RESEARCH & PRACTICE IN ASSESSMENT

RPA

RPA RESEARCH & PRACTICE IN ASSESSMENT

CALL FOR PAPERS

Manuscripts submitted to RPA may be related to various higher education assessment themes, and should adopt either an assessment measurement or an assessment policy/foundations framework. Contributions are accepted at any time and will receive consideration for publishing. Manuscripts must comply with the RPA Submission Guidelines and be submitted to our online manuscript submission system found at rpajournal.com/authors/.

RESEARCH & PRACTICE IN ASSESSMENT

The goal of Research & Practice in Assessment is to serve the assessment community as an online journal focusing on higher education assessment. It is dedicated to the advancement of scholarly discussion amongst researchers and practitioners in this evolving field. The journal originated from the Board of the Virginia Assessment Group, one of the oldest continuing professional higher education assessment organizations in the United States. Research & Practice in Assessment is a peer-reviewed publication that uses a double-blind review process. Approximately forty percent of submissions are accepted for issues that are published twice annually. Research & Practice in Assessment is listed in Cabell's Directory and indexed by EBSCO, ERIC, Gale, and ProQuest.

History of Research & Practice in Assessment

Research & Practice in Assessment (RPA) evolved over the course of several years. Prior to 2006, the Virginia Assessment Group produced a periodic organizational newsletter. The purpose of the newsletter was to keep the membership informed regarding events sponsored by the organization, as well as changes in state policy associated with higher education assessment. The Newsletter Editor, a position elected by the Virginia Assessment Group membership, oversaw this publication. In 2005, it was proposed by the Newsletter Editor, Robin Anderson, Psy.D. (then Director of Institutional Research and Effectiveness at Blue Ridge Community College) that it be expanded to include scholarly articles submitted by Virginia Assessment Group members. The articles would focus on both practice and research associated with the assessment of student learning. As part of the proposal, Ms. Anderson suggested that the new publication take the form of an online journal.

The Board approved the proposal and sent the motion to the full membership for a vote. The membership overwhelmingly approved the journal concept. Consequently, the Newsletter Editor position was removed from the organization's by-laws and a Journal Editor position was added in its place. Additional by-law and constitutional changes needed to support the establishment of the Journal were subsequently crafted and approved by the Virginia Assessment Group membership. As part of the 2005 Virginia Assessment Group annual meeting proceedings, the Board solicited names for the new journal publication. Ultimately, the name Research & Practice in Assessment was selected. Also as part of the 2005 annual meeting, the Virginia Assessment Group Board solicited nominations for members of the first RPA Board of Editors. From the nominees Keston H. Fulcher, Ph.D. (then Director of Assessment and Evaluation at Christopher Newport University), Dennis R. Ridley, Ph.D. (then Director of Institutional Research and Planning at Virginia Wesleyan College) and Rufus Carter (then Coordinator of Institutional Assessment at Marymount University) were selected to make up the first Board of Editors. Several members of the Board also contributed articles to the first edition, which was published in March of 2006.

After the launch of the first issue, Ms. Anderson stepped down as Journal Editor to assume other duties within the organization. Subsequently, Mr. Fulcher was nominated to serve as Journal Editor, serving from 2007-2010. With a newly configured Board of Editors, Mr. Fulcher invested considerable time in the solicitation of articles from an increasingly wider circle of authors and added the position of co-editor to the Board of Editors, filled by Allen DuPont, Ph.D. (then Director of Assessment, Division of Undergraduate Affairs at North Carolina State University). Mr. Fulcher oversaw the production and publication of the next four issues and remained Editor until he assumed the presidency of the Virginia Assessment Group in 2010. It was at this time Mr. Fulcher nominated Joshua T. Brown (Director of Research and Assessment, Student Affairs at Liberty University) to serve as the Journal's third Editor and he was elected to that position.

Under Mr. Brown's leadership Research & Practice in Assessment experienced significant developments. Specifically, the Editorial and Review Boards were expanded and the members' roles were refined; Ruminate and Book Review sections were added to each issue; RPA Archives were indexed in EBSCO, Gale, ProQuest and Google Scholar; a new RPA website was designed and launched; and RPA gained a presence on social media. Mr. Brown held the position of Editor until November 2014 when Katie Busby, Ph.D. (then Assistant Provost of Assessment and Institutional Research at Tulane University) assumed the role after having served as Associate Editor from 2010-2013 and Editor-elect from 2013-2014.

Ms. Katie Busby served as RPA Editor from November 2014-January 2019 and focused her attention on the growth and sustainability of the journal. During this time period, RPA explored and established collaborative relationships with other assessment organizations and conferences. RPA readership and the number of scholarly submissions increased and an online submission platform and management system was implemented for authors and reviewers. In November 2016, Research & Practice in Assessment celebrated its tenth anniversary with a special issue. Ms. Busby launched a national call for editors in fall 2018, and in January 2019 Nicholas Curtis (Director of Assessment, Marquette University) was nominated and elected to serve as RPA's fifth editor.

TABLE OF CONTENTS

ARTICLES

FROM THE EDITOR

BOOK REVIEW

Abstract

Well-developed professional development opportunities are a crucial component in ensuring that faculty engaging in assessment are equipped to do this work well. Creating these opportunities requires clear expectations of the knowledge, skills, and attitudes necessary to conduct assessment work. Additionally, the knowledge and skill requirements of faculty conducting assessment are often different from those of professionally trained assessment practitioners. Although higher education student affairs organizations have developed frameworks for assessment skills, no formal framework of knowledge, skills, and attitudes exists to drive professional development in assessment within academic affairs. This article provides a framework of assessment-related knowledge, skills, and attitudes that are important to the professional development of faculty assessment practitioners, targeting three levels of complexity. This framework can be used to evaluate current professional development offerings and plan new, intentionally designed programs in accordance with backward-design principles.

AUTHORS

S. Jeanne Horst, Ph.D.
James Madison University

Caroline O. Prendergast, M.Ed.
James Madison University

The Assessment Skills Framework: A Taxonomy of Assessment Knowledge, Skills and Attitudes

*C*onducting quality student learning outcomes assessment requires diverse knowledge, skills, and abilities. Given that people engaging in assessment work often lack formal assessment training (Hutchings, 2010; Nicholas & Slotnick, 2018), professional development (PD) opportunities are needed in areas as disparate as student learning outcomes development, curriculum mapping, data management, and reporting results. Simultaneously, the lines between campus assessment offices and faculty development offices are blurring, as employees of both focus on student learning (Kinzie, Landy, Sorcinelli, & Hutchings, 2019). To our knowledge, no formal framework of the necessary knowledge, skills, and attitudes for conducting this work has been assembled. This article aims to provide such a framework in order to develop a common language for communicating about student learning outcomes assessment and fostering high-quality work in our field.

Articulation of learning outcomes provides a natural starting point when planning educational opportunities. Professionals in both assessment (e.g., Suskie, 2018) and faculty development (e.g., McTighe & Wiggins, 2004) typically endorse a backwards design approach to the planning of learning opportunities, which begins by identifying the end goal of what the learner should know, think, or do as a result of the learning opportunity. All subsequent activities and assessments should then align with the desired student learning articulated in the student learning outcomes (McTighe & Wiggins, 2004). We argue that an identical approach—beginning with the articulation of learning outcomes and then engaging in backward design—provides a necessary foundation for building effective assessment PD for faculty members.

We expect that this process will be aided by the development of a set of learning outcomes for the field of assessment. The literature contains limited examples of the use of backward design in assessment PD offerings. For example, Burrack and Urban (2014)

CORRESPONDENCE

Email
horstsj@jmu.edu

broadly stated the following learning outcomes for assessment-related PD offerings at Kansas State University:

> Participants will possess a student-centered philosophy and knowledge of techniques to improve assessment processes. Participants will demonstrate skills in specific topics of need, such as learning to write SLOs, developing and using rubrics, and assessment planning and mapping. Participants will engage in collaborations with both internal and external partners to implement ideas beyond one department or institution. (p. 6)

We assume that many PD opportunities begin with a similar process; however, the field has not agreed upon a common set of outcomes. The lack of a common framework likely contributes to unnecessary work as campus assessment offices work to develop intentionally designed PD offerings. Each office must locally develop their own outcomes before beginning to plan their PD interventions. An exception may reside in the field of student affairs. There are, for example, several sets of standards for student affairs professionals that outline the knowledge, skills, and attitudes required to conduct high-quality assessment. The *Assessment Skills and Knowledge* (ASK) *Standards* (ACPA, 2006) includes skill descriptions that begin with the phrase "ability to" followed by a specific assessment-related skill. For example, *ASK Standards* (ACPA, 2006) Content Standard 2 ("Articulating Learning and Development Outcomes") is, "Ability to articulate intentional student learning and development goals and their related outcomes" (ACPA, 2006, p. 5). This could be easily adapted into a learning outcome for a PD activity.

In addition to the *ASK Standards* (ACPA, 2006), student affairs professionals are also held to the *Professional Competency Areas for Student Affairs Educators* (referred to as "*Professional Competencies*"; ACPA & NASPA, 2015; 2016), which include two competency areas related to outcomes assessment: the *Assessment, Evaluation, and Research* skill area and the *Student Learning and Development* skill area. The *Professional Competencies* (ACPA & NASPA, 2015; 2016) provide a bank of competencies for student affairs professionals in order to do effective assessment at foundational, intermediate, and advanced levels. Although not written specifically as outcomes, the *ASK* Standards (ACPA, 2006) and Professional Competencies (ACPA & NASPA, 2015; 2016) provide well-thought-out listings of skills necessary for conducting high-quality assessment. Moreover, although developed specifically for and by student affairs professionals, both sets of standards (ACPA, 2006; ACPA & NASPA, 2015) easily transfer to the academic "side of the house" (Finney & Horst, 2019, p. 311).

Both assessment and faculty developers would benefit from a general framework for assessment-related PD learning outcomes. Therefore, the purpose of the current manuscript is to provide a structured framework of assessment learning outcomes, the *Assessment Skills Framework (ASF)*, that includes faculty learning outcomes at novice, intermediate, and advanced levels.

Assessment Skills Framework (ASF)

In order to organize effective PD opportunities, it is key to identify and articulate the knowledge, skills, and attitudes that are important to effective assessment work for faculty members across campus. To this end, we created the *ASF*, which outlines 33 knowledge, skill, and attitude domains foundational to quality assessment practice. We characterized each domain with specific student learning outcomes. Appendix A contains a copy of the *ASF*. The knowledge, skills, and attitudes contained in the framework are necessary for professionals or practitioners conducting assessment work within their departments, offices, or divisions; however, they are not necessarily sufficient. Requirements will vary widely across institutions, positions, and portfolios of responsibility. Similarly, the level at which each of the elements in the framework should be developed will vary depending on the needs of their context.

The *ASF* was developed over the course of two years by a team of assessment professionals (faculty and graduate students) at our institution, James Madison University. During its development, the document was reviewed and revised by over a dozen assessment

The lack of a common framework likely contributes to unnecessary work as campus assessment offices work to develop intentionally designed PD offerings

and measurement professionals and graduate students within our institution's assessment office and the university assessment advisory council. The initial development of the *ASF* involved review of learning outcomes for existing PD opportunities at our institution (e.g., Waterbury, Holzman, Perkins, & Ames, 2017), assessment (e.g., Suskie, 2018), change agent (e.g., Ottaway, 1983), and cultural responsiveness (e.g., Montenegro & Jankowski, 2017) literature, and existing student affairs standards (ACPA, 2006; ACPA & NASPA, 2015; 2016). Although certain specialized assessment-related skills are absent (e.g., technical statistical skills), the *ASF* is quite comprehensive. Therefore, a single faculty member would not be expected to possess all skills. Similarly, no given PD offering can be expected to cover all skills. The information in the *ASF* can serve as a starting-point for the backwards design process when planning assessment-related PD opportunities. The *ASF* is intended as a bank from which assessment professionals can choose and adapt when planning PD opportunities.

Because people enter assessment practice at different levels of preparation, PD opportunities may be aligned to different skill levels. Therefore, the framework includes three levels of outcomes: novice, intermediate, and advanced. The *ASF* begins with a description of the knowledge, skills, and attitudes attributed to each level. Specifically, someone at the novice level is able to provide basic explanations of assessment concepts and apply that knowledge to hypothetical examples devoid of context and real-world complexity. However, the novice may express some anxiety about applying knowledge to practice. People at the intermediate level are able to provide a more thorough explanation of assessment concepts than someone at the novice level. At the intermediate level, people begin to think flexibly about assessment practice, and are able to evaluate context and determine appropriate use of their knowledge when making assessment-related decisions. Faculty at the intermediate level successfully apply their knowledge to real-life assessment projects. Relative to those at the novice level, faculty at the intermediate level express greater confidence in contributing to assessment projects but may express anxiety at the thought of taking full responsibility for parts of the assessment process. People at the advanced level are described as being able to provide nuanced explanations of assessment concepts. These individuals use reflective thinking about their assessment practice, resulting in the generation of new knowledge or useful alternative conceptions about assessment processes. These people can help, lead, encourage, or act as change agents to effectively integrate assessment into the institutional culture.

> **Because people enter assessment practice at different levels of preparation…the framework includes three levels of outcomes: novice, intermediate, and advanced.**

The *ASF* is organized into ten categories of domains. For most categories, the document provides knowledge/skills/abilities at the three skill levels (i.e., novice, intermediate, and advanced). The ten categories of domains include:

- Prerequisite Knowledge

- Foundational Assessment Knowledge and Skills

- Skill Area 1: Specify Student Learning Outcomes

- Skill Area 2: Create and Map Programming to Outcomes

- Skill Area 3: Select and Design Instruments

- Skill Area 4: Examine Implementation Fidelity

- Skill Area 5: Collect Outcomes Information

- Skill Area 6: Analyze Data, Interpret and Report Results, & Maintain Information

- Skill Area 7: Use Results to Improve Student Learning

- Skill Area 8: Assessment in Practice–Additional Skills for Assessment

Note that Skill Areas 1 through 7 align with the assessment cycle followed by most assessment offices, in some form or another (e.g., Suskie, 2018). Each Skill Area contains domains. For example, Skill Area 3, Select and Design Instruments, includes the domains of 1) evaluating instruments–alignment; 2) evaluating instruments–context and resource considerations; 3) evaluating instruments–reliability and validity; 4) designing selected response measures; 5) designing non-cognitive/attitudinal measures; and 6) designing performance assessment measures.

Skill Area 8, Assessment in Practice–Additional Skills for Assessment, includes several domains that contribute to the building of the institutional milieu. Specifically, Skill Area 8 includes domains on evaluating the quality of an assessment plan, promoting value for assessment, and promoting ethics, diversity, and inclusion. Although these topics are not explicitly part of commonly used assessment cycles, we recognize these areas as important for quality student-learning-focused assessment. Another important element of the *ASF* that reaches past the standard assessment cycle is the concept of the change agent. This person is a "visionary/believer" (Jankowski & Slotnick, 2015, p. 93) who can influence others' value for assessment. This role is critical to developing and sustaining assessment practices across an institution (Ariovich, Bral, Gregg, Gulliford, & Morrow, 2018). Not all professional development opportunities need to emphasize the creation of change agents, but their cultivation is a critical step in moving from an institution that conducts assessment to an institution that values assessment and uses results to improve student learning. Therefore, PD offerings must be equipped to create new change agents and support those who already exist across a campus.

An Example of Applying the ASF

We have applied the *ASF* for numerous purposes, including self-reflection and discussion with graduate students who study assessment. However, we would like to focus on the example of a specific PD offering at James Madison University that illustrates the benefit of applying the *ASF* in assessment work. Each year, we offer several week-long hands-on Assessment 101 workshops for faculty and staff who wish to (or are assigned to) engage in assessment. In our yearly workshop planning, we engage in an ongoing backwards-design process (McTighe & Wiggins, 2004). Because the workshop was created prior to the development of the *ASF*, it was originally mapped to a set of outcomes specifically written for the workshop. However, since the development of the *ASF*, all activities and assessment items included in the workshop have been backwards-mapped to the *ASF*. The learning outcomes, then, guide the entire curriculum and the assessment of participant learning. We then use assessment data to determine which outcomes are or are not met by the participants. Because all activities and assessments are tightly aligned to the *ASF* learning outcomes, we can use the assessment results to improve future offerings of the workshop. By focusing the PD opportunity on the outcomes specified in the *ASF*, we are able to engage in an ongoing cycle of assessment and use of results for our PD offerings.

Another use of the *ASF* at JMU is what we loosely refer to as a "fit-gap" analysis. Our institution's assessment office facilitates a wide array of professional development activities. However, we had little coordination between the PD offerings. We were unsure which skills and attitudes were fostered by our PD opportunities, and it was unclear whether we were providing opportunities that best targeted the skills we thought were most important for faculty to develop. Therefore, following the development of the *ASF*, we asked 10 assessment graduate program faculty experts to rate their perceived importance of each of the *ASF* skills for faculty who are learning to do quality assessment. In another activity, we asked assessment professionals to indicate what PD activities they offer, to what audience (i.e., academic affairs or student affairs), and indicate the skill level(s) at which they are offered. By combining faculty perceptions of each skill's importance with information about current PD activities, we were able to identify gaps that needed to be filled in our existing PD offerings. The process enabled us to plan the next level of PD offerings at the intermediate to advanced level and provided the learning outcomes from which to begin the process.

Conclusion

As we work to professionalize assessment and assessment-related PD opportunities, a carefully articulated set of skills provides competencies for faculty entry into assessment. Similar to ways in which the student affairs standards are employed, the *ASF* can provide learning outcomes for PD opportunities, contribute to position descriptions, frame conference offerings, and offer a tool for personal self-reflection (Arminio, 2009; Arminio & Gochenauer, 2004; Finney & Horst, 2019).

> **The *ASF* can provide learning outcomes for PD opportunities, contribute to position descriptions, frame conference offerings, and offer a tool for personal self-reflection.**

High-quality professional development for faculty and staff is key to developing assessment capacity in higher education (Jankowski, Timmer, Kinzie, & Kuh, 2018). The *ASF* has provided us with a vision of a campus engaged with the assessment process. In order for faculty to adopt high-quality assessment practices, they need access to high-quality assessment PD. In order to create PD offerings that both fit the needs of assessment practitioners and meet the standards of assessment professionals, we must have a common framework of assessment knowledge, skills, and attitudes. Working from a common framework allows us to clearly scaffold PD experiences, ensuring that PD offerings meet the range of faculty.

Increasing the quality of assessment practice in higher education requires formalization of skillsets. In order to ensure that faculty conducting assessment work have the tools they need to do their work well, we need to agree upon a set of knowledge, skills, and attitudes essential to that work. In other words, we need to clearly define *what* these individuals need to know, think, and do in order to conduct their work. However, outlining a framework is only the first step in building assessment skills across university campuses. PD offerings that provide faculty with opportunities to master the knowledge, skills, and attitudes outlined in the framework must be made available. We must also provide answers to the question: *How* can faculty gain the tools necessary for conducting quality assessment work? The current manuscript described the "what" in offering a framework of learning outcomes for assessment-related professional development opportunities. An upcoming *Research and Practice in Assessment* manuscript offers suggestions for "how" the field may consider professionalizing learning opportunities.

