

Student Learning Outcome Assessment in NASPAA Programs: A Review of Validity and Reliability

David C. Powell

California State University, Long Beach

Michelle Saint-Germain

California State University, Long Beach

ABSTRACT

Developments in regional and professional accreditation have brought profound changes to how public affairs programs measure student learning outcomes (SLOs). There is little doubt that these developments have resulted in an increase in the frequency of SLO assessment as well as an increase in the variety of assessment mechanisms used. As programs continue to focus resources and attention on assessment, it is instructive to examine how programs are engaging in the assessment of SLOs and what best practices have emerged from these assessment approaches. This study reviews the assessment strategies and instruments utilized by programs in the 2012–2013 NASPAA accreditation cohort, focusing on the types of validation and reliability methods used by programs.

KEYWORDS

Assessment, reliability, validity

BACKGROUND

Assessment in Accreditation

Accreditation has historically been based on standards. However, the basis of these standards has changed profoundly over the past several decades. Most accrediting agencies have progressed from an input- and resource-based orientation to a focus on outputs and outcomes. This progression is evident in many regional and national accrediting bodies (e.g., WASC, ABET, AACSB, and NASPAA)¹ and is endemic to higher education today (Weiss, Cosby, Habel, Hanson, & Larsen, 2002).

This shift to outputs and outcomes has changed the discourse about quality in higher education and had encouraged many institutions and

programs to fundamentally redefine quality. Part of this redefinition has resulted in a focus on assessment across the curriculum in many programs. The clear result has been an increase in instances of assessment as well as the variety of assessment approaches used.

Assessment of Competencies

As accrediting organizations moved toward emphasizing the assessment of outcomes, a renewed focus on competencies became prevalent. The movement toward competency-based education has been chronicled in other places by more skillful hands (Getha, Hummert, Nalbandian, & Silva, 2013; Rissi & Gelmon, 2014; Tompkins, Laslovish, & Greene, 1996; Voorhees, 2002), but it is clear that many higher-education institutions have embraced a competency-

based focus. At the extreme end of this spectrum are schools such as Southern New Hampshire University, which has gone so far as to identify 120 competencies that students must demonstrate proficiency in to earn the school's Bachelor of Arts degree. These competencies are not linked to any specific courses, and students are not required to complete any mandated credit-hour requirements. However, most programs and institutions do systematically connect competencies to specific courses and credit completion requirements. Indeed, systematic development and assessment of competencies is considered a best practice in education (Arinto, 2013).

While it is clear that competency development and assessment should be a systematic process, the nature of these competencies varies from discipline to discipline and even program to program. For example, Green (1999) identified four categories (knowledge of processes; values and priorities; job skills and work habits; and communication, leadership, and teamwork) as

elemental competencies for employees entering organizations. Ewell (1994) focused on knowledge and how knowledge is used; he identified the dichotomy of knowledge reproduction and knowledge construction as the basis of competency determination. Jennings (1989) defined competencies within public affairs programs by focusing on knowledge, values, skills acquired, and effectiveness at doing jobs. It is clear that competencies may be defined and measured in a variety of ways.

NASPAA's 2009 Standards

The core of assessment in NASPAA's 2009 standards is clearly competency-based. However, NASPAA provides programs with flexibility in how competencies are defined and how outcomes are measured. This approach requires that programs define what type of professionals they want to produce and translate NASPAA's five universal required competencies (to lead and manage in public governance; to participate in and contribute to the policy process; to analyze, synthesize, think critically, solve prob-

TABLE 1.
Examples of Direct and Indirect Evidence of Student Learning

Direct Evidence of Student Learning	Indirect Evidence of Student Learning
Annotated bibliographies	Exit interviews with graduating students
Appraisals of live or videotaped presentations	Focus groups with students, alumni, etc.
Assessment center/mock interviews	National ratings or rankings
Case study analysis	Student self-assessment
Completed course assignments, homework	Surveys of students or alumni
Community service project report	Surveys of employers in general
Comprehensive exam, midterms, finals	
Culminating projects or papers, thesis	
Internship supervisor report	
Journals, discussion board posts, blogs	
Oral exam judged by faculty/practitioners	
Policy analysis reports	
Role playing, simulation	

lems, and make decisions; to articulate and apply a public service perspective; and to communicate and interact productively with a diverse and changing workforce and citizenry) into specific student learning outcomes (SLOs) relevant to the program's mission and student population. This places a premium on alignment of assessment criteria and performance objectives (Aristeguieta & Gomes, 2006).

