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology, 60*(5), 969-983. doi: 10.1002/asi.21030
- Olsen, P. E. (2009). The use of portfolios in leadership education. *Journal of Leadership Education, 7*(3), 20-27.
- Rensselaer Polytechnic Institute [RPI] (2010). *Objectives vs. Outcomes*. <https://provost.rpi.edu/learning-assessment/learning-outcomes/objectives-vs-outcomes>
- Rhodes, T. L. (2015). Assessment: Growing up is a many-splendored thing. *Journal of Assessment and Institutional Effectiveness, 5*(2), 101-116.
- Riggio, R. E., Ciulla, J., & Sorenson, G. (2003). Leadership education at the undergraduate level: A liberal arts approach to leadership development. In S. E. Murphy & R. E. Riggio (Eds.) *The future of leadership development* (pp. 223-236). Mahwah, NJ; Lawrence Erlbaum Associates.
- Salili, F., Chiu, C. Y., & Lai, S. (2001). The influence of culture and context on students' motivational orientation and performance. In *Student motivation* (pp. 221-247). Springer.
- Shavelson, R. J. (2007). *A brief history of student learning assessment: How we got where we are and a proposal for where to go next*. Association of American Colleges and Universities: Washington, D.C.

- Stemler, S.E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation*, 9(4).
- Uchiyama, K.P. & Radin, J.L. (2009). Curriculum mapping in higher education: A vehicle for collaboration. *Innovative Higher Education*, 33, 271-280. doi: 10.1007/s10755-008-9078-8
- Upcraft, M. L., & Schuh, J. H. (1996). *Assessment in student affairs: A guide for practitioners*. San Francisco: Jossey-Bass.