In order to ensure that faculty conducting assessment work have the tools they need to do their work well, we need to agree upon a set of knowledge, skills, and attitudes essential to that work.

References

American College Personnel Association (2006). *ASK standards: Assessment skills and knowledge content standards for student affairs practitioners and scholars*. Washington, D.C.: Author.

American College Personnel Association & National Association of Student Personnel Administrators (2015). *ACPA/ NASPA professional competency areas for student affairs educators*. Washington, DC: Authors.

American College Personnel Association & National Association of Student Personnel Administrators (2016). *ACPA/ NASPA professional competencies rubrics*. Washington, DC: Authors.

Ariovich, L., Bral, C., Gregg, P., Gulliford, M., & Morrow, J.A. (2018, May). *The Assessment Profession in Higher Education: Addressing the Varied Professional Development Needs of Practitioners*. Lexington, KY: Association for the Assessment of Learning in Higher Education and Watermark.

Arminio, J. (2009). Chapter Ten: Applying professional standards. In G. S. McClellan & J. Stringer (Eds.), *The handbook of student affairs administration* (3rd ed., pp. 187–205). San Francisco, CA: Jossey-Bass.

Arminio, J., & Gochenauer, P. (2004). After 16 years of publishing standards, do CAS standards make a difference? *College Student Affairs Journal*, *24*, 51-65.

Burrack, F., & Urban, C. (2014). Strengthening foundations for assessment initiatives through professional development. *Assessment Update*, *26*(6), 5-12.

Finney, S. J., & Horst, S. J. (2019): Standards, standards, standards: Mapping professional standards for outcomes assessment to assessment practice. *Journal of Student Affairs Research and Practice*, 56, 310-325.

Hutchings, P. (2010, April). *Opening Doors to Faculty Involvement in Assessment*. (Occasional Paper No. 4). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Jankowski, N. A., & Slotnick, R. C. (2015). The five essential roles of assessment practitioners. *Journal of Assessment and Institutional Effectiveness*, 5, 78-100.

Jankowski, N. A., Timmer, J. D., Kinzie, J., & Kuh, G. D. (2018, January). *Assessment that matters: Trending toward practices that document authentic student learning*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Kinzie, J., Landy, K., Sorcinelli, M. D., & Hutchings, P. (2019). Better together: How faculty development and assessment can join forces to improve student learning. *Change: The Magazine of Higher Learning*, *51*, 46-54.

Montenegro, E., & Jankowski, N. A. (2017, January). *Equity and assessment: Moving towards culturally responsive assessment*. (Occasional Paper No. 29). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

McTighe, J., & Wiggins, G. (2004). *Understanding by design: Professional development workbook*. Alexandria, VA: Association for Supervision and Curriculum Development.

Nicholas, M. C., & Slotnick, R. C. (2018, April). *A Portrait of the Assessment Professional in the United States: Results from a National Survey*. (Occasional Paper No. 34). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Ottaway, R. N. (1983). The change agent: A taxonomy in relation to the change process. *Human Relations*, *36*(4), 361-392.

Suskie, L. (2018). *Assessing student learning: A common sense guide*. (3rd Ed.). San Francisco, CA: Jossey-Bass.

Waterbury, G., Holzman, M., Perkins, B., & Ames, A. (2017, Fall). The crucial role of faculty development in promoting quality assessment practices. *AALHE Intersection*, 17-21.

Appendix A: Assessment Skills Framework

Competency Level Overview

Competency Levels	General Characterization of Knowledge, Attitudes, and Skills	Cognitive Level (Bloom's Taxonomy)
Novice	**Knowledge**: • Able to provide basic explanations of assessment concepts. • Characterized by more rigid or "black and white" thinking related to assessment practice. **Attitudes**: • May have anxiety about applying knowledge to practice. **Skills**: • Can apply assessment knowledge to simplified/hypothetical examples devoid of context and "messiness".	Low-level Bloom's Taxonomy: **Describe**, Identify, Define, Distinguish, Recognize
Intermediate	**Knowledge**: • Able to provide thorough explanations of assessment concepts. • Characterized by more flexible thinking related to assessment practice. • Can evaluate context and determine the appropriate use of knowledge/skills when making assessment-related decisions. **Attitudes**: • Confidence in ability to participate in an active project with some leadership or guidance. • May have anxiety about leading or taking full responsibility for parts of the assessment process. **Skills**: • Can apply assessment knowledge to the planning and implementation of real-life assessment projects.	Mid-level Bloom's Taxonomy: **Apply**, Conduct, Demonstrate, Analyze, Compare/Contrast, Evaluate, Examine, Integrate
Advanced	**Knowledge**: • Able to provide nuanced explanations of assessment concepts. • Characterized by reflective, insightful thinking related to assessment practice. • Generates new knowledge and useful alternative conceptions of assessment processes (e.g., may no longer view the assessment process as "linear"). **Attitudes**: • Confidence in ability to lead various parts of the assessment process. • Embodies the spirit of assessment & promotes the value of assessment. **Skills**: • Can help or lead others in completing assessment tasks. • Can serve as a change agent or leader in assessment to effectively integrate assessment into the culture of a program or institution.	High-level Bloom's Taxonomy: **Design**, Develop, Propose, Plan, Synthesize, Review, Anticipate, Solve, Reflect

Prerequisite Knowledge

Trait/Domain	Knowledge, Attitudes, Skills
Knowledge of program/understanding of context	When applicable: • Articulates the mission and vision of the program. • Articulates the mission and vision of the institution/department/office within which the program is situated. • Describes the history of the program (e.g., impetus for program, changes to programming/curriculum, previous assessment results). • Describes student population served by program; aware of the needs of these students. • Describes departmental resources (e.g., staffing, time, money, knowledge and skills of colleagues, buy-in of colleagues) that may impact assessment practice. • Describes sociopolitical factors (e.g., departmental and institutional hierarchy, accreditation/accountability requirements) that may impact assessment practice.

Foundational Assessment Knowledge and Skills

Trait/Domain	Novice	Intermediate	Advanced
Distinguishing student learning outcomes assessment from other assessment and evaluation processes	• Differentiates between student learning outcomes assessment and other types of assessment (e.g., needs assessment, use assessment, satisfaction assessment). • Differentiates between student learning outcomes assessment and other aspects of program evaluation and/or program review. • Describes how student learning outcomes assessment is situated within program evaluation.	• Considers the ways in which student learning outcomes assessment is situated within their own program's program evaluation/review process.	• Creates a program evaluation/review process that integrates student learning outcomes assessment.
Describing the assessment cycle	• Provides basic descriptions of each step of the assessment cycle (e.g., can define implementation fidelity and describe the purpose of collecting implementation fidelity data).	• Provides detailed descriptions of each step of the assessment cycle (e.g., can describe how to collect implementation fidelity data).	• Provides nuanced descriptions of the steps of the assessment cycle, including the ways in which various parts of the assessment cycle interact with one another (e.g., can describe how articulating program theory in Step 2 informs the development of an implementation fidelity checklist, and how implementation fidelity data informs how assessment results are interpreted in Step 5). • Explains the steps of the assessment cycle to others.
Identifying the purposes of assessment	• Can describe the differences between assessment for accountability and assessment for improvement.	• Can identify factors within their own program that may indicate the primary driver for assessment is accountability rather than improvement or vice versa.	• Promotes assessment for improvement vs. assessment for accountability within their own program and/or across the institution.

Foundational Assessment Knowledge and Skills, Cont.

Trait/Domain	Novice	Intermediate	Advanced

Distinguishing between various levels of assessment	• Can identify the various levels of assessment (e.g., section, course, program, department, division, institution) and describe how they are situated within one another.	• Identifies relevant assessment considerations given the level at which assessment is conducted (e.g., curricular alignment, implementation fidelity, reporting).	• Facilitates multi-level communication and engagement to promote alignment of assessment activities across levels, including within the following domains:
			  • Developing SLOs
			  • Collecting data
			  • Reporting results
			  • Using results
			• Navigates the socio-political structure of their own program/institution to promote multi-level assessment that provides fruitful information at each level.

Skill Area 1: Specify Student Learning Outcomes

Trait/Domain	Novice	Intermediate	Advanced
Developing student learning outcomes	• Recognizes the need for clear student learning objectives. • Describes various taxonomies used to classify student learning objectives (e.g., Blooms, Finks, SOLO). • Writes a student learning objective and identifies a measurable verb at an appropriate level of Bloom's or another taxonomy.	• Writes SLOs including appropriate level of Bloom's or another taxonomy for their own program with some guidance. • Independently identifies common issues with SLOs (e.g., double-barreled SLOs, vague language, unmeasurable verbs, or inappropriate level of verb usage). • With guidance, considers and incorporates relevant theories in the development of SLOs.	• Independently develops SLOs for their own program according to best practice and relevant theories. • Captures the spirit of the program in the SLOs. SLOs are aligned with the mission and vision of the program.

Skill Area 2: Create and Map Programming to Outcomes

Trait/Domain	Novice	Intermediate	Advanced
Developing theory-based programs	• Describes the concept of program theory. • Recognizes the importance of theory-based programming. • Identifies components of a logic model for program development (e.g., inputs, activities, outputs, short-term outcomes, intermediate outcomes, long-term outcomes).	• Clearly articulates program theory underlying a given program or intervention. • With assistance, develops a logic model to aid in articulating program theory for a given program or intervention. • Identifies various bodies of literature to assist with program development, such as academic motivation, learning theories, or cognitive theories.	• Integrates clearly articulated theory when developing programs. Independently develops clear logic models underlying theory-based programs. • Thoroughly explains why programming should produce the intended outcome(s) based on clearly articulated program theory. • Directs others to appropriate bodies of literature to aid in building programs, such as academic motivation, learning theories, or cognitive theories. • Synthesizes multiple sources of information when developing programming (e.g., construct theories, motivation, learning theories, cognitive theories).
Mapping of SLOs with curriculum	• Recognizes the importance of mapping curriculum to SLOs. • Given a clearly articulated set of SLOs and curriculum, maps curriculum to SLOs.	• Clearly articulates their own SLOs and program curriculum. • With assistance, maps their own program curriculum to SLOs.	• Independently and effectively maps their own program curriculum to SLOs. • Uses curriculum map to note gaps in programming or redundant programming. Collaborates with others to generate plans to close identified gaps or reduce redundancies.

Skill Area 3: Select and Design Instruments

Trait/Domain	Novice	Intermediate	Advanced
Evaluating instruments – alignment	• Describes basic types of instruments and their intended uses (e.g., indirect/direct, selected response/constructed response, cognitive/non-cognitive). • Matches appropriate instrument to SLO on a superficial level.	• Describes different item types that can be used within each type of instrument (e.g., multiple choice, true/false, and matching items as types of selected response items). • Effectively matches appropriate instruments to SLOs. • Chooses an instrument appropriate for the cognitive process indicated in a given SLO.	• Critically considers options to select appropriate measures for their own assessment plan and provides a well-developed rationale. • Considers alignment of cognitive processes and breadth and depth of content coverage when evaluating instruments for a specific SLO.
Evaluating instruments – context & resource considerations	• Describes the pros and cons of selecting an existing instrument versus developing a new instrument. • Describes the advantages and disadvantages of using each type of instrument (e.g., indirect/direct, selected response/constructed response, cognitive/non-cognitive). • Describes the pros and cons of using commercial versus non-commercial instruments.	• Weighs pros and cons of selecting vs. designing instruments, using commercial vs. non-commercial instruments, and using each type of instrument into evaluating an instrument for their own assessment plan. • Identifies the various resource and contextual considerations necessary to evaluate an instrument within their own assessment plan (e.g., time, population served, data access, scoring).	• Effectively evaluates instruments for their own assessment plan based on knowledge of context and resource considerations within their own program.
Evaluating instruments – reliability & validity considerations	• Acknowledges the importance of considering reliability and validity when selecting a measure. • Describes commonly used types of reliability and validity evidence.	• Identifies specific reliability and validity evidence appropriate for different types of instruments (e.g., rater agreement for performance assessments). • Explains the importance of rater training and reliability issues related to rater agreement.	• Evaluates appropriate reliability and validity evidence when selecting a measure. Seeks help in conducting reliability analyses for their own data, if necessary. • Conducts literature search for validity evidence when appropriate.

Skill Area 3: Select and Design Instruments, Cont.

Trait/Domain	Novice	Intermediate	Advanced
Designing selected response measures	• Identifies components of a basic multiple-choice item (i.e., stem, alternatives, correct answer, distractors). • Identifies best practices for constructing selected response measures, including: • Developing clear, measureable objectives • Using a test blueprint • Writing items • Piloting items with students and revising	• Writes items according to best practices for a selected response measure that is appropriately mapped to a set of their own SLOs. • Applies best practices for constructing selected response measures to own program.	• Regularly writes or revises items that are appropriately mapped to their program SLOs and program curriculum. • Leads others in applying best practices for constructing selected response measures.
Designing non-cognitive/attitudinal measures	• Identifies characteristics of non-cognitive/attitudinal instruments (e.g., inclusion of negatively worded items, variety of response options, response option labels). • Identifies best practices for constructing non-cognitive/attitudinal measures (e.g., avoiding loaded items, avoiding double-barreled items)	• Follows best practices for constructing non-cognitive/attitudinal measures within their own program.	• Leads others in applying best practices for constructing selected response measures. • Incorporates concerns about student motivation into non-cognitive assessment design.

Skill Area 3: Select and Design Instruments, Cont.

Trait/Domain	Novice	Intermediate	Advanced

Designing performance assessment measures	• Identifies basic rubric components (e.g., elements, rating scale, scoring criteria).
	• Distinguishes between holistic and analytic rubrics and identifies the advantages and disadvantages for each.
		• Develops an appropriate prompt or task that will be rated by a rubric for their own program.
		• Designs a rubric mapped to their own program's student learning outcome(s) that appropriately implements the following skills:
		  • Selects the appropriate rubric type
		  • Effectively describes elements/traits
		  • Determines rating scale and score levels
		  • Develops scoring criteria
		• Assists in conducting a rater training.
			• Independently designs and implements a rubric within their own assessment plan.
			• Leads others in designing a prompt or task and rubric mapped to student learning outcome(s).
			• Consults with experts to design and conduct a rater training.

Skill Area 4: Examine Implementation Fidelity

Trait/Domain	Novice	Intermediate	Advanced
Designing implementation fidelity studies	• Defines implementation fidelity and explains the rationale behind implementation fidelity research. • Explains the five components of implementation fidelity (i.e., program differentiation, adherence, quality, exposure, responsiveness).	• With assistance, creates an implementation fidelity plan for their own program (e.g., creates materials for recording data on each of the five components, determines method of collection).	• Develops, conducts, and/or leads an implementation fidelity study for their own program. • Articulates the importance of implementation fidelity to others who express resistance (change agent).
Collecting implementation fidelity data	• Explains the pros and cons of collecting implementation fidelity data via self-report versus observation.	• Collects or aids in collection of implementation fidelity data for their own program.	• Develops a sustainable plan for ongoing collection of implementation fidelity data for their own program and/or leads others in the development of a sustainable data collection plan.
Interpreting implementation fidelity results	• Recognizes the importance of evaluating data for each of the five components of implementation fidelity. • Explains how implementation fidelity data are interpreted in conjunction with student learning outcomes assessment findings (e.g., how low implementation fidelity affects interpretation of poor assessment results).	• With assistance, explains findings from each of the five components for their own program. • Interprets student learning outcomes assessment data for their own program in light of implementation fidelity results.	• Makes recommendations for program improvement, based on integration of student learning outcomes assessment and implementation fidelity results.

Skill Area 5: Collect Outcomes Information

Trait/Domain	Novice	Intermediate	Advanced
Planning a data collection design	• Recognizes the types of questions that can be asked when developing a research design (e.g., differences, relationships, change, competency). • Defines validity threats in the context of research design (e.g., internal validity, external validity, statistical conclusion validity).	• Selects an appropriate research design for their own program, based upon the types of research questions asked (e.g., differences, relationships, change, competency) and desired inferences. • Evaluates the threats to validity associated with the research design chosen for their own program. • Compares and contrasts the appropriate inferences that can be drawn from a chosen research design. Acknowledges the limitations of reasonable assessment designs in which random assignment is not feasible.	• Anticipates the appropriate inferences that may be drawn from a research design and uses that information to propose a strong and clearly articulated rationale for their own (or others') research design. • Encourages others to reflect upon the appropriate inferences and threats to validity associated with various research designs. • When appropriate, challenges others' rigid views of causality. Clearly articulates the need for tentative conclusions when others make causal claims.

Skill Area 5: Collect Outcomes Information, Cont.

Trait/Domain	Novice	Intermediate	Advanced
Selecting a data collection method	• Recognizes the importance of collecting meaningful and credible data. • Recognizes the importance of factors such as data collection mode, frequency, and sample size in choosing a data collection method. • Identifies the importance of specifying who, how, what, where, and when related to data collection. • Recognizes factors related to student motivation and how they apply to data collection methods. • Identifies factors to consider when designing a survey (e.g., question format, timing, layout), including both electronic and paper/pencil formats.	• Develops a data collection plan for their own program, considering costs and benefits of data collection mode, frequency, and sample size • When appropriate, evaluates sampling methods (e.g., stratified sampling) in order to address the representativeness of the sample. • Evaluates factors that may contribute to students' motivation to complete measures. • Creates surveys in a variety of formats (e.g., constructed response, Likert-type, technology-enhanced). • Considers usability of the survey.	• Maintains a sustainable and efficient data collection plan for their own program assessment (or leads others in doing so). • Integrates knowledge of sampling methods into data collection plan (e.g., stratified sampling, multistage sampling) when appropriate. • Reflects upon the representativeness of the sample. • Makes theory-based recommendations for increasing students' motivation and encourages others to consider student motivation when designing a data collection plan. • When appropriate, takes advantage of electronic survey capabilities (e.g., skip logic, technology-enhanced items).

Skill Area 6: Analyze Data, Interpret and Report Results, & Maintain Information

Trait/Domain	Novice	Intermediate	Advanced

Item analysis	• Understands that items vary in quality.
	• Identifies general, non-technical characteristics of high- and low-quality items.
		• Investigates the quality of individual items using basic item analyses, including:
		  • Item difficulty
		  • Item discrimination
			• Investigates the quality of items within the context of a scale or measure, including:
			  • Inter-item correlations
			  • Internal consistency reliability (coefficient alpha)
			  • Alpha if item is deleted

Reliability of score inferences	• Recognizes reliability as the "relative consistency of responses" and provides basic definitions of different types of reliability estimates (e.g., internal, test-retest, inter-rater, alternate forms).	• Explains the concept of reliability and applies it to their own program assessment.
		• Describes the following forms of reliability estimates, when each would be appropriate, and basic interpretation of a given numeric estimate:
		  • Internal consistency (coefficient alpha, split-half reliability)
		  • Inter-rater reliability
		  • Test-retest reliability
		  • Alternate forms reliability
			• Evaluates reliability evidence associated with their own assessment data and assists others in evaluating results from their programs.
			• Given a journal article or output that includes reliability estimates, interprets the estimate.
			• Given a measurement context, recommends appropriate means for investigating reliability.