The implications of a competency-based, mission-driven assessment process are profound for programs. This type of process requires programs to engage in a sustained, iterative process of articulating learning outcomes and to link these outcomes to the curriculum. This iterative process typically consists of several steps:

1. Interpreting the required competencies relative to program mission and population
2. Mapping the competencies into expected SLOs
3. Mapping the SLOs into the curriculum
4. Identifying and adopting teaching and learning methods to promote the achievement of competencies
5. Developing and/or selecting assessment methods to evaluate student accomplishment
6. Analyzing the results of the assessment process and applying these results for program improvement

This iterative process will vary from program to program and may involve both direct and indirect evidence of student learning (Saint-Germain & Powell, 2012). Table 1 lists examples of direct and indirect evidence of student learning that programs have used in the past.

The development of these different types of evidence of student learning will likely involve multiple stakeholders who may possess a variety of interpretations and approaches to measuring competencies and SLOs. Therefore, the involvement of these stakeholders places an imperative on developing consistent, valid, and reliable measures.

Reliability in Assessment

One of the primary goals in creating mission-based, competency-grounded assessment mechanisms should be the creation of reliable measures. Reliability typically consists of three dimensions: stability, equivalence, and internal consistency (O'Sullivan, Rassel, & Berner, 2008).

Stability requires measures that yield similar results over time (O'Sullivan et al., 2008). This suggests the need for a longitudinal approach to assessing student learning outcomes in which students are exposed to similar assessment measures during their time in the program. Programs may achieve this through pre- and postcourse or -program measurements. A test-retest method is often useful in assessing the stability aspect of reliability (Heise, 1969).

Equivalence requires that two or more "assessors" assign the same or similar values to the same phenomenon (Shrout, 1998). Interrater reliability is a common measure for assessing the equivalence dimension of reliability (Armstrong, Gosling, Weinmann, & Marteau, 1997). Programs can incorporate this type of measurement through the use of multiple "assessors" of student work and the application of grading rubrics.

Internal consistency refers to the homogeneity of measures and indicators that are meant to represent the same concept (O'Sullivan et al., 2008). A common measure of internal consistency is the coefficient alpha (Streiner, 2003). Programs may apply internal consistency through the use of a rubric and through measuring the consistency of indicators attributed to grades and evaluative measures. These measures are often linked to behaviors exhibited by students and can range from complex to simple scales. For example, Arinto (2013) recommended a simple three-tiered assessment for basic, intermediate, and advanced competencies. These discrete categories may be anchored to behaviors exhibited by students in the assessment artifacts that the students create.

Validity in Assessment

In addition to reliability, assessment measures should also be operationally valid. Two of the most common validation methods are content and criterion validation (Messick, 1990). Ostensibly, the content of the assessment artifact created by students should correlate with the competency measured. Programs may achieve this through a variety of mechanisms, including course-embedded assignments, case studies, and the like.

Criterion-based validation may be achieved through concurrent or predictive approaches, among others. A concurrent approach could correlate the results of an assessment measure at the end of a student’s program to the student’s performance in a course that delivered the competency being measured. For example, the performance of a student in a budgeting case study completed at the end of the program could be compared to the performance of the student on assessment measures within a previously completed budgeting course. The inverse of this approach (using student performance on assessment measures in a budgeting course to predict performance on a capstone budgeting assignment at the end of the program) could constitute a more predictive approach to this type of criterion-based validation.

The Current State of Assessment in Graduate Public Affairs Programs

It is clear from the preceding discussion that programs can pursue many options as they develop assessment systems in accordance with more competency-based accreditation standards.

Since the current NASPAA standards were adopted recently, it is instructive to examine the progress that programs have made in developing their assessment systems and the role that reliability and validity play in this development.