Skill Area 6: Analyze Data, Interpret and Report Results, & Maintain Information, Cont.

Trait/Domain	Novice	Intermediate	Advanced

Validity of score inferences	• Defines validity as the extent to which evidence supports the interpretations made from the data.
	• Conceptualizes validity as a continuum, rather than an absolute property (i.e., all-or-none).
	• Identifies the importance of ongoing validity research.
	• Defines validity within the context of score inferences, rather than the context of the measure itself.
		• Accurately explains the concept of validity and applies it to their own program assessment.
		• Describes the sources of validity evidence outlined in the Standards for Educational and Psychological Testing:
		  • Evidence based on test content
		  • Evidence based on response processes
		  • Evidence based on internal structure
		  • Evidence based on relations to other variables
		  • Evidence for validity and consequences of testing
		• Identifies and debunks basic validity myths and misconceptions.
			• Evaluates validity evidence associated with their own assessment data.
			• Given an assessment scenario, recommends appropriate methods of investigating validity.
			• Identifies research articles or assessment plans in which strong validity evidence is presented, and/or which explicate a strong rationale for collecting specific validity evidence.

Analyzing data – quantitative	• Understands and provides basic interpretations of common descriptive statistics, including measures of central tendency, variability, and association. • Identifies research questions that can be best investigated using quantitative research methods.	• Selects appropriate descriptive statistics for a given research question. • Generates effective quantitative research questions. • Conducts (basic) descriptive statistical analyses, including measures of central tendency, variability, and association. • Defines and interprets the results of (basic) inferential statistics, such as t-tests, bivariate linear regression, and ANOVA.	• Selects appropriate (basic) inferential statistics for a given research question. • Conducts (basic) inferential statistical analyses, including t-tests, bivariate linear regression, and ANOVA. • Identifies the ways in which sample size can influence analytical findings. • Distinguishes between statistical and practical significance when interpreting results.

Skill Area 6: Analyze Data, Interpret and Report Results, & Maintain Information, Cont.

Trait/Domain	Novice	Intermediate	Advanced

Analyzing data – qualitative	• Identifies differences between major approaches to qualitative research:
	  • Narrative research
	  • Phenomenological inquiry
	  • Grounded theory
	  • Ethnography
	  • Case study
	• Identifies research questions that can be best investigated using qualitative research methods.
		• Develops basic coding schemas and applies them to qualitative data.
		• Identifies important considerations in developing an interview protocol.
		• Generates effective qualitative research questions.
		• Selects appropriate qualitative approaches for a given research question or scenario.
			• When applicable, identifies resources for and plans a qualitative study for their own program's assessment process.
			• Creates effective interview protocols for a given purpose and context.
			• Provides appropriate descriptions and interpretations of qualitative data.
			• When applicable, uses software programs for transcription and analysis of qualitative data (e.g., NVivo, HyperResearch).

Analyzing data – mixed methods	• Identifies research questions that can be best investigated using mixed methods. • Identifies various mixed methods research designs (e.g., sequential explanatory, concurrent triangulation).	• Identifies components of a mixed methods study (e.g., Creswell's components) that includes quantitative, qualitative, and mixed research questions. • Identifies the appropriate type of mixed methods design for a particular research study. • Appropriately displays research methods via design diagrams.	• Identifies resources for and plans a mixed methods study for their own program, writing appropriate quantitative, qualitative, and mixed research questions and diagramming the design. • Integrates quantitative and qualitative data using methods such as joint display tables to inform interpretation.
Displaying data	• Appropriately interprets basic data displays (e.g., bar graphs, histograms, line graphs, scatterplots). • Identifies essential components of effective data displays.	• Identifies and explains common errors in displaying data (e.g., inappropriate axes, missing labels). • Determines the appropriate graph or table for a specific data visualization need. • With assistance, creates basic data displays using their own program's assessment data.	• Independently creates data displays that accurately portray their own program's assessment data. • Chooses appropriate data display methods for the type of data collected. • Effectively integrates data displays with text when creating reports.

Skill Area 6: Analyze Data, Interpret and Report Results, & Maintain Information, Cont.

Trait/Domain	Novice	Intermediate	Advanced
Interpreting results	• Recognizes the types of information needed to make accurate inferences about program effectiveness (e.g., implementation fidelity, research design, quantitative/qualitative/mixed methods results).	• Draws narrow and/or limited conclusions based on isolated sources of assessment data. • Lists common mistakes in interpretation and articulates why they are wrong.	• Integrates multiple sources of information to draw nuanced conclusions about program-level outcomes and tell a cohesive story about program effectiveness. • Demonstrates appropriate caution when interpreting results and does not draw spurious conclusions. • Makes recommendations about additional information that could be collected to strengthen the interpretation of results. • Directs others in interpretation of their own assessment results.
Writing reports	• Identifies key components of effective assessment reports. • Identifies intended audience(s) and their relevant considerations and needs.	• With assistance, constructs a report of their own program's assessment plan and findings. • Tailors content, tone, and style of a report to accurately and effectively convey information to a given audience.	• Inspires action, motivates, and tells compelling stories through report writing. • When appropriate, uses reports to move beyond simple accountability to focus on student learning. • Creates nuanced assessment reports with awareness of political and other contextual factors, such as timing of assessment reporting (adapted from ASK Standard 12).
Maintaining information	• Recognizes the importance of maintaining a database that spans across years. • Recognizes the importance of securing sensitive data.		• Designs data maintenance systems that allow for comparison across years. • Considers data security, applicable laws and policies, consistency of coding, and clarity of documentation when collecting and archiving data (adapted from CAS Standard 10).

Skill Area 7: Use Results to Improve Student Learning

Trait/Domain	Novice	Intermediate	Advanced
Using results to improve student learning	• Articulates the importance of using assessment results to make evidence-based changes to programs. • Identifies learning improvement as one of the primary reasons for conducting assessment. • Describes the steps of the simple model for learning improvement (Fulcher et al., 2014): • Identifies an appropriate SLO • Conducts baseline assessment • Proposes and implements a coordinated intervention • Conducts a post-test assessment • Creates a data collection/intervention implementation timeline • Articulates the importance of each step of the simple model (i.e., the impact of removing a given step). • Differentiates between a change and an improvement in the context of a fictional program assessment process.	• Identifies key indicators of program readiness to embark on a learning improvement project (e.g., administrative support, faculty cohesion, quality assessment practices). • Accurately evaluates their own program's readiness to embark on a learning improvement project. • Feels confident serving as a participant on a learning improvement team.	• Integrates knowledge of one's own program (e.g., program theory, implementation fidelity results, outcomes data) to formulate an evidence-based plan for using results to improve the program. • Serves as the lead on a learning improvement project.

Skill Area 8: Assessment in Practice - Additional Skills for Assessment

Trait/Domain	Novice	Intermediate	Advanced
Evaluating the quality of an assessment plan	• Articulates best practices for assessment at each step of the assessment cycle.	• Identifies the strengths and weaknesses of their own program's assessment process at each step of the assessment cycle (e.g., strong SLOs, poor research design, no implementation fidelity information).	• Develops and implements a plan to improve their own program's assessment process based on knowledge of best practices and identified weaknesses.
Promoting value for assessment	• Recognizes the value of assessment for program improvement and increasing student learning and development. • Recognizes the need for assessment education (generally and personally). • Identifies common barriers (e.g., organizational, attitudinal, political) to conducting high-quality assessment and using results.	• Communicates the value of assessment to others within their own program and/or institution. • Independently seeks additional training related to assessment. • Collaborates with others and shares ideas related to assessment practice. • Identifies the primary barriers to conducting high-quality assessment and using results within their own program and/or institution.	• Creates a positive climate within their own program and/or institution that encourages and supports assessment practice. • Encourages others in their assessment practices and serves as an accessible resource. • Develops strategies to overcome barriers to conducting high-quality assessment and using results within their own program and/or institution.

Skill Area 8: Assessment in Practice - Additional Skills for Assessment, Cont.

Trait/Domain	Novice	Intermediate	Advanced

Promoting ethics, diversity, and inclusion	• Recognizes responsibilities related to ethics, diversity, and inclusion as they relate to instruments, technology, confidentiality, and reporting.
	• Expresses awareness of IRB protocols and human research principles.
		• Applies principles of ethics, diversity, and inclusion as they relate to assessment:
		  • Instruments: Reviews assessment instruments' inclusivity and accessibility. Considers the needs of students with disabilities.
		  • Technology: Considers accessibility of technology when collecting assessment data.
		  • Confidentiality: Protects confidentiality of data, when appropriate.
		  • Reporting: Avoids the misrepresentation of student groups in reporting.
		• Has completed IRB training and is familiar with IRB submission procedures.
			• Serves as a change agent by promoting accessible and ethical use of instruments, considering the needs of all students, maintaining confidential data, and creating accurate and representative reports.

Abstract

The *Vision 2030* agenda was recently adopted as a roadmap and methodology for developmental and economic action throughout the Kingdom of Saudi Arabia. *Vision 2030* includes support for universities' academic and administrative operations through the collection, rigorous analysis, and reporting of a wide range of data. Prefaced with an overview of the Saudi economic, policy, and educational landscape, the paper's main contribution is a case study of Imam Abdulrahman Bin Faisal University (IAU), chosen because of its recent attempt to institutionalize academic assessment protocols, procedures and culture. It institutionalized a directorate focused on academic assessment, launched a Decision Support Unit dashboard, and developed key performance indicators (KPIs) to assess students' academic performance, course performance and employability. The IAU has since become a regional leader in higher education assessment. Emulating their approach affords other Saudi higher education institutions the opportunity to increase Saudi graduates' ability to directly contribute to the country's economy, ultimately promoting economic growth, diversification, and development as envisioned in *Vision 2030*.

AUTHORS

Amani K. Hamdan
Alghamdi, Ph.D.
*Imam Abdulrahman
Bin Faisal University*

Ghazi Alotaibi Ph.D.
*Imam Abdulrahman
Bin Faisal University*

Osama Ibrahim Ph.D.
*Imam Abdulrahman
Bin Faisal University*

Institutional Academic Assessment and Effectiveness in Higher Education: A Saudi Arabia Case Study

*E*very educational system strives to ensure academic quality, enhance students' learning, and make their institutions more accountable to stakeholders. This paper seeks to shed light on a Saudi Arabian (SA) policy program aimed at developing an evidence- and effectiveness-oriented institutional culture that will contribute to increasing the performance of academic institutions in general and faculty members and students in particular. To that end, in the context of the worldwide movement toward institutional academic assessment and effectiveness in higher education, this paper presents a case study of a Saudi university's experience with bolstering and sustaining efforts to obtain external accreditation and move forward on a successful academic trajectory.

Saudi Arabia is a noteworthy case because it is experiencing high unemployment within the context of a bourgeoning academic environment and context. The education system is often chastised for inadequately preparing students for the Saudi labour market (Alrasheedy, 2017; Lindsey, 2010). Universities are expected to come forward to address this issue. One strategy is to focus on assessing the academic effectiveness of higher education (HE) institutions. Prefaced with a literature review capturing an overview of the Saudi economic, policy, and educational landscape, the paper's main contribution is a case study of Imam Abdulrahman Bin Faisal University (IAU) chosen because of its recent attempts to institutionalize academic assessment protocols, procedures, and culture.

CORRESPONDENCE

Email
akhalghamdi@iau.edu.sa

Overview of the Saudi Landscape

Over the last decade, dramatic and positive changes have taken place in the Kingdom of Saudi Arabia (KSA), and many of these changes are happening in the area of education. The nation has witnessed exceptional growth in the number of HE institutions. This phenomenon was recently highlighted by the KSA's Ministry of Higher Education when it reported an 86% expansion in the number of universities from 2006 to 2016 (Pavan, 2016).

In 2016, 1.7 million students (total SA population is 34 million) were enrolled mostly in the Kingdom's 28 public universities with others placed in the 10 private-sector institutions. The latter proportion is projected to account for 12% of all enrollment in 2020 and continues to grow exponentially. Most institutions comprise colleges and departments with a complement of research centers. Universities offer programs ranging from humanities and social sciences to medicine, engineering, science, technology, and business. Despite this range of academic offerings, the Saudi HE system is not keeping up with market demands. Only two Saudi HE institutions received top placement (top 100) in major global ranking systems although *Vision 2030* aspires to raise this to five by 2030. *Vision 2030* also aims to modernize curricula and forge closer links between the academy and industry, which will require a focus on assessments, student outcomes, and institutional capacity (ICEF Monitor, 2018).

> It is important to recognize that the expansion and improvement of HE constitute a need rather than a luxury—in part as a result of the gap between education demand and supply.

The dramatic growth in the number of Saudi HE institutions is consistent with two global trends. First, over the last 20 years, growth in HE enrollments has closely followed world trade growth and has far outpaced the world's GDP growth. Second, this expansion is viewed by many governments as a vehicle for achieving national priorities and contributing to economic growth (British Council, 2012). In Saudi Arabia, the expansion of HE institutions has also been consistent with a significant increase in the number of high school graduates—individuals who are seeking opportunities to complete their education and launch their career (Organization for Economic Cooperation and Development [OECD], 2019). Another factor has been the amalgamation of the country's women's colleges, which had previously been under the auspices of the General Presidency for Girls' Education, the supervisory authority for all aspects of female education from 1960 until 2002.

Recent studies (Hamdan, 2015a; 2015b; 2016; 2017) have identified several factors interacting with the state of HE in Saudi Arabia: high birth rate, high unemployment rate among Saudi females, vast number of expatriate workers (also prevalent in other *Gulf Cooperation Council* [GCC] countries), changes in the global economy and its economic structures and, most recently, the launch of the KSA's new national development plan's *Vision 2030* policy program (Kingdom of Saudi Arabia, 2018; to be discussed). The government's growing investment in human capital and human development is bringing higher expectations to the labor market, which is increasing the level of pressure on Saudi universities to improve the caliber of their graduates.

Compared to a 39% average for OECD countries, only 24% of Saudi citizens graduate from higher education. This trend may be reversing because first-time entry rates into bachelor degree programs rose to 66% in 2017 (higher than the 58% OECD average). That said, SA graduate level enrollments are far below OECD rates, especially master-level, first-time entry rates, 3% compared to 24%. Upon graduation from HE, a quarter (26%) of Saudi graduates is unemployed, which makes unemployment is low compared to partner countries and OECD nations. This number reflects the hard reality that while 84% of male graduates are employed, nearly two thirds (59%) of female tertiary graduates are not (OECD, 2019).

Saudi Arabia as an Oil-Dependent Country

The development of a modern HE system in Saudi Arabia is well underway. The expansion of access to opportunities has occurred rapidly and on a massive scale. The next challenge of "improving the quality of learning and teaching within the universities has been acknowledged" (Alnassar & Dow, 2013, p. 59). It is important to recognize that the expansion and improvement of HE constitute a need rather than a luxury—in part as a result of the gap between education demand and supply associated with the country's massive oil-driven demographic growth. Despite the fact that SA "has amassed great wealth from oil, developments in the country have been subject to the unpredictable cycle of rising and

and falling oil prices" (Alexander, 2000, p. 412). This vulnerability has intensified the pressure on the government to find ways to diversify the economy and thus reduce the level of dependence on the petroleum industry.

KSA is not unique in its drive toward higher levels of economic diversification and performance. This phenomenon is well known around the world as "governments are increasingly looking to the different sectors of higher education to augment learning skills and improve workers' ability to develop and use technology, thus enhancing productivity and strengthening state's economic position" (Alexander, 2000, p. 412). The contemporary world economy is rapidly changing; meaning, national governments will come under more and more pressure to raise their respective economies' level of diversification in order to minimize their level of exposure to unexpected shifts in demand for specific commodities as well as for specific types of goods and services. In Saudi Arabia, that commodity is oil.

Unemployment in KSA

Paradoxically, notwithstanding the massive oil-driven economic expansion, including the impressive growth of the private sector, the KSA continues to be burdened by a high rate of unemployment. According to the KSA General Authority for Statistics (GaStat; 2016), the 2016 rate of unemployment of Saudi nationals was 12.2%. Moreover, nearly 34.5% of all Saudi adult females and 5.9% of all adult males were unemployed. The highest levels of unemployment were found in the youth population as 40% of all citizens under the age of 35 had no stable form of employment (Alrasheedy, 2017). This rate varied among different age groups with the highest being among people aged 20 and 24 (GaStat, 2016).

But unemployment among Saudi nationals is different from unemployment in most other countries, which typically results from poor economic conditions and general poverty. The KSA's problem with unemployment is intertwined with the fact that the majority of private-sector positions are held by expatriates (i.e., people who have left their own country to live in another, often for a prolonged period). Moreover, the preponderance of high-status private-sector positions, which require highly qualified employees, is currently filled by non-Saudis who constitute approximately one third of the country's population. The seriousness of this situation is highlighted by the fact that less than 10% of the private-sector workforce comprised Saudi citizens in 2009, a trend that still continues (Alrasheedy, 2017). The longstanding problem of Saudi graduates' lack of job-related skills is one of the main reasons for the severe underrepresentation of Saudi nationals in the KSA workforce. Indeed, "one of the main issues that the private sector has is the fact that there aren't enough well-trained Saudis for the kinds of jobs that are needed" (Lindsey, 2010, p. 10).

Even though Saudi Arabia faces many challenges related to ensuring that its university graduates are equipped with the skills and knowledge required to succeed in the workforce, it should be remembered that, in this particular regard, the KSA is by no means unique. Many countries around the world—including relatively advanced countries like the United States—are being forced to confront the reality of substantial disparities between what its university students are learning and what its employers are seeking. In the late 1990s, this problem was brought into stark relief by the Boyer Commission (1998), which was charged with investigating and analyzing the U.S. situation. Twenty years ago, the commission found that

> many students graduate having accumulated whatever number of courses is required, but still lacking a coherent body of knowledge, or any inkling as to how one sort of information might relate to others. And all too often they graduate without knowing how to think logically, write clearly, or speak coherently. The university has given them too little that will be of real value beyond a credential that will help them get their first jobs. And with larger and larger numbers of peers holding the same papers in their hands, even that credential has lost much of its potency. (Boyer Commission, 1998, p. 6)

The widespread and pervasive global nature of the problems associated with ensuring that graduates are able to meaningfully contribute to the workforce and economy should help

Many countries around the world—including relatively advanced countries like the United States—are being forced to confront the reality of substantial disparities between what its university students are learning and what its employers are seeking.

Saudi policymakers and educators to avoid becoming pessimistic about the country's current state of affairs and future prospects. They are not in it alone; there is potential and opportunity to address the issue in their contemporary context.

Saudi Higher Education Assessment and Accountability

The KSA government unveiled the *National Transformation Plan* (NTP) on April 25, 2016 as part of the wider "Vision for the Kingdom of Saudi Arabia," otherwise known as *Vision 2030* (Kingdom of Saudi Arabia, 2018). This is both a roadmap and methodology for developmental and economic action throughout the KSA. Associated with the government's new strategy—to decrease the level of reliance on oil production and increase the level of emphasis on human capital—is a change (new emphasis) in the relationship between the government and the HE system. It appears that the emerging economic reality is driving the Saudi government to redefine the role of HE, including exerting pressure on institutions to become more productive and gear their programs toward fulfilling the needs of the labor market.

> Indeed, according to the Saudi government, *Vision 2030* seeks to ensure that higher education outcomes are in line with the requirements of the job market. In the year 2030, we aim to have at least five Saudi universities among the top 200 universities in international rankings. We shall help our students achieve results above international averages in global education indicators (Saudi Arabia Council of Economic and Development Affairs [SACEDA], 2016, p. 40; see also The National, 2016).

It is important to note that any efforts to pursue higher international rankings must relate to workforce development outcomes as well.

In a nutshell, the key education-related elements of *Vision 2030* are to move forward with the following: (a) provide education services for all student levels; (b) improve the recruitment, training, and development of teachers; (c) improve the learning environment to stimulate creativity and innovation; (d) improve the curricula and teaching methods; (e) improve students' values and core skills; (f) enhance the education system's ability to address national development requirements and meet labor market demands; (g) develop creative financing methods and improve the education system's financial efficiency; and (h) increase private-sector participation in the education sector (Kingdom of Saudi Arabia, 2016).

The importance of HE assessment extends beyond students' performance in examinations and assignments to include the overall performance of the institution itself and any major components relative to clearly defined benchmarks.

In order to ensure that it achieves the education-related objectives set out in the *Vision 2030* strategy, the Saudi government needs to closely monitor the performance of the various participating stakeholders. According to Kuh and Ewell (2010), human capital investment and national economic growth cannot happen without first assessing students' learning to ensure that graduates are acquiring the skills and competencies required for success in the 21st century. The importance of HE assessment extends beyond students' performance in examinations and assignments to include the overall performance of the institution itself and any major components relative to clearly defined benchmarks. Indeed, the assessment of students' learning and the institution's overall effectiveness is essential because it provides the foundation for better strategic planning and decision making, including with respect to the allocation of human and financial resources (Middaugh, 2010). As Middaugh (2010) emphasizes, "successful colleges and universities in the twenty-first century will be characterized by effective assessment and planning" (p. 13).

The logical next step after assessing institutional performance and incorporating assessment data into strategic planning is holding key stakeholders responsible or accountable for their acts and omissions and, in particular, for the degree to which they have met their respective objectives. According to Trow (1996),

> accountability is the obligation to report to others, to explain, to justify, to answer questions about how resources have been used, and to what effect…. The fundamental questions with respect to accountability are: who is to be held accountable, for what, to whom, through what means, and with what consequences (p. 2).

In the academic context, accountability for organizational performance offers multiple benefits to policymakers and the public as a whole, one of which is the creation of incentives for internal stakeholders to pursue excellence (including faculty members, senior leaders, and administrators). Thus, even though accountability is retrospective in the sense that it mostly focuses on the assessment of previous performance, it is also prospective in that it tends to structure future behavior in positive ways. The anticipation of being held accountable casts a forward-looking shadow over future action (Trow, 1996).

Similarly, Rabovsky (2012), in referring to accountability-generated incentives in public organizations, points out that, in many cases, leaders and administrators respond to performance-based incentives by implementing approaches to management that raise efficiency and improve performance. Moreover, incentive structures that make distinctions between organizations based on their level of performance can have additional beneficial effects on institutional outcomes. Indeed,

> by rewarding organizations that perform well and sanctioning those that perform poorly, policymakers can provide strong incentives for public agencies to reduce or eliminate wasteful activities and to employ entrepreneurial strategies in developing new technologies and methods to improve service delivery (Rabovsky, 2011, p. 2).

This in turn can lead to a stronger return on investment or to greater 'bang for the buck.' In the HE sector this would help to demonstrate that a national investment in HE is worthwhile, which in turn enhances faith in higher education.

Potential challenges of assessment and accountability. Notwithstanding the many benefits offered by assessment and accountability in the university environment, these approaches present a number of hazards and limitations. This is especially the case when assessment and accountability are unilaterally imposed with few if any distinctions made to account for significant variation in access to resources, the difficulty of tasks, or both (Rabovsky, 2011). Moreover, some Western institutions have observed that administrators sometimes respond to ambitious accountability requirements by both "gaming the system" to manipulate the data provided to evaluators and focusing heavily on activities that are likely to elevate scores in the short term despite the potential to undermine long-term performance (Abernathy, 2007).

In commenting on universities in the United Kingdom, Trow (1996) points out that many academic departments carefully manage their reporting to the relevant government authorities. This can be seen in

> the care with which they sort out the sheep from the goats on their staffs (with what effect on the morale of the goats?); the intense interest that has arisen around gaining certified publication before closing date—an interest that in some cases has involved the withdrawal of scholarly papers from one journal to place them with another solely on the ground of publication date; the recruitment of stars trailing clouds of publications and glory in their train. And on the teaching side, the anxious rehearsals for a forthcoming site visit, whole days given to walking through the visit, with every moment and conversation choreographed and planned for fullest effect; the even more anxious employment of consultants on how best to present themselves to those review committees. (p. 5)

Even when assessment and accountability measures are undertaken in good faith and with high diligence by all relevant stakeholders, there remains potential for bias to interfere with the assessment process and especially with any subsequent implementation of the resulting conclusions. This is largely a reflection of the heavy element of subjectivity in all forms of assessment. According to Rabovsky (2011), subjectivity leads to a lack of neutrality around the use of performance information, largely because (a) these data must be interpreted and given meaning by human decision makers and (b) there sometimes is disagreement within the policy community about the legitimacy of specific indicators. This objectivity and neutrality deficit can in turn lead key stakeholders to view performance information and the conclusions

Even when assessment and accountability measures are undertaken in good faith and with high diligence by all relevant stakeholders, there remains potential for bias to interfere with the assessment process.

generated from that information with distrust, ultimately with negative consequences for organizational learning. One way to offset this situation is to ensure that raters come from HE, business, industry, government, and other stakeholders.

A well-designed collegiate assessment program could and should include collection points after graduation. Just looking at learning within the undergraduate period is very limiting.

Another problematic scenario involving good-faith assessment and accountability is related to attempts to improve organizations based on the assessment conclusions. As Romzek (2000) explains,

> in seeking to change the operations of agencies and employees, it is essential to have corresponding shifts in accountability relationships so that behavioral expectations are appropriately aligned with managerial emphases. The new behaviors sought need to be reflected in the administrative measures used to evaluate agency and individual performance (p.35).

Higher education institutional effectiveness (i.e., its ability to achieve its mission, vision, and goals) will depend on these new accountability relationships.

Finally, although some may not agree, Trow (1996), a renowned scholar on this topic, suggested that even the best assessment systems are unable to capture the true depth and breadth of an educational institution's impact on its students. This is yet another reason why policymakers should be conservative in setting expectations for any particular assessment and accountability program. Trow asserted that "our impact on students can never be fully known" (1996, p. 13). He softens the implication that this statement may come across as weak with the caveat that an institution's impact

> emerges over [students'] whole lifetimes and takes various forms at different points in their lives. Those effects are mixed up with many other forces and factors over which we in higher education have no control—and among these are the student's character and life circumstances. Moreover, our influence on their lives takes many different forms, the most important of which are unmeasurable. One of the major functions of higher education which evades all measurement is our ability to raise the horizons of our students, to encourage them to set their ambitions higher than if they had not come under our influence. Colleges and universities at their best teach students that they can actually have new ideas, ideas of their own rather than merely the manipulation of ideas produced by others. That is not a conception of self very often gained in secondary school, and yet it lies at the heart of most of what people who gain a post-secondary education achieve in their lives. (p.13)

Although Trow (1996) provides salient provocative points, a well-designed collegiate assessment program could and should include collection points after graduation. Just looking at learning within the undergraduate period is very limiting. As well, future assessment initiatives could expand beyond knowledge attainment and measure nontraditional outcomes such as hope, curiosity, and entrepreneurship.

Internal versus external assessment. In order to understand the growing emphasis on HE assessment around the world, including Saudi Arabia, it is important to appreciate the distinction between internal assessment and external assessment. Internal assessment refers both to the gathering of data pertaining to the performance of an institution and any subsequent analysis of those data by the same institution for its own internal purposes. In the academic context, this kind of assessment is typically related to the performance of students and faculty members, although it can also pertain to institutional performance as a whole (e.g., meeting various financial targets). External assessment involves one or more external organizations analyzing institutional data and then drawing conclusions about the latter's performance of the same (in many cases relative to the performance of similar institutions within the same country or even on the international level; Trow, 1996).

Trow (1996) captures an important distinction between these two categories of assessment. External accountability is much like an audit and is aimed at ensuring that universities and colleges are fulfilling their responsibilities to their key stakeholders and, ultimately, to society as a whole. Internal accountability strives to ensure that various

institutional components (e.g., administration, faculty, students, and custodians) are fulfilling their respective responsibilities to one another with a view to meeting and even exceeding performance targets, identifying areas for improvement, and optimizing the use of resources, among other objectives.

Within Saudi Arabia, external assessment is conducted by various national-level bodies, one of which is the *National Center for Assessment and Academic Accreditation* (NCAAA, https://www.ncaaa.org.sa/en/). This organization grants both institution- and program-level accreditation and thus helps to ensure the quality of KSA post-secondary education. In addition to receiving national, domestic accreditation, some Saudi universities and colleges choose to pursue international-level accreditation, with a notable example being IAU's College of Engineering, which received accreditation from ABET, and College of Medicine, which received accreditation from Canada's Royal College of Physicians and Surgeons.

Institutional effectiveness. The distinction between internal and external assessment should also be understood within the broader context of institutional effectiveness (i.e., how well it is achieving its goals). Welsh and Metcalf (2003) explain that the improvement of institutional effectiveness requires the assessment of program learning outcomes, a review process for academic programs, and the assessment of the various performance outcomes set by the institution. Furthermore, the implementation of associated measures is aided by incorporating assessment activities into daily academic operations (Sheldon, Golub, Langevin, St. Ours, & Swartzlander, 2008).

Academic organizational effectiveness is closely related to organizational effectiveness in other contexts, including business settings. Variables that affect the implementation of HE institutional-effectiveness measures are similar to those that affect continuous improvement in corporate and industrial contexts including but not limited to adequacy of both training and time, degree of stakeholder buy in, and the extent to which data are used to inform decision making (Sheldon et al., 2008). Indeed, many opportunities exist for exploring the programs, strategies, and techniques that are designed to increase performance within corporate and industrial settings, and transferring and adapting those programs, strategies, and techniques to university and college environments. To go one step further, future HE external assessments could be expanded to include stakeholders in the assessment process like using business and industry representatives to evaluate students, faculty, and curricula.

To wrap up the overview of the Saudi context, there is no doubt that accreditation is a significant issue in Saudi Arabia, and its level of importance will only increase as the desire to improve (and expectations of improvement) become more intense among key stakeholders, including the institutions' respective leadership teams, faculty members, students, parents, and alumni as well as the KSA's government and major employers. Alshayea (2012) points to the growing emphasis on accreditation in the country by noting that "the Saudi Arabian higher education system has taken major actions to improve itself and maintain international standards. These actions are in response to the perceived low quality of the system which has affected its graduates" (p. 2).

Couched within this Saudi Arabian context, we now present a case study of IAU, a university leader in the KSA ranked 6th among Saudi higher education institutions in Quacquarelli Symonds ranking. We highlight how IAU paved the way for what may be eventual widespread use of a new approach to assessment that ultimately increases institutional performance and competitiveness, in house and beyond. Results pursuant to performance, processes, and products may also serve as a way to generalize assessment findings beyond higher education, a topic for future research.

Method

Case studies investigate contemporary phenomena within their real-life context (Yin, 1984), in this case institutional academic assessment and effectiveness in higher education as it played out at IAU. The university developed a new approach for assessing academic and institutional effectiveness (including administrative processes) in association with applying for and receiving external accreditation. This case especially focuses on how IAU dealt with faculty members and students' academic performance components (via the

We highlight how IAU paved the way for what may be eventual widespread use of a new approach to assessment that ultimately increases institutional performance and competitiveness, in house and beyond .

establishment of Key Performance Indicators [KPIs]) relative to IAU's primary mission, objectives, and strategic plan.

The authors judged IAU to be a good example of both a critical and unique case (Yin, 1984) of institutional academic assessment and effectiveness in the KSA. Critical because insights into its innovative approach to institutional assessment will constitute a significant contribution to knowledge and any theory building about higher education in KSA. It is unique in the sense that IAU is one of the first among a few KSA universities to engage in the proactive assessment of the effectiveness of its academic success as a higher education institution.

IAU is one of the first among a few KSA universities to engage in the proactive assessment of the effectiveness of its academic success as a higher education institution.

Per the tenets of case study research design protocol, a holistic single case study was developed and characterized as such because it examines the global nature of the phenomenon with no attention to subunits, which would have lead to an embedded case study (Yin, 1984). This merits clarification because IAU has three campuses and many specialized colleges that accommodate over 45,000 students (IAU, 2019), which would amount to subunits in an embedded case.

The chain of evidence used to develop the case included documents from multiple sources, among them the university and its related units and websites, internal and external reports, and personal communication with key informants. These steps collectively addressed issues of case study-related validity and reliability (Yin, 1984).

Finally, we strived to achieve analytical generalizability instead of statistical generalizability to a particular population. The intent was to facilitate generalizing the IAU case to (a) the broader phenomenon of institutional academic assessment and effectiveness in the context of KSA higher education (i.e., to similar situations) and (b) any eventual theories or models pursuant to that phenomenon (Yin, 2010).

Case Study of Imam Abdulrahman Bin Faisal University

A chronological approach is used to present the IAU's case study (Yin, 1984), starting in 2013 and culminating in 2017. The chronological case reporting approach serves to document changes over time thereby informing an evolving context for the phenomenon under study. As a caveat, equal attention was given to reporting each key event (five in total occurring during three specific years, 2013, 2015, and 2017). These time periods contain different events that unfolded culminating in the final assessment protocol (Yin, 1984). The 2017 termination date for the case was chosen because it represents when the university's assessment protocol was finalized. Future studies will explore and report on its effectiveness and efficacy. This case study served only to document the emergence of the IAU's assessment system and its attempts to establish protocols, procedures, and an assessment culture.

Creation of the IAU's Internal Decision Support Unit (DSU; 2013)

Imam Abdulrahman Bin Faisal University or IAU (formerly known as University of Dammam) is a NCAAA-accredited public university located in the Eastern Province of Saudi Arabia. It comprises four campuses (Dammam, Khobar, Qatif, and Jubil). The university houses 21 colleges, nearly 80 majors, and five research centers with upwards of 45,000 students and 3200 faculty members (local and international). The colleges are clustered under four overarching disciplines: health, engineering, science and management, and arts and education (IAU, 2019; University of Dammam, 2015). IAU ranks in the top six universities in Saudi Arabia and 582nd in the world out of 1000 (Quacquarelli Symonds [QS], 2019).

The IAU's Decision Support Unit (DSU) was established in 2013 on the initiative of the university's Vice President for Studies, Development, and Community Services (Alnouman, 2017). The Unit deals with the collection, processing, and analysis of data and the presentation of its findings to the university's decision makers in order to provide them with insights into both the institution's true state of affairs and various alternative courses of action. The IAU-DSU's mandate also includes responsibility for the development and implementation of decision-support systems related to the educational, research, and professional services provided by the university to internal and external parties.

The IAU-DSU intends to be a model for universities and other higher education institutions throughout Saudi Arabia. It handles an impressive range of university data including information related to education delivery, academic performance, human resources, financial performance, research, ranking, accreditation, student services, community services, hospital services, and all types of operational systems (personal communication, Alnouman, 2017; spring 2018). Furthermore, Alnouman (current head of the IAU-DSU) explained that the Unit's work takes into consideration the objectives, requirements, and interests of beneficiaries and stakeholders in order to: help anticipate, formulate, and forecast the future; develop models that facilitate the generation of alternatives; and develop proposals for appropriate development policies in all sectors of the university.

The Unit's work does not stop at this point, as it continues to both monitor the changes and transformations that occur as a result of development decisions. It presents the results of follow-up to officials and decision makers. The IAU-DSU's mandate encompasses multiple areas and themes (e.g., students,' faculty members,' and administrative members' performance; research outcomes; rankings; and alumni). These all point to institutional effectiveness. All of the Unit's work is undertaken in accordance with the authority conferred on it by the university, as set out in its *Internal Policies and Regulations Manual* (personal communication, Alnouman, spring 2018).

Creation of the Internal IAU-DSU Dashboard (2013)

The IAU appreciated that quantitative data were necessary for informing management decisions. To facilitate the managerial aspect of assessment work, the DSU now uses analytic dashboard software as a toolbox that processes massive quantities of data and makes them ready for analysis and interpretation. Dashboards connect to multiple data sources to ensure that the institution has a clear snapshot of its key metrics. These systems help to provide a solid basis for decision making related to improving faculty members' teaching, students' learning outcomes, and overall institutional performance. They are especially useful for provosts, deans, department chairs, and administrative directors in that they help these individuals to manage their institutions with greater effectiveness and efficiency (Middaugh, 2010). At IAU, deans and upper administration have access to the dashboard data and use this information to make strategic decisions. Select data are made available at the IAU's website for public consumption.

Data mining and subsequent analysis undertaken using the DSU dashboard provide information that supports decision making aimed at improving the overall performance of the university and its 19 colleges and faculties as well as its faculty members and students. Indeed, given the many challenges confronting Saudi HE institutions, it is clear that quantitative, evidence-based analysis is essential for increasing institutional effectiveness, thus ensuring further progress in teaching and learning. This kind of ongoing assessment should help to bring greater transparency and accountability to Saudi post-secondary institutions' operations while raising their domestic and international reputations.

Actually, IAU's assessment objectives are very similar to those articulated by an American accreditation organization, the Middle States Commission on Higher Education. Much like this organization, IAU believes that "assessment of student learning demonstrates that an institution's students have knowledge, skills, and competencies consistent with institutional goals, and that students at graduation have achieved appropriate higher education goals" (Middaugh, 2010, p.11).

Obtaining External NCAAA Institutional Accreditation (2015)

To facilitate the external assessment process and thus obtain institutional accreditation, IAU developed several key policies and support structures, as noted above. These policies and structures enabled the NCAAA to better gather the information that it required to grant IAU institutional accreditation in 2015. IAU's accreditation has enhanced its reputation and legitimacy, both nationally and regionally. It now ranks 5th among 28 public KSA universities and 12th among the Middle East and North Africa (MENA) region (detailed at IAU's website, https://www.iau.edu.sa/en).

Given the many challenges confronting Saudi HE institutions, it is clear that quantitative, evidence-based analysis is essential for increasing institutional effectiveness, thus ensuring further progress in teaching and learning.