METHODS

This study examined the self-studies and supporting documentation for 29 programs that were part of NASPAA’s 2012–2013 accreditation cohort. The primary question asked was, “What are NASPAA programs doing for assessment of student learning on the five universal competencies?” We collected data on the extent of progress on assessment efforts, the actors involved, the assessment strategies and instruments employed, and whether the validity and reliability of assessment were explicitly mentioned. We also conducted interviews with representatives from a sample of programs on their experiences with assessment.

FINDINGS

Extent of Program Progress on Assessment

The majority of programs in NASPAA’s 2012–2013 accreditation cohort did not report the completion of an entire assessment cycle for all five universal required competencies. Only five programs reported a completed assessment cycle for all competencies. However, it was not clear that all five competencies had actually been measured by these programs. This concern was cited by NASPAA’s Commission on Peer Review and Accreditation (COPRA) in four of these five programs. Table 2 shows the progress of programs in completing assessment cycles.

TABLE 2.
Number of Programs Completing an Assessment Cycle for Each Competency

To lead and manage in public governance	To participate in and contribute to the policy process	To analyze, synthesize, think critically, solve problems, and make decisions	To articulate and apply a public service perspective	To communicate and interact productively with a diverse and changing workforce and citizenry
10	8	15	9	9

The vast majority of programs ($n=18$) reported completing an entire cycle for only one competency. Most programs ($n=15$) reported completing an assessment cycle for the competency “to analyze, synthesize, think critically, solve problems, and make decisions.” Ten programs completed an assessment cycle for the first competency, “to lead and manage in public governance.” An equal number of programs reported completing a cycle for the fourth competency, “to articulate and apply a public service perspective” ($n=9$); and for the fifth competency, “to communicate and interact productively with a diverse and changing workforce and citizenry” ($n=9$). Eight programs reported completing a cycle for the competency “to participate in and contribute to the policy process.”

Therefore, it appeared that the vast majority of programs had not completed an entire cycle for all five competencies and most completed an entire cycle for only one competency. However, a number of programs received only a one-year (re)accreditation due to insufficiencies in assessment of student learning vis-à-vis the universal required competencies. Programs in the 2014–2015 accreditation cohort will be expected to complete an entire assessment cycle for three of the five universal competencies.

Actors Involved in Assessment

Several programs ($n=11$) included students in assessing their own work. Some programs required students to assess their work utilizing the same rubric used by faculty. Seven programs reached outside the program to involve external stakeholders in assessment. For programs that utilized external stakeholders, the majority used practitioners and advisory board members to provide feedback on student achievement. Despite these attempts to involve external stakeholders, the dominant modality was to utilize only full-time faculty for assessment. As discussed above, this is one area where most programs are not following the recommended best practices to ensure fair, valid, and useful assessment data. The involvement of multiple reviewers would afford programs, not only a variety of feedback, but also the opportunity to conduct inter-rater reliability measures across stakeholder groups.

Assessment Tools and Strategies

Table 3 shows the types of evidence of student learning collected by the programs, as reported in their self-studies. The majority of programs used direct evidence of student learning, mostly through course-embedded assignments and exams. Outside of courses, programs also used comprehensive exams, pre- and postprogram essays or exams, portfolios, and internships or experiential learning experiences. The majority of programs are using types of direct evidence listed as best practices for assessment of student learning.

Several programs supplemented direct evidence of student learning with indirect evidence, the latter generally consisting of student or alumni surveys or focus groups. However, some programs used only indirect evidence. In addition to surveys, these programs used such measures as reviews of the curriculum or syllabi, student ratings of instruction, summaries of faculty comments on students, and reviews of student progress toward the degree. COPRA was more likely to cite concerns with assessment of student learning for those programs using only indirect evidence, and these programs were also more likely to have this concern carry over into continued monitoring of assessment on the program’s annual accreditation maintenance report to NASPAA.