Creation of the IAU's Directorate of Monitoring and Academic Performance (MAP; 2017)

In 2017, the university's president established the Directorate of Monitoring and Academic Performance (MAP) under the VP Office for Academic Affairs (IAU, 2018). The MAP's main responsibility is to monitor the university's then 19 constituent colleges for academic performance, areas for improvement, and the ways in which (and to what extent) their mission and vision are being accomplished. MAP's institutional assessment mission comprises the following four elements:

- responding to mandated institutional data-reporting requirements emanating from the IAU's higher administration for the purposes of improving the institution and responding to requests from regional and specialized accreditation organizations;

- developing and monitoring performance indicators that measure overall institutional effectiveness and the university's and its constituent colleges' progress toward the achievement of the university's strategic planning goals, with an emphasis on the quality of teaching and learning;

- conducting special research studies to assist academic and administrative departments in meeting program-review, assessment, and accreditation requirements; and

- contributing to ongoing institutional initiatives to improve methods of storing, managing, analyzing, and reporting data.

Creation of Internal Key Performance Indicators (KPIs; 2017)

MAP's work is facilitated through the use of KPIs which were developed in 2017 after receiving NCAAA accreditation, to monitor the colleges' academic performance.

MAP's work is facilitated through the use of KPIs (see Table 1), which were developed in 2017 after receiving NCAAA accreditation, to monitor the colleges' academic performance. These KPIs emerged as a result of a three-pronged initiative that included (a) identification of both best practices and (b) IAU's priorities followed by (c) a pilot test of the draft indicators. This initiative was led by one of the authors, working with IAU faculty members and another professor. All three stages are now discussed.

Phase 1: Identification of international, regional, and local KPIs and best practices. The goal during this first phase was to identify best practice; that is, any KPIs used to measure the academic performance of universities and their component colleges and faculties around the world. The team (led by one of the authors) began by benchmarking with overseas universities in Europe, the United States, Japan, and Australia, as well as regional universities in Egypt, and local universities in the KSA, using their respective websites. The aim was to understand how these institutions monitor their own academic performance and identify what KPIs they employed.

Relevant and common KPIs were listed and grouped into four categories: teaching, learning, assessment, and feedback. These KPIs were subsequently compared to those of the AFAQ (a long-term Educational Plan approved by Royal Decree; Kingdom of Saudi Arabia Ministry of Education, 2015) and the NCAAA. Then, various subcategories were developed (e.g., student and course performance and employability) and the KPIs were distributed under each subcategory. Based upon the KPIs, a Balanced Score Card (BSC) was created to assess both performance and progress.

Phase 2: Identification of IAU's priorities. The goal during this second phase was to determine whether the list of categories, subcategories, and KPIs developed in the first phase was appropriate to IAU's context. To determine this, the monitoring-related objectives within IAU's strategic plan were analyzed, including their respective KPIs. The outcomes of this analysis then provided a basis for enhancing the original list. The final KPIs (before pilot testing) were deemed to be closely aligned with IAU's context and needs. Strategic IAU objectives included assessing the impact of new learning and teaching methods that can be

linked with MAP. In particular, the draft of the strategic plan that the university shared with faculty members included the following objectives (IAU, 2017):

- 1.1 Expansion of teaching and learning opportunities;

- 1.2 Development of programs, curricula, teaching, and evaluation methods;

- 6.1.2 Activation of the role of the decision-making unit; and

- 6.1.5. Evaluation of the performance and motivation of the university employees.

Phase 3: KPI piloting. In the third phase, the list of KPIs developed in phases one and two was first implemented at IAU's College of Computer Science and Information Technology, the first college to be assessed for its level of academic performance. The results and feedback were subsequently used to change, amend, delete, and add various KPIs. The final list of KPIs, profiled in Table 1, has been approved for use in the initial assessment of the remaining colleges within the university. It is organized by student performance (seven dimensions), course performance (one dimension), and alumni employability (one dimension).

Table 1
Key Performance Indicators for Assessing Academic Effectiveness of IAU Colleges

A. Student Performance	A.1 Degree Completion	A.1.1	On-time degree completion ratio compared with IAU's ratio (over the last four years)
		A.1.2	On-time degree completion ratio of the various programs (over the last three years)
	A.2 Faculty-to-Student Ratio (FSR)	A.2.1	Faculty-to-Student Ratio (FSR) per college (compared with national/international ratios)
		A.2.2	Faculty-to-Student Ratio (FSR) per program
	A.3-Student Warning Analysis	A.3.1	Development of the number of students on probation over the last four regular semesters
		A.3.2	Number of students per college/program who received first, second, and third warnings during the last four semesters
	A.4 Program Drop Outs	A.4.1	Number of students who dropped out (dropped out from all courses) during the last four semesters
		A.4.2	Number of students who dropped out (dropped out from some courses) during the last four semesters
		A.4.3	Number of students who changed program during the last four semesters
		A.4.4	Number of students who dropped out from the university during the last four semesters
	A.5 Program/College Performance	A.5.1	Distribution of students' CGPA on the bell curve per college
		A.5.2	Distribution of students' CGPA on the bell curve per program
		A.5.3	Percentage of students whose CGPA is 3.75 or above
		A.5.4	Development of students' academic achievement (CGPA) per college [average] over the last two years
		A.5.5	Development of students' academic achievement (CGPA) per program [average] over the last four semesters
		A.5.6	Development of students' academic achievement (CGPA) by batch
	A.6 Student and Faculty Load	A.6.1	Development of students' unit load per college [average] over the last two years
		A.6.2	Development of students' unit load per program [average] over the last two years
		A.6.3	Faculty load per academic staff member by college/department (average: number of course sections / total number of faculty)
	A.7 Active Students	A.7.1	Percentage of international/non-Saudi students, if applicable (compared with the IAU percentage)
		A.7.2	Percentage of female to male students, if applicable (compared with the IAU percentage)
B. Course Performance	B.1 Course/Program Performance	B.1.1	Percentage of Fs per college/program – List of the five topmost percentages in courses per program
		B.1.2	Percentage of A^+s and As per college/program – List of the five topmost percentages in courses per program
C. Alumni	C.1 Employability	C.1.1	Employability rate
		C.1.2	Gender employability rate / program

MAP's work also includes ongoing verification of the recently established KPIs based on benchmarking with national institutions, namely the NCAAA and the AFAQ and with international institutions such as the National Institute for Learning Outcomes Assessment (NILOA). The KPI template used by MAP respects the following variables, which guide data mining and analysis for each of the 18 colleges:

- *undergraduate students' profile* (in terms of total headcounts, breakdown by gender, degree completion period/rate, cumulative GPA by program and by college, and students' withdrawal rate per program and per college relative to that of the university as a whole);

- *undergraduate students' success* and progress rates (in terms of retention and graduation rates for specific cohorts of students);

- *classroom environment* (in terms of the student/faculty ratio and class sizes, among other variables);

- *full- and part-time faculty distribution* (in terms of gender and other characteristics);

- *degrees awarded*; and

- *employment outcomes* of recent graduate cohorts (in terms of jobs, professions, and/or enrollment in graduate programs).

Any KPI data gathered by the IAU-DSU are converted into information that can be used in making important decisions about admissions, institutional plans, and the allocation of human and financial resources in support of activities related to teaching and learning. And, as Middaugh (2010, p. 230) recommends, "effective institutions continue to monitor the ongoing relevance of the college or university experience of graduates beyond the receipt of a degree."

Case Wrap Up and Discussion

Given the many changes that are taking place in the global economy, Saudi Arabia has no choice but to reduce its level of reliance on oil production. The KSA government's new strategy, as articulated in various official documents including *Vision 2030*, emphasizes a significant realignment of priorities in the country's higher education system. The accomplishment of these priorities depends to a great extent on the implementation of effective assessment (internal and external) within Saudi universities and colleges.

This case study illustrated that Imam Abdulrahman Bin Faisal University (IAU) is already distinguishing itself as a leader, as it has begun to move forward with the performance measurement and analysis of all key aspects of its operations. The creation of the internal dashboard and KPIs was linked with IAU's NCAAA external accreditation. The intent was to ensure that the university continued to grow and be informed by critical and current information about the institution's academic effectiveness. The KPIs were especially designed so the university could monitor and collect data about the individual colleges. Subsequent to receiving NCAAA accreditation and the implementation of these academic assessment structures and tools, the IAU is now a national and regional leader, as noted (QS, 2019).

A culture of informed decision making based on evidence collected through KPI assessments is a precondition for the continued infusion of resources into assessment.

The implementation of institutional assessment represents a milestone for IAU in terms of raising its performance and quality on many fronts. Over time, the implementation of assessment will contribute to improving the quality of teaching and learning, thus elevating not only the university's academic performance but also its national and international ranking. These improvements will have a significant impact on IAU's constituent 19 colleges by raising their level of accountability for their teaching performance and learning outcomes, an approach that is closely aligned with the objectives of the Commission on the Future of Higher Education (2006).

A culture of informed decision making based on evidence collected through KPI assessments is a precondition for the continued infusion of resources into assessment. The determination of each institution's strategic objectives with respect to students' lives and experiences is required in order to frame credible discussion about institutional effectiveness

and provide analytical results that can be used to build strategies. This cycle could be most effective in translating data into information that can be made available to both internal and external constituencies.

HE assessment and planning are necessities and should be inclusive not just of reporting on performance but also the preparation of enrollment planning models, administration and analysis of institution-wide surveys, and dissemination of information obtained from those surveys to the university community and wider audience. Furthermore, other universities can follow the lead of both the IAU-DSU and MAP, which are encouraged to oversee and maintain a portrait of each college as part of their participation in a rigorous accountability system. The presence of accountability measures in the university culture will ideally become a cornerstone of the institution. Among other benefits, this culture will help to ensure that the constituent colleges, faculty members and students will appreciate the importance of assessment and will make a sincere effort to ensure that it becomes and remains meaningful.

Although centralized decision-making by administrators helps to avoid some of the challenges associated with collective decision-making, there is no perfect method of discharging the monitoring role without ongoing revision to the KPIs. The key to the successful assessment of institutional effectiveness is to ensure that the information obtained from assessments is actually used to inform strategic planning and decision-making, including resource allocation. Moreover, communicating information and ideas quickly and comprehensively is essential for conveying the central message of assessment—that is, improving teaching and learning (Kuh et al., 2014). The gathering and use of evidence of students' learning in making decisions and strengthening institutional performance and public accountability are collectively known as "student learning outcomes assessment" (Ikenberry & Kuh, 2014, p. 2).

The ongoing institutional assessment initiative that is being undertaken at IAU, one of the Saudi's leading HE institutions, includes the creation of a culture of institutional assessment aimed at improving the quality of teaching and learning. This case has illustrated that, by implementing these kinds of practices, institutions can become more accountable and transparent in their operations and ongoing strategic planning. The case provides useful insights for academic mangers and decision makers who are striving to shape a better future for their institution in general and to improve students' experiences and outcomes in particular.

Looking toward the future, accreditation must not be the ultimate goal—although being nationally accredited certainly is a milestone for IAU. Such an approach would be analogous to a student aggressively studying for days leading up to a final exam and then forgetting virtually everything after the end of the semester. Rather, the ultimate goal beyond accreditation should be ongoing systemic improvement, including the institutionalization of performance-management and quality control standards and practices, together with the continuous identification of areas for further development. That is, the institution has to evergreen its assessment protocol, keeping it fresh, current, and relevant.

In conclusion, in *Vision 2030* the Saudi government committed to (a) closing the gap between HE outputs and the requirements of the job market by working with the private sector; (b) having at least five KSA universities among the top 200 in international rankings; and (c) helping university students to achieve academic results that are above international education indicators (Saudi Arabia Council of Economic and Development Affairs, 2016). To that end, growing institutions like IAU must engage in continuous planning to determine what to measure and how to frame meaningful discussions about institutional effectiveness. They also need to develop improvement strategies that can be effectively communicated to internal and external stakeholders. The IAU case herein reflects the great potential to raise the quality of Saudi Arabia's higher education outcomes by assembling, deploying, and evergreening evidence and analysis of institutions' teaching and learning performance and effectiveness in light of national goals and visions.

HE assessment and planning are necessities and should be inclusive not just of reporting on performance but also the preparation of enrollment planning models, administration and analysis of institution-wide surveys, and dissemination of information obtained from those surveys to the university community and wider audience.

References

Abernathy, S. F. (2007). *No Child Left Behind and the public schools: Why NCLB will fail to close the achievement gap- and what we can do about it*. Ann Arbor, MI: University of Michigan Press.

Alexander, K. (2000). The changing face of accountability. *Journal of Higher Education, 71*(4), 411–431.

Alnouman, A.M. (2017). *Decision Unit Development at Imam Abdulrahman bin Faisal University*. Unpublished manuscript.

Alnassar, S. A., & Dow, K. L. (2013). Delivering high quality teaching and learning for university students in Saudi Arabia. In L. Smith & A. Abouammoh (Eds.), *Higher education in Saudi Arabia: Achievements, challenges and opportunities* (pp. 49–61). New York, NY: Springer.

Alrasheedy, A. (2017). *The cost of unemployment in Saudi Arabia* [Economic Research Working Paper 17/5]. Riyadh, Saudi Arabia: Saudi Arabian Monetary Authority.

Alshayea, A. (2012, February 8–10). *Improvement of the quality assurance in Saudi higher education*. Paper presented at the Cyprus International Conference on Educational Research, North Cyprus.

Boyer Commission. (1998). *Reinventing undergraduate education: A blueprint for America's research universities*. Stony Brook, NY: State University of New York. Retrieved from https://files.eric.ed.gov/fulltext/ED424840.pdf

British Council. (2012). *The shape of things to come: Higher education global trends and emerging opportunities to 2020*. London, England: Author.

Commission on the Future of Higher Education. (2006). *Final report: A test of leadership: Charting the future of U.S. higher education*. Jessup, MD: US Department of Education. Retrieved from https://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf

Hamdan A.K. (2015a). *Teaching and learning in Saudi Arabia: Perspectives from higher education*. New York, NY: Sense Publisher.

Hamdan A.K. (2015b). Reforming higher education in Saudi Arabia: Reasons for optimism. In F. Badry & J. Willoughby (Eds.), Higher Education Revolutions in the Gulf: Globalization and Institutional Viability. London, England: Routledge.

Hamdan A.K. (2016). The making of world class universities: Saudi Arabia's higher education system In K. Gray, H. Bashir, & S. Keck (Eds.), *Western higher education in Asia and the Middle East: Politics, economics, and pedagogy* (pp. 155–175). London, England and New York, NY: Lexington.

Hamdan A.K. (2017). Saudi Arabia: Higher education reform from 2005 onwards and the implications for women In S. Kirdar (Ed.), *Education in the Arab world* (pp. 197–210). London, England and New York, NY: Bloomsbury.

ICEF Monitor. (2018, July 18). *Saudi Arabia's expanding higher education capacity*. Retrieved from ICEF website: https://monitor.icef.com/2018/07/saudi-arabias-expanding-higher-education-capacity/

Ikenberry, S. O., & Kuh, G. (2014). From compliance to ownership. Why and how colleges and universities assess student learning. In G. Kuh, S. O. Ikenberry, N. A. Jankowski, J. Kinzie, P. Hutchings, & T. R. Cain (Eds.), *Using evidence of student learning to improve higher education* (pp. 1–23). New York, NY: Jossey-Bass.

Imam Abdulrahman Bin Faisal University. (2017). *Strategic plan: 2016-2030*. Dammam, Saudi Arabia: Author. Retrieved from https://www.iau.edu.sa/sites/default/files/content box/iau_updated_strategic_plan_english_30april2017.pdf

Imam Abdulrahman Bin Faisal University. (2018). *Directorate of Monitoring and Academic Performance*. Retrieved from IAU website: https://www.iau.edu.sa/en/administration/offices-of-the-vice-presidents/office-of-the-vice-president-for-academic-affairs/directorates/directorate-of-monitoring-academic-performance

Imam Abdulrahman bin Faisal University. (2019). *About*. Retrieved from IAU website: https://www.iau.edu.sa/en/about-us

Kingdom of Saudi Arabia. (2016). *National transformation program 2020*. Riyadh, Saudi Arabia: Author. Retrieved from http://vision2030.gov.sa/sites/default/files/NTP_En.pdf

Kingdom of Saudi Arabia. (2018). *Governance model for achieving Saudi Arabia's Vision 2030*. Retrieved from KSA website: http://vision2030.gov.sa/en/node/259

Kingdom of Saudi Arabia General Authority for Statistics. (2018). *Home*. Retrieved from https://www.stats.gov.sa/en

Kingdom of Saudi Arabia Ministry of Education. (2015, April 15–18). *IECHE 2015*. Media documents for International Exhibition & Conference on Higher Education, Riyadh, Saudi Arabia. Retrieved from http://www.ieche.com.sa/documents/media-room-2015.pdf

Kuh, G., & Ewell, P. (2010). The state of learning outcomes assessment in the United States. *Higher Education Management and Policy, 22*(1): https://doi.org/10.1787/hemp-22-5ks5dlhqbfr1

Kuh, G., Ikenberry, S. O., Jankowski, N. A., Kinzie, J., Hutchings, P., & Cain, T. R. (2014). *Using evidence of student learning to improve higher education*. New York, NY: Jossey-Bass.

Lindsey, U. (2010, October 3) .Saudi Arabia's education reforms emphasize training for jobs [Web log post]. Retrieved from https://www.chronicle.com/article/Saudi-Arabias-Education/124771

Middaugh, M. F. (2010). *Planning and assessment in higher education: Demonstrating institutional effectiveness*. New York, NY: Jossey-Bass.

Organization for Economic Cooperation and Development. (2019). *Country note: Education at a glance: Saudi Arabia*. Paris, France: Author. Retrieved from https://www.oecd.org/education/education-at-a-glance/EAG2019_CN_SAU.pdf

Pavan, A. (2016). Higher education in Saudi Arabia: Rooted in heritage and values, aspiring to progress. *International Research in Higher Education, 1*(1), 91–100.

Quacquarelli Symonds. (2019). *QS Arab region university rankings 2019*. Retrieved from Top Universities website: https://www.topuniversities.com/university-rankings/arab-region-university-rankings/2019

Rabovsky, T. M. (2011, June 2–4). *Accountability in higher education: Exploring impacts on state budgets and institutional spending*. Presented paper at the Public Management Research Conference, Syracuse, NY.

Rabovsky, T. M. (2012). Accountability in higher education: Exploring impacts on state budgets and institutional spending patterns. *Journal of Public Administration Research and Theory, 22*(4), 675–700.

Romzek, B. S. (2000). Dynamics of public sector accountability in an era of reform. *International Review of Administrative Sciences, 66*(1), 21–44.

Saudi Arabia Council of Economic and Development Affairs. (2016). *Saudi Vision 2030*. Riyadh, Saudi Arabia: Author. Retrieved from www.vision2030.gov.sa/download/file/fid/417

Sheldon, M. R., Golub, A. J., Langevin, J. R., St. Ours, P. A., & Swartzlander, B. J. (2008). Improving institutional effectiveness: Description and application of an implementation model. *Planning for Higher Education, 36*(2), 17–26.

The National. (2016). *How Saudi Arabia is planning a new economic era*. Retrieved from The National website: https://www.thenational.ae/business/how-saudi-arabia-is-planning-a-new-economic-era-1.142169

Trow, M. (1996). Trust, markets and accountability in higher education: A comparative perspective [*Research & Occasional Papers Series: CSHE.1.96*]. Berkeley, CA: University of California, Berkeley.

University of Damman. (2015). *The University of Dammam: At a glance*. Copenhagen, Denmark: Issu Publishing. Retrieved from http://issuu.com/universityofdammam/docs/the_university_of_dammam-__at_a_gla?e=14885385/12285085

Welsh, J. F., & Metcalf, J. (2003). Faculty and administrative support for institutional activities: A bridge across the chasm. *Journal of Higher Education, 74*(4), 445–468.

Yin, R. (1984). *Case study research*. Thousand Oaks, CA: SAGE.

Yin, R. (2010). Analytic generalization. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research* (pp. 21–23). Thousand Oaks, CA: SAGE.

Abstract

Technology use is increasing in higher education, particularly for test administration. In this study, Capaldi's (1994) sequential theory, which postulates that the specific order of reinforcements and nonreinforcements influences persistence in the face of difficulty or failure, was applied to online multiple choice testing situations in regard to easy and difficult questions. Preliminary data appear to confirm that the order of easy and difficult questions on a test has an impact on student efforts in completing the test. These data may be especially important in the context of online learning, where the teacher is physically absent, as well as in situations where tests are administered through technology-based learning platforms. In all cases, test performance plays a role in student motivation.

AUTHORS

Ronald Mellado Miller, Ph.D.
Utah Valley University

Maureen Snow Andrade, Ph.D.
Utah Valley University

The Effects of Test Question Order on Task Persistence

*I*ncreasingly, teaching and learning processes in higher education institutions are technology infused. Technology is a valuable resource for instructors, both in and outside of the classroom, particularly as used within learning management systems. These systems provide learners with access to a range of learning materials and activities and allow faculty to track student participation and progress through various assessments such as assignments and tests (Falvo & Johnson, 2007).

Such systems are also used as the platform for online courses. Over six million students in U. S. higher education institutions are taking at least one online course. CEO survey respondents indicated that their institutions offer approximately 40% of all courses in online or blended delivery modalities, with 67% of these courses taught by full-time faculty (Magda, 2019). These CEOs represent public institutions, which are experiencing a constant increase in online course enrollments (Seaman, Allen, & Seaman, 2018).

Student success, regardless of course delivery modality, is a significant issue for both faculty and administrators. Concerns about lack of progress, typically measured by retention and graduation rates, are paramount. Approximately 61% of incoming students are retained to the second year at their starting institution, while about 73% persist to their second year at any institution (NSC Research Center, 2018). Six-year graduation rates for full-time students obtaining a bachelor's degree are 60% nationally (National Center for Education Statistics, n.d.)

CORRESPONDENCE

Email
maureen.andrade@uvu.edu

Assessment of learning impacts student success. As such, examining how tests are structured and administered and how students manage testing environments is critical. Technology is frequently used across course delivery modes to administer assessments. While instructors may review tests for standards of reliability and validity, a variable that may be overlooked in test construction is that of student persistence in completing assigned tasks. In other words, when faced with difficult test questions or problems, what factors influence a student to make repeated attempts to solve the problem at hand?

Given the increase in numbers of students taking online courses where they are physically separated from an instructor, as well as increased numbers of students taking online assessments for face-to-face courses through a learning management system, this information could be particularly important. The purpose of this study was to examine if the order of easy and difficult questions on a test had an impact on student persistence in completing the test.

Literature Review

Faculty members across higher education institutions are responsible for assessing students' knowledge and skills. A common mechanism for this is formal testing. Faculty typically do not receive training in test construction as part of their PhD work, which is discipline-based, and may therefore rely on publisher-provided tests or create their own with varying degrees of success. Thus, increasing awareness of issues that impact effective test construction is advantageous in terms of obtaining accurate measurements of students' learning and constructing tests in ways that support student persistence in test completion.

A key issue related to testing is the impact of various question order strategies on performance (Bard & Weinstein, 2017; Caudill & Gropper, 1991; Pettijohn & Sacco, 2007; Tal, Akers, & Hodge, 2008). Random or sequential question order appears to have little effect (Cordero, Layson, Martinez, & Quindoza, n.d.; Tal et al., 2008) nor does reverse question order (Pettijohn & Sacco, 2007). Practice prior to an exam accompanied by predictions and postdictions on performance similarly has little impact; however, high achieving students have been found to be more accurate in predicting their performance than low achieving students although the former were underconfident while the latter were overconfident (Bol, Hacker, O'Shea, & Allen., 2005). The order of multiple-choice response items also appears to have little effect on the difficulty of items, although when the correct answer occurs last, the effect on difficulty is slightly increased (Hohensinn & Baghaei, 2017).

Studies have also examined variations such as easy-difficult, randomized, and difficult-easy scenarios (Bard & Weinstein, 2017; Weinstein, & Roediger, 2010, 2012), often in conjunction with student predictions of performance or self-evaluations (Bard & Weinstein, 2017; Bol et al., 2005; Hacker, Bol, & Bahbahani, 2008; Weinstein, & Roediger, 2010, 2012). These studies may involve having students pause between sets of questions to make judgments about their performance. When the easiest questions are presented first, students have more positive perceptions of their performance than when the opposite occurs, These perceptions are maintained throughout the test, perhaps due to students anchoring their positive self-evaluations based on the initial test questions (Weinstein & Roediger, 2010, 2012). This phenomenon also occurs with multiple test attempts—students anchor their self-evaluations based on question difficulty on the initial test attempt and do not adjust them to account for changes in question order on additional test attempts (Bard & Weinstein, 2017).

Capaldi's (1994) sequential theory also provides insights into research on question order and is particularly applicable to online multiple choice testing situations. The theory postulates that the specific order of reinforcements and nonreinforcements influences persistence in the face of difficulty, whether impossible or simply difficult questions. To apply this to testing, a reinforced trial would be one where the question is easy and a nonreinforced question would be a difficult question. For example, when faced with a series of difficult questions followed by an easy question, DDEE (where D indicates a difficult question trial and E an easy one), a student would learn the relationship SD→E, where SD represents the stimulus memory of a difficult problem. Thus, the student would learn that the memory of difficult problems (SD) is followed by easy problems (E).

In the face of many difficult problems, sequential theory seems to suggest that the student will persist in the task in anticipation of the easier problems to come (Capaldi, 1994). If trained in the EEDD sequence, however, the student would learn that SE→D, or that the stimulus memory of easy problems (SE) is followed by difficult problems (D). In this case, when faced with many difficult problems, sequential theory seems to suggest that the student will not persist in the task, as only difficult problems from then on are anticipated (Capaldi, 1994). As such, the student is less motivated to continue. Interestingly, this appears to the most common way for tests and texts to arrange their problem sets, with the easiest problems first and the most difficult problems being presented later.

In addition, sequential theory suggests that the amount of training as the result of test taking (or learning based on the patterns encountered) can have a dramatic effect on the persistence effects discussed earlier. For example, given a series, DEDE, a short test versus a long test should give different results. In the series DEDE, SD→E, where difficult problems lead to easy ones, this pattern would be learned. However, on a short test, the pattern would not be expected to be learned to the extent expected on a long test. Thus, it would be anticipated that a student trained with DEDE would be more persistent when faced with many difficult problems after a longer test than a shorter test.

While other experiments (e.g., Skinner, 1999; Cizek, 1994) have examined the question of order effects on student performance, no study has examined them utilizing the specific predictions given by sequential theory. For example, Skinner (1999) found that students who did well on difficult questions given first, did better on easy questions later on the test than students who had easy questions first. The test takers did not deal with question order on an item by item basis, but were affected by patterns of questions. Cizek (1994) described the results of his experiment dealing with order effects as unpredictable. Others, like Perlini, Lind, and Zumbo, (1998) found that arrangements of test question difficulty had little effect on overall performance. Thus, sequential theory, with its unique item-by-item predictions, appears to offer new insights into the effects of easy and difficult test question arrangements.

Method

The experiment was designed to examine sequential theory in the context of academic technology-based testing. Participants in this study were 38 undergraduate psychology students at a university in the Western United States. Participants were given extra credit for their participation.

Four multiple choice tests were created, an Easy Test (EEEEE), an Alternation Test (EDEDE), an ED Test (EEEDD), and a DDE Test (EDDEE). The tests consisted of either ten (short test) or fifty (long test) questions. An additional Impossible Test of thirty questions was created which consisted of impossible questions for which correct answers were removed and substituted by incorrect answers. Each subject performed one of the regular tests and was then given the Impossible Test. As with most tests, it was anticipated that the students would complete the entire test. Thus, the dependent variable was the amount of time spent trying to solve the Impossible Test questions.

In terms of predictions, it should be noted that each of the tests, the Alternation Test (EDEDE), the ED Test (EEEDD), and the DDE Test (EDDEE), except for the Easy Test (EEEEE), was equated for the percent of easy and difficult questions, with each condition consisting of 60% easy and 40% difficult questions and each condition beginning with an easy question. Only the order of the questions was manipulated. Once a sequence was given (for example, EEEDD), it was then repeated until either 10 trials were given (in the short training) or 50 trials were given (long training). A difference was hypothesized for each of the compared tests regarding time spent solving the Impossible Test questions.

For the Easy Test, it was hypothesized that the long test students would spend less time on the impossible questions as they were only trained S^E→E, which represents that the memory (S^E) of an easy question predicts that another easy question is coming. Therefore, when faced with difficult questions, the student would spend less time trying to solve them, especially in the long test condition where students would be more familiar with the S^E→E pattern. It was thought this would occur because, while trained S^E→E , the memory of difficult

questions, S^D, would not have been conditioned to any stimulus, and so would result in less response time.

For the Alternation Test, it was hypothesized that the training of $S^D{\rightarrow}E$ would lead the long test students to spend more time on the impossible questions than the short test students. This is because the stimulus $S^E{\rightarrow}E$, or the condition that difficult questions would be followed by easy questions, would have been more strongly conditioned through greater training.

For the ED Test, it was hypothesized that the training of $S^E{\rightarrow}E$ would lead to the long test students spending less time on the impossible questions than the short test students as the stimulus $S^E{\rightarrow}D$, that easy questions would be followed by difficult questions, and that difficult questions are only followed by difficult questions.

For the DDE Test, it was hypothesized that, since $S^{2D}{\rightarrow}E$ would become the most persistent in the impossible tasks as they were trained that two difficult questions would be followed by easy questions.

Materials

To form the tests used, a number of questions were extracted from various databases. The databases included the Myers (2003) and the Nairne (2005) introduction to psychology test databanks. The questions were labeled as either easy or difficult as judged by a panel of psychology students. The impossible questions were formed by removing the correct answer from a question and substituting it with an incorrect answer.

Examples of easy and difficult questions follow.

Easy Question:
Deafness refers to the inability to:
    a. walk

    b. see

    c. hear

    d. talk

Difficult Question:
Concept of the id, ego, and superego is best regarded as a theory about 3:
    a. Separate personalities inhibiting one body

    b. Different sets of reaction patterns within each personality

    c. Separate stages in personality development

    d. Distinct types of unconscious conflicts

Procedure

The students were given the web address of the test and were required to finish it in one sitting. The test consisted of the schedule already outlined, given in either two (short test) or ten (long test) sets of five questions. After completing that portion of the experiment, all students were given six sets of five (for a total of thirty) impossible questions. As per Teevan, Zarrillo, & Greenfeld (1983), the dependent variable measured was time, in this case, the amount of time taken to answer each of the impossible questions.

Results

A 2 (Training: Short versus Long) X 4 (Schedule: Easy, Alternation, ED, DDE) X 6 (Sets) X 5 (Questions) repeated measures mixed model analysis of variance (ANOVA) was run with Training and Schedule utilized as between subjects measures and Sets and Questions as the repeated measures. The dependent variable was the time in seconds to answer each of the impossible questions. Descriptive statistics for the study can be found in Table 1. Inferential statistics from the ANOVA can be seen in Table 2. A summary of the results can be seen in Figure 1.

The order of easy and difficult questions on a test appears to have an impact on a student's efforts to complete the test, as suggested by sequential theory.

Table 1
*Estimated Marginal Means-Training *Schedule*

Schedule	Training	Mean	SE	95% Confidence Interval	
				Lower	Upper
Easy	Long	23.6	7.55	8.16	39.0
	Short	27.1	9.80	7.13	47.2
Alternation	Long	38.8	7.17	24.21	53.5
	Short	26.2	11.63	2.46	49.9
ED	Long	25.2	7.55	9.74	40.6
	Short	38.7	8.75	20.81	56.5
DDE	Long	25.3	7.55	9.85	40.7
	Short	41.9	8.75	24.01	59.7

Table 2
Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Sets	24961	5	4992	2.512	0.032
Sets ✳Training	7664	5	1533	0.771	0.572
Sets ✳Schedule	62640	15	4176	2.101	0.013
Sets ✳Training ✳ Schedule	26358	15	1757	0.884	0.583
Residual	298150	150	1988		
Questions	10784	4	2696	1.562	0.189
Questions ✳Training	3771	4	943	0.546	0.702
Questions ✳Schedule	38878	12	3240	1.877	0.044
Questions ✳Training ✳ Schedule	38462	12	3205	1.857	0.047
Residual	207101	120	1726		
Sets ✳Questions	86747	20	4337	2.738	< .001
Sets ✳Questions ✳ Training	59130	20	2957	1.866	0.013
Sets ✳Questions ✳ Schedule	147325	60	2455	1.550	0.007
Sets ✳Questions ✳ Training ✳Schedule	148949	60	2482	1.567	0.005
Residual	950601	600	1584		

Note. Type 3 Sums of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Training	6730	1	6730	0.717	0.404
Schedule	10200	3	3400	0.362	0.781
Training ✳Schedule	28432	3	9477	1.010	0.402
Residual	281447	30	9382		

Note. Type 3 Sums of Squares

Figure 1
Mean Seconds Spent on the Impossible Questions
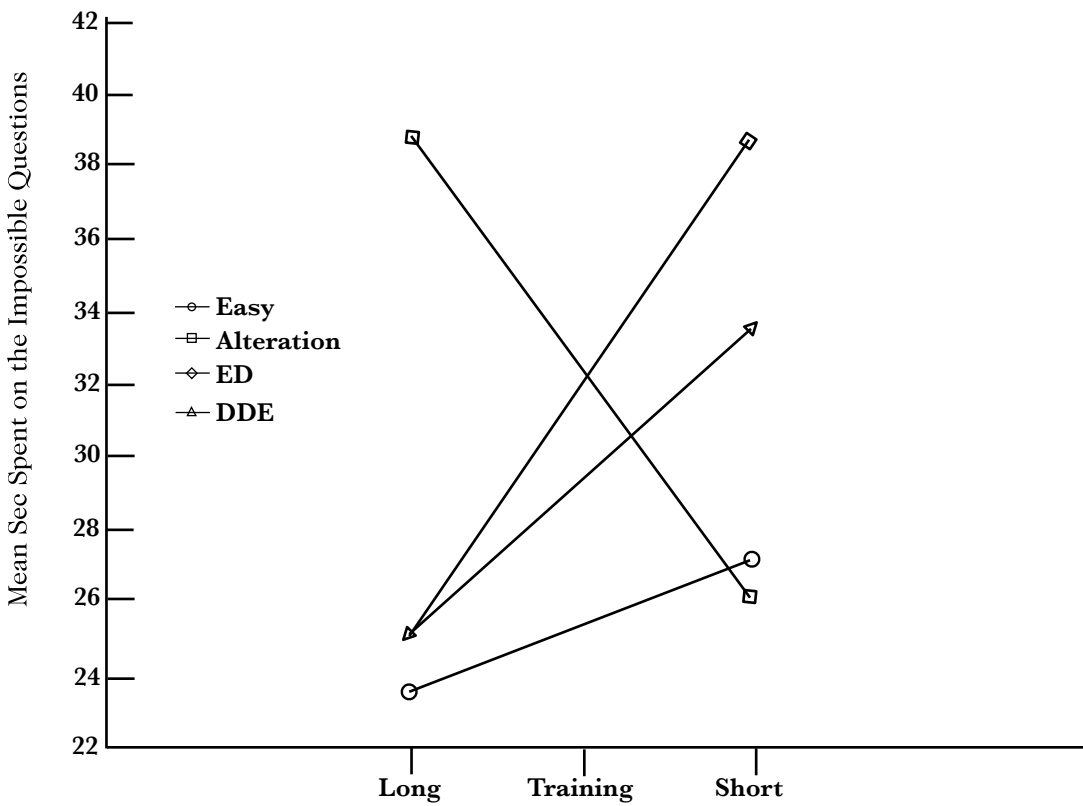
Overall, for the Easy Test, the mean seconds decreased slightly from the short to the long test while the Alternation Test increased from the short to the long test. Additionally, the ED Test and the DDE Test both decreased in their mean seconds spent on the impossible questions from the short to the long tests.

Discussion and Conclusion

Overall, all hypotheses appear to have been supported. The Easy Test and the ED Test did decrease in the time spent on the impossible questions, and the Alternation Test did increase time when comparing performance on the short versus the long tests. However, the DDE Test decreased in a way very similar to that of the ED Test.

Thus, the order of easy and difficult questions on a test appears to have an impact on a student's efforts to complete the test, as suggested by sequential theory (Capaldi, 1994). It is important to emphasize that each of the tests was equated for percentage of easy and difficult questions. Additionally, the Alternation Test became the most persistent of all of the tests in the long test condition, presumably resulting from the established pattern of being conditioned to anticipate difficult questions being followed by easy questions. While the DDE test did not perform in the long test as anticipated, this may simply be the result of the long test not being long enough to train the necessary relationship. While further investigation seems to be required, these data seem to support sequential theory and may be especially important in the context of distance learning, where the presence of the teacher is physically absent and the context of given tests may play a greater role in student persistence and motivation.

Additionally, as most textbook-based tests give the easiest questions first and build up to the most difficult, this essentially creates a situation like the Easy Test with easy questions followed by what (to the student) may seem to be impossible questions. It would be better for texts and exams to mix the easy and difficult questions such that difficult questions are followed by easy ones and not vice versa.

In sum, when constructing tests, faculty members should avoid having easy questions followed by difficult questions and should ensure that difficult questions are followed by easy ones.

While we acknowledge that the sample size for this study is small, another implication is the standard way of giving exams, where either easy and difficult questions are randomly given or where easy ones are generally followed by difficult ones, should be re-examined. Given that the order a student encounters of easy and difficult questions, some students may be given an advantage or be disadvantaged by current practice.

In sum, when constructing tests, faculty members should avoid having easy questions followed by difficult questions and should ensure that difficult questions are followed by easy ones. Pedagogically, faculty members often simply increase the difficulty of problems on a test; however, easy or easier questions should be interspersed with the difficult questions.

In terms of limitations of the study, the number of subjects was small, and a follow-up study with more participants and perhaps other conditions would add clarity. Additionally, this test was not a regular course exam; therefore, the motivations of the students for doing well may have been other than they would be on an exam that counts toward a grade. While this does not seem to have had an effect on the theorized outcomes, it may have played a role in the patterns students focused on and may have enhanced the effects seen here. In spite of these limitations, the study provides practical applications for test construction.