Reliability Efforts

The initial analysis revealed that nine programs mentioned attempts to ensure reliability in their self-studies. The primary focus of these programs was to ensure equivalency in their ratings of student work. Not surprisingly, most programs approached the measurement of equivalence through inter-rater reliability. These programs tried to achieve inter-rater reliability by involving multiple faculty members in the development of rubrics and then utilizing these rubrics to evaluate student work. One program spent considerable time reading books on the construction of fair and objective rubrics and then presented the rubric for each universal required competency to students for their comments. Another program utilized a consultative committee in addition to faculty members to

TABLE 3.
Direct and Indirect Evidence of Student Learning Used by Programs

Direct Evidence	Indirect Evidence
Course embedded assignments (<i>n</i> = 15) (case study analysis, blog posts, course paper, policy memo, evaluation proposal, Statistical Package for Social Sciences (SPSS) exercise)	Surveys (<i>n</i> = 11) (students, alumni, exit interviews, focus groups)
Course-embedded exams (<i>n</i> = 5) (common set of questions across sections)	Student ratings of instruction (<i>n</i> = 3)
Peer review of group members (<i>n</i> = 1)	Curriculum or syllabi reviews (<i>n</i> = 3)
Capstone course project or paper (<i>n</i> = 7)	Student progress toward degree (<i>n</i> = 1)
Written or oral comprehensive exam (<i>n</i> = 2)	Faculty comments on students (<i>n</i> = 1)
Pre- and postprogram memos (<i>n</i> = 1)	
Pre- and postcourse standardized exam (<i>n</i> = 1)	
Internship (<i>n</i> = 2)	
Experiential learning project (<i>n</i> = 1)	
Portfolio (<i>n</i> = 3)	

score student work in accordance with the developed rubric. However, no programs reported any statistical analysis of rater scores to establish formal inter-rater reliability. While most programs currently have the data necessary to formally estimate inter-rater reliability, they have not engaged in a systematic attempt to accomplish this. Interviews revealed that the primary obstacles to creating a more rigorous system of inter-rater reliability were resource limitations and lack of faculty support.

Finally, all programs that measured reliability focused on equivalence rather than stability. This was not surprising, as equivalence can be measured by utilizing one cohort of students, while assessing stability would require a more longitudinal approach across cohorts of students over time. Viewing assessment as an ongoing process and adhering to an assessment schedule could allow programs to assess multiple cohorts and draw conclusions regarding patterns in student learning outcomes over time.

Validity

An analysis of the self-studies revealed that 10 programs mentioned procedures that could be

construed as validation measures. Of these programs, the focus seemed to be content validity. Most of these programs utilized faculty-developed rubrics that were directly connected to course-embedded assignments. A few programs had assessment mechanisms in place that allowed for concurrent validation by comparing student performance on capstone and course-embedded assignments.

Programs appeared to be using course-based assessments as well as more programmatic-level assessments. Course-embedded assessments occurred primarily in core courses, while programmatic-level assessments consisted of pretest and post-test observations of students as they enter and leave the program. One large program endeavored to establish a valid pre-program and postprogram multiple-choice test to assess student learning. Beginning with one of NASPAA’s universal required competencies, the program spent 18 months developing a test bank of 25–28 items and validated it with a sample of political science honors undergraduate students. The first year of implementation provided useful evidence that served as the basis for some program changes. However, the

program also cautioned that there are many challenges to using this approach.

As was the case with reliability, most programs were collecting assessment information that could permit either content or concurrent validation methods. However, interviews also revealed that the primary obstacles to validation were scarce resources and lack of faculty support. It is likely that programs are focusing time and energy on completing assessment cycles and have not had the time or ability to devote much effort to measuring validity and/or reliability. More pronounced movement toward measuring validity and reliability will likely occur after programs have met the minimum expectations of COPRA in completing cycles for all five competencies.

CONCLUSION

This article presents a review of the progress that public affairs programs have made in assessing student learning outcomes. The majority of programs reported completing an entire assessment cycle for only one of NASPAA's five universal required competencies. While some programs reported complete cycles for multiple competencies, feedback from COPRA indicated that COPRA did not consider many of these cycles complete. The progress of programs in completing assessment cycles indicates that programs are still in the early stages of developing assessment systems capable of measuring student learning across all five competencies.

While many programs did not directly address efforts to measure reliability and validity, programs have generated assessment evidence that could be used to measure the reliability and validity of assessment artifacts. For example, some programs currently utilize multiple stakeholders (e.g., faculty, consultative committees, etc.) to assess student work. This work is often scored using grading rubrics and the results are reported back to stakeholders at faculty and committee meetings. Therefore, programs are developing databases that can be used in the future to measure inter-rater

reliability. Some programs also reference pre- and postcourse and -program assessments that could be used to measure stability as well.

Many programs reported using course-embedded and programmatic assessment mechanisms as well as alumni and employer surveys. The data generated from these assessments could possibly be used for either concurrent or predictive validation strategies. Performance on course-embedded assessments could be used to predict performance on programmatic capstone requirements or future performance in public or nonprofit organizations. Programs could also correlate student performance on capstone requirements with student performance on course-embedded assignments, as the data that would permit this type of analysis exists in some programs.

It is clear that programs are progressing in the development of their assessment databases that would enable a more robust assessment of both reliability and validity in the future. As programs begin to complete assessment cycles for all required competencies, resources and attention may become available to focus on issues of reliability and validity. Therefore, it will be instructive to revisit this research in the future after programs have completed their initial assessment cycles for all competencies. Once rubrics and processes have been established, the marginal costs of implementing these procedures should decrease, and it will be interesting to see if programs reallocate resources to issues of reliability and validity in the future.

NOTE

- 1 These accrediting bodies are as follows: Western Association of Schools and Colleges (WASC), Accreditation Board for Engineering and Technology (ABET), Association to Advance Collegiate Schools of Business (AACSB), and Network of Schools of Public Policy, Affairs, and Administration (NASPAA).

REFERENCES

- Arinto, P. (2013). A framework for developing competencies with open and distance learning. *International Review of Research in Open and Distance Learning, 14*(1), 167–185.
- Aristeguieta, M., & Gomes, K. (2006). Assessing performance in NASPAA graduate programs. *Journal of Public Affairs Education, 12*(1), 1–18.
- Armstrong, D., Gosling, A., Weinmann, J., & Marteau, T. (1997). The place of interrater reliability in qualitative research: An empirical study. *Sociology, 31*(3), 597–606.
- Ewell, P. T. (1994). *The self-regarding institution: Information for excellence*. Boulder, CO: National Center for Higher Education Management Systems.
- Getha, H., Hummert, R., Nalbandian, J., & Silva, C. (2013). Competency model design and assessment: Findings and future directions. *Journal of Public Affairs Education, 19*(1), 141–171.
- Green, P. (1999). *Building robust competencies: Linking human resources systems to organizational strategies*. San Francisco, CA: Jossey-Bass.
- Heise, D. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review, 34*(1), 93–101.
- Jennings, E. (1989). Accountability, program quality, outcomes assessment, and graduate education for public affairs and administration. *Public Administration Review, 42*(5), 438–446.
- Messick, S. (1990). *Validity of test interpretation and use*. Princeton, NJ: Education Testing Services.
- O'Sullivan, E., Rassel, G., & Berner, M. (2008). *Research methods for public administrators*. New York, NY: Pearson.
- Rissi, J., & Gelmon, S. (2014). Development, implementation, and assessment of a competency model for a graduate public affairs program in health administration. *Journal of Public Affairs Education, 20*(3), 335–332.
- Saint-Germain, M., & Powell, D. C. (2012). *An analysis of standard 5 in 2012 self-studies*. NASPAA white paper.
- Shrout, P. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research, 7*(3), 301–317.
- Streiner, D. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103.
- Tompkins, J., Laslovish, M., & Greene, J. (1996). Developing a competency-based MPA curriculum. *Journal of Public Affairs Education, 2*(2), 117–130.
- Voorhees, R. (2002). Competency-based learning models: A necessary future. *New Directions for Institutional Research, 2001*(110), 5–13.
- Weiss, G., Cosbey, J., Habel, S., Hanson, C., & Larsen, C. (2002). Improving the assessment of student learning: Advancing a research agenda in sociology. *Teaching Sociology, 30*(1), 63–79.

ABOUT THE AUTHORS

David C. Powell is a professor and the director of the Graduate Center for Public Policy and Administration at California State University, Long Beach. His research areas include student learning outcomes assessment and m-government.

Michelle Saint-Germain is professor emerita of the Graduate Center for Public Policy and Administration at California State University, Long Beach. She remains active in research, publishing, and consulting concerning assessment, accreditation, and accountability for graduate programs. She teaches in the fall semester and indulges in her love of bicycling, travel, and watercolor painting the rest of the year.