References

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education, 73*(4), 269–290.

Capaldi, E. J. (1994). The sequential view: From rapidly fading stimulus traces to the organization of memory and the abstract concept of number. *Psychonomic Bulletin & Review, 1*(2), 156-181.

Caudill, S. B., & Gropper, D. M. (1991). Test structure, human capital, and student performance on economics exams. *The Journal of Economic Education, 22*(4), 303–306.

Cizek, G. J. (1994). The effect of altering the position of options in a multiple-choice examination. *Educational and Psychological Measurement, 54*, 8-20.

Cordero, H. M., Layson, J. K., Martinez, B. M., & Quindoza, M. P. (n.d.). The effect of question order and paper color on reading comprehension of undergraduate students of UP Diliman: An empirical investigation.

Falvo, D. A., & Johnson, B. F. (2007). The use of learning management systems in the United States. *TechTrends, 51*(2), 40-45.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning, 3*(2), 101–121.

Hohensinn, C. & Baghaei, P. (2017). Does the position of response options in multiple-choice tests matter? *Psicológica, 38*(1), 93-109.

Magda, A. J. (2019). *Online learning at public universities: Recruiting, orienting, and supporting online faculty*. Louisville, KY: The Learning House, Inc.

National Center for Education Statistics. (n.d.). Fast facts. Graduation rates. Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=40

NSC Research Center. (2018, June 27). Persistence & retention – 2018. Retrieved from https://nscresearchcenter.org/snapshotreport33-first-year-persistence-and-retention/.

Perlini, A. H., Lind, D. L., & Zumbo, B. D. (1998). Context effects on examinations: The effects of time, item order and item difficulty. *Canadian Psychology, 39*(4), 299-307.

Pettijohn, T. F., & Sacco, M. F. (2007). Multiple-choice exam question order influences on student performance, completion time, and perceptions. *Journal of Instructional Psychology, 34*(3), 142–149.

Seaman, J. E., Allen, I. E., & Seaman, J. (2018). Grade increase: Tracking distance education in the United States. Oakland, CA: Babson Survey Research Group. http://www.onlinelearningsurvey.com/highered.html

Skinner, N. F. (1999). When the going gets tough the tough get going: Effects of item difficulty on multiple-choice test performance. *North American Journal of Psychology, 1*(1), 79-82.

Tal, I. R., Akers, K. G., & Hodge, G. K. (2008). Effect of paper color and question order on exam performance. *Teaching of Psychology, 35*(1), 26–28.

Teevan, R. C., Zarrillo, D., & Greenfeld, N. (1983). Relationship of fear of failure to amount of time spent in a final examination. *Psychological Reports, 53*(2), 343-346.

Weinstein, Y., & Roediger III, H. L. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition, 38*(3), 366–376.

Weinstein, Y., & Roediger III, H. L. (2012). The effect of question order on evaluations of test performance: How does the bias evolve? *Memory & Cognition, 40*(5), 727–735.

Abstract

Oral communication is an important learning outcome in higher education that can be difficult to assess. This article presents how one institution paired an online survey of instructors with a more traditional direct assessment to conduct a comprehensive assessment of oral communication in a General Education program, illuminating both how students are performing and how instructors are teaching oral communication. Detailed methodology and a full survey instrument are provided so that readers can translate this model to assessing oral communication or other major competency areas at their own institutions. Highlights of the survey results are also presented. The authors ultimately explain how the survey has led to faculty development around oral communication and to new oral communication resources that are aligned with the specific obstacles identified by survey respondents. They also describe how survey results helped to further explain the findings of the direct assessment.

AUTHORS

Molly Sapia, M.A.
Temple University

Gina Heller, B.A.
Temple University

Dana Dawson, M.A.
Temple University

Nicholas Carmack, B.A.
Temple University

CORRESPONDENCE

Email
molly.sapia@temple.edu

Using an Online Instructor Survey as Part of a Comprehensive Assessment: An Example in Oral Communication

Oral communication is a ubiquitous college-level learning outcome among core curricula and General Education (GenEd) programs, as well as within specific majors or degree programs. Like any major learning outcome, assessing oral communication should be included in regular assessment efforts. Unlike other major learning outcomes, however, scholars have noted that assessing communication skills is atypical, and study design can be difficult. Morreale et al. (2011), who presented a meta-analysis of oral communication assessment from 1975 to 2009, explained that "Methods of assessment used in other academic areas cannot always be adapted to communication, particularly to the assessment of oral communication skills" (p. 259). They also noted that since the mid 2000s, published articles on the topic have declined, making it difficult to find recent literature to guide study design for assessing oral communication at the college level (Morreale et al., 2011, p. 265).

One common approach to assessing major competency areas follows the framework of "direct assessments," or when "students…display their knowledge and skills" and "are evaluated using an assessment instrument. Objective tests, essays, presentations, and classroom assignments all meet this criterion" (Morreale et al., 2011, p. 257). To do this, one may collect anonymized student work items submitted as part of class work or course review efforts, and have faculty apply validated rubrics to assess them. Direct assessments allow for a program to directly test student performance and are often required by accreditation agencies. For example, Middle States accreditation requires "organized and systematic

assessments… evaluating the extent of student achievement of institutional and degree/program goals" (Middle States Commission on Higher Education, 2020).

How to directly assess student performance of oral communication is not immediately obvious, however. What "items" of student work are there to be collected? Dunbar et al. (2006) and Avanzino (2010) videotaped formal oral presentations by students in required GenEd public speaking courses as an adaption of a typical direct assessment design. While fruitful, this method seems particularly limited when studying oral communication for various reasons. For one, communication skills exist in a given moment and are context-dependent; "the appropriateness and effectiveness of communication education is generally based on the situation and in the perceptions of the viewer or the impression made by the communicator on the observer" (Morreale et al. 2011, p. 259). In addition, many universities do not require specific courses in oral communication or public speaking, and therefore it can be unclear from which courses one should gather student presentations for a direct assessment.

Formal oral presentations also do not represent the full spectrum of oral communication skills. Studies have found students gain oral communication skills via other teaching methods, such as group work (Crebert, Bates, Bell, Patrick, & Cragnolini, 2004) and in-class or extracurricular debates (Kennedy, 2007; Williams, McGee, & Worth, 2001). Additionally, employers often report informal oral communication skills as more commonly used or desired in the workplace than formal presentation skills. The National Association of Colleges and Employers (NACE) 2016 Job Outlook survey of 200 employers found the "ability to verbally communicate with persons inside and outside the organization" was ranked as the most important skill in job candidates (NACE Staff, 2016). Crosling and Ward (2002) surveyed employers of business graduates in Australia, who reported that the most common forms of oral communication in the workplace were informal work-related discussions among peers (with 83% of employers rating it as *constantly* or *often*) and responding orally to a supervisor's instructions (66%). On the other hand, only 36% of employers reported oral presentations as a constantly or often-used mode of communication in the workplace.

> **The National Association of Colleges and Employers (NACE) 2016 Job Outlook survey of 200 employers found the "ability to verbally communicate…" was ranked as the most important skill in job candidates.**

The GenEd program at Temple University set out to conduct an assessment of oral communication in the 2018-2019 academic year. Following the literature and guidance described above, they conducted a direct assessment of 100 formal oral presentations given by students in GenEd courses, videotaped student presentations (with consent), and had faculty score them using the AAC&U (Association of American Colleges & Universities) VALUE rubric for oral communication (see Maki, 2015). The project is not described in detail here, but examples of this type of project can be seen in Dunbar et. al (2006) or Avanzino (2010). This direct assessment illuminated how students are performing on the skill of formal oral presentations and engaged faculty in the topic by having them apply the rubric to those student presentations.

However, looking at the work of 100 students cannot deliver a comprehensive picture of oral communication within the GenEd curriculum at Temple, which includes a course library of around 140 courses and serves approximately 15,000 students in a given fall semester. In order to gain a better understanding of what types of oral communication were being taught and assessed throughout the general education curriculum, GenEd staff conducted an online survey of instructors. The survey was designed to explore such research questions as:

1. What forms of oral communication, if any, do GenEd instructors incorporate into their classes?

2. What forms of oral communication are seen the most frequently? The least frequently?

3. When instructors evaluate oral communication, how do they arrive at the grades?

4. What forms of feedback do instructors give on student oral communication?

5. What skill levels do GenEd instructors generally perceive their students to have reached on a variety of elements of oral communication?

6. What resources do instructors need and want regarding oral communication?

Uncovering the answers to these questions proved immensely fruitful; the authors were able to learn much more about oral communication in the Temple GenEd program through the survey than they would have from the direct assessment alone. Indirect assessments are often thought of as those which ask students to "reflect on learning rather than demonstrate it" (Morreale et al., 2011, p. 257); however this article aims to show that, alternatively, asking instructors to reflect on how they teach a skill can be equally as productive. The survey's findings helped further explain the quantitative findings from the direct assessment, and also helped illuminate issues with the alignment between how the program defines, teaches, and assesses oral communication. This has led to faculty development and resource dissemination, explained further below. In addition, the act of surveying faculty seemed to have a positive impact on the respondents, some of whom wrote that they were inspired to revamp their teaching of oral communication in their GenEd courses after going through the reflective process of answering the survey questions.

In this article, the authors present a detailed description of their online instructor survey as part of a comprehensive assessment of oral communication. The aim is to both present an updated and unique concrete example for assessing oral communication in GenEd or a similar program, and also to serve as a guide for programs looking to add an online instructor survey into a comprehensive assessment of a key competency area. Methodological details and tips for survey design are included, as well as highlights from the results. The authors then discuss how survey results were used to design and implement faculty development efforts related to the promotion of oral communication competency in GenEd classes.

Designing the Survey Instrument

In designing an instructor survey for assessment, one must operationalize the skill area(s) at the heart of the survey (here, oral communication). Decisions also need to be made about the parameters used to guide respondents. In other words, what time period should instructors be asked to reflect upon? Should instructors be asked to answer the survey about all courses they have taught, or a specific one? This therefore also involves selecting the unit of analysis, or the entity that is being analyzed for the study. In other words, will the survey be designed to gather information on instructors' general practices, or on what they did in specific courses?

For concepts to be translated into actual survey questions, they need to be operationalized, or defined, in concrete terms that can be translated to empirical measurement (Saris & Galhofer, 2007). Survey design experts Bradburn et al. (2004) explained that "one simple reason for making each question as specific as possible is to make the task easier for the respondent, which, in turn will result in more accurate reports of behavior" (p. 61). If you ask respondents very general questions—take, for example, what types of oral communication do you include in your courses?—this requires the respondent to define oral communication as a term before they can answer. This means that survey respondents may be answering with different definitions of oral communication in mind, which can generate messy data that are difficult to analyze.

Therefore, regardless of the competency area being assessed, one must create a concrete definition for that term and translate it to various, specific survey questions. The theoretical framework developed and applied to oral communication in the present survey includes three parts (presented in Table 1): 1) informal or everyday uses of the skill, such as group work or break-out discussions; 2) formal, graded uses of the skill where students are assigned specifically to demonstrate it, such as oral presentations or debates; and 3) explicit instruction of the skill. For oral communication, this third part may mean instructing students on how to design an oral presentation, rather than assigning the presentation itself. This framework can be applied in operationalizing other competency areas.

One must also define the timeframe about which respondents should answer. The current study uses the previous two academic years. While a shorter time period, such as the previous semester or one academic year, may have yielded more accuracy, expanding the timeframe to two academic years included more eligible instructors and protected their anonymity. For example, if only the previous year was selected, then only instructors who

> In addition, the act of surveying faculty seemed to have a positive impact on the respondents, some of whom wrote that they were inspired to revamp their teaching of oral communication in their GenEd courses after going through the reflective process of answering the survey questions.

Table 1
Operationalized Definition of Oral Communication for Survey Design

Table 1: Operationalized Definition of Oral Communication for Survey Design	
Type	**Definition**
Encouragement	Oral communication that is not required or graded, such as ungraded presentations or breaking the class out into small group discussions
Assessment	Any situation where students are required to communicate orally and are graded on it
Instruction	Any form of specifically teaching students how to communicate effectively, such as giving a demonstrative presentation

have taught a GenEd in the previous year would be eligible to take the survey, making the potential number of participants smaller. Additionally, when working within a small institution, choosing a shorter timeframe may make it easier to identify a respondent through their answers.

The authors also declared the unit of analysis to be courses, not instructors, and would recommend this to others looking to measure the prevalence of specific forms of oral communication or another skill. In other words, the survey asked respondents to answer about how they incorporated oral communication into a specific course, rather than asking about their general practices. These specificities aid participant memory recall, which can both increase the accuracy of the data and lower the burden on respondents by giving them a specific anchor from which to draw memories (Bradburn et al., 2004). Because the survey was anonymous, the participants were not asked to specify the course, but rather only the area of the GenEd curriculum in which the course falls. Therefore, participants were asked to "think of **one specific in-person GenEd course** that [they were] currently teaching or have taught in the past two years. Answer the questions with that course in mind." The full questionnaire can be seen in the appendix, and an outline of the sections of the questionnaire is presented in Table 2.

The survey asked respondents [about] a specific course... which can both increase the accuracy of the data and lower the burden on respondents by giving them a specific anchor from which to draw memories.

Table 2
Content of the Oral Communication Instructor Survey

Table 2: Content of the Oral Communication Instructor Survey	
Section	**Topic**
1	Consent form
2	Identifying a particular course about which to answer (but not reporting it)
3	Measuring encouragement, instruction, and assessment of oral communication
4	Class participation grades
5	Instructor perception of oral communication skills in that class
6	The option to answer the above sections 2-5 again for another course
7	Oral communication in online GenEd courses
8	Characteristics about the respondent

The survey proceeded by measuring encouragement, instruction, and assessment of oral communication within the specific course. Question 3 measured the presence of various forms of encouragement of oral communication. Question 4 defined instruction of oral communication and asked respondents if their course included this. If yes, the respondent was asked more about those forms of instruction in Questions 5 and 6. Similarly, Question 7 measured the presence or absence of assessment of oral communication, and in the case of presence, respondents were asked for more details on those assessments in Questions 8-12.

This is referred to as "skip logic," or using filter questions to determine if subsequent questions are relevant to the respondent, which should be employed to ask respondents as few questions as possible.

Definitions of encouragement, assessment, and instruction within this context were and should be provided to the respondent, and a list of closed-ended response options was and should be as exhaustive as possible. If, in drafting the question, one does not know what the response options should be, more background research is necessary. Discussion with faculty can be helpful in this endeavor, as well as reviewing the pre-test data, discussed in the next section.

Pre-test

Definitions of encouragement, assessment, and instruction within this context were and should be provided to the respondent, and a list of closed-ended response options was and should be as exhaustive as possible. If, in drafting the question, one does not know what the response options should be, more background research is necessary.

Before arriving at the final version of the questionnaire, the authors conducted a pre-test, or "initial testing of one or more aspects of the study design: the questionnaire, the sample design, a computer program for analysis, and so forth" (Babbie, 1973, p. 205). Once a draft of the survey was programmed into SurveyMonkey (https://www.surveymonkey.com), faculty volunteers took the survey in advance and provided their feedback on its clarity, technological functioning, and overall design. This endeavor proved immensely helpful and is a necessary step in conducting any survey for assessment purposes.

Five pre-test interviews with faculty were conducted between February 6 and 14, 2019. A research staff person met with the faculty member in their office, remained available for questions or issues while the faculty member completed the survey, and then conducted a brief interview. The interviewer instructed the participant to take the survey as they normally would. This strategy allows participants to point out anything confusing or technologically flawed in real time and to provide feedback while it is still fresh in their minds.

Pre-test participants helped to identify questions that were unclear, confusing, or too wordy, and therefore likely to be skimmed over or skipped by busy respondents. For example, the survey originally asked, "What is the total number of semesters in which you have taught a GenEd course?" with a dropdown menu of numbers 1-50. Pre-test participants expressed confusion over whether this includes summer semesters and explained that, because GenEd at Temple is only a ten-year-old program, 50 was too high of an upper limit. It also became apparent in their data that respondents generally could not remember exactly how many semesters they had taught a GenEd course and were estimating (e.g., selecting 5 or 10). This led to clarifications within the question in the final version of the survey – respondents should indeed include summer semesters, response options were changed to ranges, and the maximum number of semesters was lowered to 15.

After the pre-test interviews were complete, the authors reviewed the data file. Many closed-ended questions included an "Other" choice where respondents could type in their own answer. Pre-test data were reviewed for common answers that were added as response options in the final questionnaire. If using skip logics, the data file can be checked to make sure those are working. For example, if a selection of "No" should skip the next two questions, but instead showed the next two questions, the skip logic needs to be fixed within the software. Lastly, the authors noticed that SurveyMonkey recorded pre-test participants IP addresses, and were able to fix the anonymity settings before launching the full survey.

Data Collection and Number of Respondents

The survey was programmed in SurveyMonkey, and respondents were recruited via email. The survey was open from March 13 to April 1, 2019. Three emails went out, an initial email and two reminders, to invite instructors to participate in the survey. The email included an explanation of the project, that it had been approved by the Temple IRB, that the survey was voluntary and anonymous, and that it would take about 15 minutes to complete.

The 1,022 instructors who had taught a GenEd course at Temple in the two academic years prior and for whom a university email address was on record were invited to participate. Overall, 287 individuals reported on 301 sections of GenEd courses. Fourteen instructors responded about two separate courses. Seven instructors reported that they had only ever

taught online GenEds, in which case they skipped ahead to Question 21, or the section on oral communication in online GenEds. Therefore, the bulk of the results relate to the 294 face-to-face courses. Overall, this yielded a response rate of 287/1,022, or about 23%.

Data Analysis

Analysis of the closed-ended questions was done using Stata 15, a statistical software package. Tables and graphs were made in Microsoft Excel. Verbatim survey answers in response to open-ended questions were stored in individual Excel files and coded by a team of coders using *grounded theory* (see Charmaz, 2014). Analysis under grounded theory involves cyclical rounds of open- and focused- coding. In other words, one does not bring predetermined categories to the open-ended survey responses, but rather lets the analytic categories emerge organically. The research team conducted this process separately for each of the eight qualitative questions in the survey. Two coders independently applied the eight coding schemes to their respective survey answers. At that time, the Kutools package in Excel was used to identify discrepancies in the two sets of codes. The discrepancies found by Kutools provided an opportunity for the team to discuss and strengthen the coding scheme (Extend Office, 2009). The two coders went through each discrepancy and decided on a final code together.

About 98% of the 294 face-to-face courses included at least one form of encouragement of oral communication. The most prevalent forms were "encouragement of questions," which was present in 83% of classes, and "encouraging participation in discussion," which occurred in 77% of classes.

Results

Forms of Oral Communication in GenEd Courses

As a reminder, oral communication was measured within the three areas of encouragement, instruction, and assessment. These results are presented below in Table 3. About 98% of the 294 face- to-face courses included at least one form of encouragement of oral communication. The most prevalent forms were "encouragement of questions," which was present in 83% of classes, and "encouraging participation in discussion," which occurred in 77% of classes. While less common, many respondents had students in their course deliver discussion facilitations, debates, and/or presentations that were ungraded; about 78% of respondents required at least one of these.

About 43% of courses included some instruction on oral communication. The most common form was providing a lesson on how to give oral presentations or oral communication more broadly, which occurred in about 27% of classes. Further, 59% of courses (or 173) included some form of formal assessment of oral communication. Among the 294 courses, 55% required a formal oral presentation, the most common form of oral communication assessment. Table 3 shows the breakdown in occurrence by solo or group presentations, as well as prevalence for other forms of assessment. In addition, 14% of courses included student facilitation of class discussion, 11% of courses included debates, and 4% of courses included some other form of graded oral communication assessment. There was also much variety in how instructors graded these assessments and how they presented those grades to students; Table 4 presents these results among the 173 courses that assessed oral communication.

Instruction/Assessment Gap

The survey asked respondents if they teach or assess any of the five criteria from the AAC&U Oral Communication VALUE rubric: *organization, language, delivery, central message, and supporting materials* (see Maki, 2015). Question 6 asked respondents who included some form of instruction whether they instruct on these specific criteria. Question 11 asked respondents who include some form of assessment whether they assess these specific criteria. The question did not specifically refer to the rubric itself, however they were provided with a definition of each term that stemmed from the rubric.

Graph 1 below shows that there are significant differences between the elements of oral communication that respondents assess, and those on which they instruct. The percentage of all 294 courses in which instructors teach these criteria ranges from 22.8% to 31% of classes, while the percentage of classes in which instructors assess them ranges from 31% to 51%. The average difference in percentage points between instruction and assessment for any given criteria is 15.17. The largest discrepancy between assessment and instruction

Table 3

Prevalence of Various Forms of Oral Communication in In-Person GenEd Courses

Encouragement	Frequency	Percent
Encouragement of questions	245	83.30%
Encouraging participation in discussion	225	76.50%
Break-out discussions/exercises	216	73.50%
Ungraded student presentations	58	19.70%
Ungraded discussion facilitations	54	18.40%
Ungraded debates	52	17.70%
Other	12	4.10%
Instruction		
Gave a lesson on oral presentations or oral comm.	79	26.90%
Gave specific feedback on oral presentations	69	23.50%
Gave a formal oral presentation as an example	36	12.20%
Other	35	11.90%
Assessment		
Student facilitation of class discussion	41	13.90%
Debate	32	10.90%
Other	12	4.10%
Formal Oral Presentations:		
Group presentations only	81	27.60%
Single-student presentations only	35	11.90%
Both	45	15.30%
Neither	133	45.20%
Total Number of Courses:	**294**	

Note: Percentages do not total 100% because respondents could select as many answer choices as they liked.
Source: Survey Questions 3, 4, 5, 7 & 8

Table 4

Prevalence of Grading Methods and Feedback Given regarding Oral Communication Among Courses including Assessment

Methods of Grading	Frequency	Percent
A rubric created by the instructor	105	60.70%
A general sense of student performance	73	42.20%
Peer evaluation	36	20.80%
Pass/fail based on certain requirements	19	11.00%
A rubric created by another source	17	9.80%
Forms of Feedback		
A letter or number grade	109	63.00%
Verbal feedback	81	46.80%
Written feedback	81	46.80%
A completed rubric	65	37.60%
Did not provide personalized feedback	12	6.90%
Other	9	5.20%
Total Number of Courses:	**173**	

Note: Percentages do not total 100% because respondents could select as many answer choices as they liked.
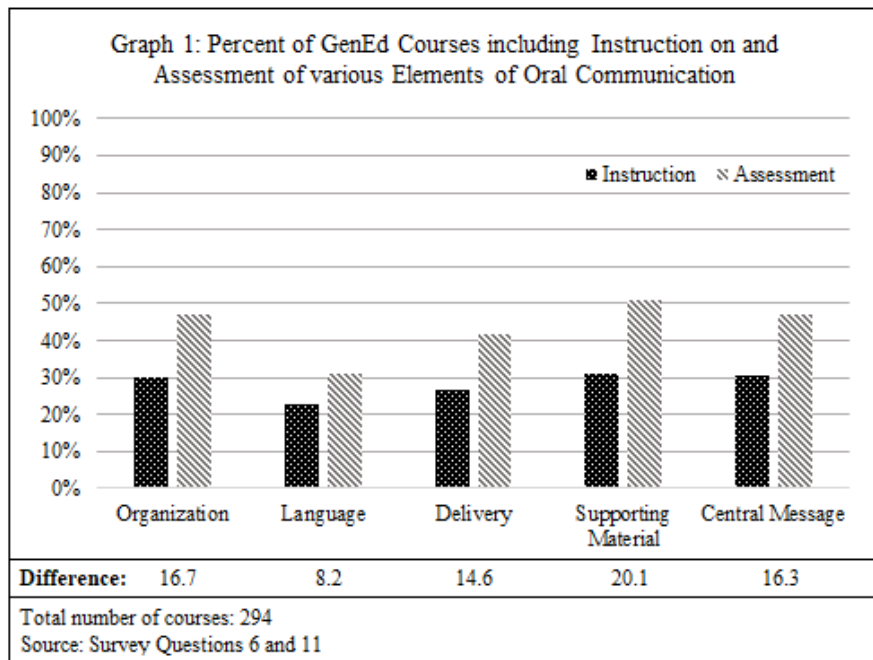Source: Survey Questions 7, 9, &10

was the difference around supporting material; more classes assessed supporting material than instructed on it by 20.07 percentage points. The rest of the percentage point differences can be seen in Graph 1. These data demonstrate that, in the GenEd program at large, students are being assessed on elements of oral communication for which they may have received little to no instruction, a very important finding.

Graph 1: Percent of GenEd Courses including Instruction on and Assessment of various Elements of Oral Communication

| Difference: | 16.7 | 8.2 | 14.6 | 20.1 | 16.3 |

Total number of courses: 294
Source: Survey Questions 6 and 11

Student Performance as Perceived by GenEd Instructors

Questions 17 and 18 asked respondents how they would rate the students' oral communication skills overall in the specific course they had in mind, and those questions can be seen in the survey instrument in the appendix. Overall, most respondents rated the students in their course as being either *inconsistent* in their oral communication skills (46.2%), defined as "partially proficient" or having a "wide inconsistency in *proficiency*," or proficient (41.0%), defined as "consistently proficient across a large portion of the class."

If they chose, the respondents could then provide comments in their own words, and various themes emerged. The most common theme among respondents was noticing variation in skill levels. Some noted that they could tell that "Upperclassmen (Juniors/Seniors) are more proficient in oral communication. Underclassmen (Freshmen/Sophomores) seem inconsistent… [some] can be introverted or 'afraid' to participate… Other times, they are hesitant to share an opinion…" Other respondents said they saw variation by college or department. One respondent who teaches students from various colleges and departments stated that "Some programs encourage and foster oral communication skills and others don't."

This leads to a second common theme of student reluctance to participate orally in class. Anxiety and fear of saying something wrong were commonly identified as key issues perceived by instructors. Respondents reported that some students are fearful or reticent as a result of being "discouraged from challenging opinions" in prior education. However, it was mentioned that once students overcome their fear, oral communication becomes easier for them. This issue, instructors explained, is exacerbated for those students for whom English is a second language (ESL). Respondents explained that "Non-native English speakers need extra help in their presentation skills." Some instructors specifically expressed a desire for more resources to help with teaching oral communication to the ESL student population.

Obstacles and Suggestions for Improvement

Question 29 asked if respondents had any more comments on this topic. One major theme was that oral communication is difficult to incorporate into large classes, mainly

These data demonstrate that, in the GenEd program at large, students are being assessed on elements of oral communication for which they may have received little to no instruction, a very important finding.

because the lack of time to assign oral communication-based projects, and because the grading can be so time consuming. One instructor stated that it is "actually impossible to assess or grade oral communication" in large classes and difficult to foster meaningful conversations. Respondents said that while small classes make grading more feasible, some still cannot find the time to assign oral presentations.

Respondents also expressed a need for general guidance for instructors on teaching and assessing oral communication. For example, one participant said, "It would be useful if instructors understood that having specific objectives and aligning assessments with those objectives is a standard and effective pedagogical practice. Even just having them understand the term assessment is important." Others cited the need for additional information regarding the GenEd program competency areas and program goals as a whole, saying, for example, "I would like to know more about what the aims are for the general education courses and how they differ from other courses."

This theme carried over into their requested resources. Some respondents expressed desire for a rubric, template, or some syllabus language that they could use to better incorporate the competency areas, including oral communication, into their classes. An additional suggestion to address this confusion was to provide workshops, seminars, or an instructional video for GenEd instructors. Lastly, as mentioned, a few respondents specifically requested additional resources for oral communication for instructors and students for whom English is their second language.

Discussion

Implementing an instructor survey in conjunction with direct assessment of oral communication proved highly useful in illuminating quantitative findings and suggesting the most effective means of acting upon those findings, or "closing the loop" (Roscoe, 2017). Surveying instructors highlighted issues with alignment between how oral communication competency is defined and taught, as well as issues with the rubric selected to assess the competency. As was shown in Table 3, the survey found that only about 55% of courses included formal oral presentations. As explained, the authors aimed to study a three-part theoretical framework of oral communication competency: informal and everyday uses, formal and graded uses, and the actual instruction or explicit lessons on the topic. Studying formal oral presentations alone would cover only about one half of one of these areas. As a result of this finding, the faculty committee tasked with overseeing the GenEd curriculum discussed whether instruction on and assessment of formal presentations ought to be required, or whether a modified assessment rubric was needed. They ultimately decided to promote a more expansive approach to teaching oral communication and to develop a rubric more closely aligned with that approach.

Survey results have also been used to inform faculty development efforts. As shown in Graph 1, the survey found that, for each of the five rubric criteria, a higher percentage of courses assess them than instruct upon them. This is partially because many courses (57%) did not have any explicit oral communication instruction. Some respondents made it apparent that they had never thought of instructing on oral communication. For example, one respondent wrote that "Explicit discussion of oral communication techniques and expectations is not something I had considered doing, but I will consider it for future courses." Another said taking the survey itself sparked them to incorporate oral communication instruction it into their class, saying "Your reminder about oral skills encouraged my demonstration interview in class. I had prepared to send my students off to do their own interviews without any instruction in how to do it." Therefore, even though oral communication instruction was not as popular as its more common counterparts of encouragement and assessment, instructors seemed hungry to hear more about this third prong and incorporate it into their courses.

Beyond instruction, participants shared feedback on other struggles including: incorporating oral communication into larger classes; aligning assessments with clear learning objectives; teaching oral communication to ESL students; and wanting some sort of template, syllabus language, or rubric to help them incorporate oral communication.

Although not discussed in the results highlighted, professors also expressed explicit struggles with incorporating oral communication into online GenEds. While some smaller classes incorporated oral communication through VoiceThread or similar software, many did not because instructors were unfamiliar with software options, had technological issues, found the grading to be too burdensome, or found students to be reluctant to film themselves.

Learning about these obstacles has led the GenEd program to convene a group of GenEd faculty volunteers to create a guide for addressing the oral communication competency, the goal of which is to walk instructors through selecting a specific oral communication skill, providing instruction on the skill, and designing an assessment. The list of oral communication skills was approved by the faculty committee overseeing the curriculum and aligned with the description of the competency for GenEd courses at the institution. The hope is that providing an expanded list of approaches to addressing oral communication will allow for flexibility in adapting to varying class sizes and instructional formats. Attaching the skill explicitly to instruction and assessment will promote the importance of giving students the opportunity to practice and develop the competency prior to assessment. The guide will be shared via a site for all GenEd instructors in the institutional Learning Management System. The next goal is to create and validate an assessment rubric that establishes benchmarks and can be disseminated and used to assess oral communication throughout GenEd.

The online instructor survey illuminated many forms of encouragement, instruction, and assessment of oral communication happening on campus. It also revealed what struggles instructors have around incorporating oral communication and therefore was able to inform faculty and resource development on campus.

Finally, findings from the survey also helped to flesh out findings from the quantitative direct assessment. In the direct assessment, not covered in depth here, the authors collected 100 videotapes of student formal oral presentations in GenEd classes and had faculty volunteers score those presentations using the AAC&U Oral Communication VALUE Rubric. From this, they were able to deduce averages from 0 to 4 for the five criteria of *language, central message, organization, supporting materials,* and *delivery*, finding that students performed the strongest on *organization* and *supporting materials*, and the weakest on *delivery*. *Delivery* also had the most variation in scores across students. Further, students in classes of fewer than 30 people performed significantly better on all four skills than students in larger classes, except for on *delivery*, for which class size had no impact.

Instructor comments from the survey helped explain why students performed the weakest on the rubric element of *delivery*, and why class size did not help explain the large variation in scores. Components of the AAC&U VALUE rubric element delivery include demeanor, eye contact, and expressivity, all of which can enhance the presentation and make the speaker seem authoritative. Respondents explained how students struggle with confidence and speaking anxiety, and that this was particularly true for first- and second-year students, as well as ESL students. Instructors also frequently commented on class size, noting how larger class sizes made it difficult to assign oral presentations. These narratives help explain why students in larger sections in which oral presentations were assigned did not perform as well as students in smaller classes on the elements of *language, central message, organization*, and *supporting materials*. Alignment between respondents' narratives around *delivery* and quantitative findings also helps to validate findings of the direct assessment.

In sum, when looking at the limited literature available on how to assess oral communication in a GenEd or similar program, Temple University was led to conduct a direct assessment of videotaped student presentations. However, they wanted to pair this with another method to paint a fuller picture of oral communication throughout their very large GenEd program. The online instructor survey illuminated many forms of encouragement, instruction, and assessment of oral communication happening on campus. It also revealed what struggles instructors have around incorporating oral communication, and therefore was able to inform faculty and resource development on campus. It is the authors' hope that this concrete example will help other programs pair a direct assessment of student work with an indirect assessment that gathers faculty practices and opinions to provide a comprehensive understanding of the competency and lead to improved teaching and learning across the board.

References

Avanzino, S. (2010). Starting from scratch and getting somewhere: Assessment of oral communication proficiency in General Education across lower and upper division courses. *Communication Teacher*, *24*(2), 91-110.

Babbie, E. (1973). *Survey research methods*. Belmont, CA: Wadsworth Pub Co.

Bradburn, N., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design—for market research, political polls, and social and health questionnaires*. San Francisco, CA: Jossey-Bass.

Charmaz, K. (2014). *Constructing grounded theory*. London: Sage Publications.

Crebert, G., Bates, M., Bell, B., Patrick, C.J & Cragnolini, V. (2004). Developing generic skills at university, during work placement and in employment: graduates' perceptions. *Higher Education Research & Development*, *23*(2), 147-165.

Crosling, G. & Ward, I. (2002). Oral communication: The workplace needs and uses of business graduate employees. *English for Specific Purposes*, *21*(1), 41-57.

Dunbar, N.E., Brooks, C.F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, *31*(2), 115-128.

Kennedy, R. (2007). In-Class debates: Fertile ground for active learning and the cultivation of critical thinking and oral communication skills. *International Journal of Teaching and Learning in Higher Education*, *19*(2), 183-190.

Extend Office. (2009). *Kutools-combines more than 300 advanced functions and tools for Microsoft Excel*. https://www.extendoffice.com/product/kutools-for-excel.html

NACE Staff. (2016). *Employers: Verbal communication most important candidate skill*. National Association of Colleges and Employers. https://www.naceweb.org/career-readiness/competencies/employers-verbal-communication-most-important-candidate-skill/

Maki, P. L. (2015). *Assessment that works: A national call, a twenty-first-century response*. Association of American Colleges and Universities.

Morreale, S., Backlund, P., Hay, E. & Moore, M. (2011). Assessment of oral communication: A major review of the historical development and trends in the movement from 1975 to 2009. *Communication Education*, *60*(2), 255-278.

Roscoe, D. D. (2017). Toward an improvement paradigm for academic quality. *Liberal Education*, *103*(1).

Saris, W. E. & Gallhofer, I.N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons, Inc.

Middle States Commission on Higher Education. (2020, January 2). *Standards for Accreditation and Requirements of Affiliation*. https://www.msche.org/standards/

Williams, D.E., McGee, B.R., & Worth, D.S. (2001). University student perceptions of the efficacy of debate participation: An empirical investigation. *Argumentation and Advocacy*, *37*(4), 198-209.

Appendix

Oral Communication in GenEd Courses: Instructor Survey

1. Do you consent to participating in this survey? (full consent form available upon request to authors)
    1. Yes (allows them to move onto the survey)
    2. No (does <u>not</u> allow them to move onto the survey) **[TO DISQUALIFICATION PAGE]**

Introduction

2. Please think of **one specific in-person GenEd course** that you are currently teaching or have taught in the past two years. Answer the questions with that course in mind. In which GenEd area is the course? (*Require response)*

    1. First Year Writing
    2. Quantitative Literacy
    3. Intellectual Heritage I
    4. Intellectual Heritage II
    5. Arts
    6. Human Behavior
    7. Race & Diversity
    8. World Society
    9. Science & Technology
    10. U.S. Society
    11. I have only taught online GenEd courses **[SKIP TO Q21]**
    12. I have not taught any GenEd courses in the past two academic years **[TO DISQUALIFICATION PAGE]**
    13. Don't recall/not sure **[TO DISQUALIFICATION PAGE]**

[DISQUALIFICATION PAGE MESSAGE:]
Thank you for interest, but this survey is for faculty who have taught a GenEd course in the past two years and willingly consent to participation.

The next few questions will ask you about whether the GenEd course you selected included:
1) Encouragement of,
2) Instruction on, or
3) Assessment of oral communication.

<u>Please do not include when a portion of the grade is dedicated to participation.</u> We will ask you about this separately.

Encouragement and Instruction

3. Did your course include any of the following that might **encourage** (but not require or assess) oral communication? (Select all that apply) (*Require response)*
    1. Encouraging, but not requiring, participation in class discussion
    2. Break-out (small group) discussions or exercises during class
    3. Regular encouragement of questions
    4. Ungraded/non-assessed student presentations
    5. Ungraded/non-assessed debates

5. Ungraded/non-assessed debates
6. Ungraded/non-assessed student-led discussion facilitations
7. None of the above
8. Other(s):

4. Did your course include any **instruction on oral communication**? This would be any form of specifically teaching students how to communicate effectively in verbal form. (*Require response)*

   1. Yes
   2. No **[SKIP TO Q7]**
   3. Don't recall/not sure **[SKIP TO Q7]**

Instruction Follow-Up

5. In what way(s) did you instruct on oral communication in this course? (Select all that apply)
   1. I gave a formal oral presentation specifically as a demonstration of what they should look like
   2. I gave a lesson on how to give good oral presentations (This could include discussing proper elements of oral presentations or how to overcome nerves while speaking, for example)
   3. I gave specific feedback on oral presentations aimed at student improvement
   4. Other(s):

6. When you instructed on oral communication, which elements did you address? (Select all that apply)

   1. Organization- Grouping and sequencing of ideas
   2. Language- Vocabulary, terminology, and sentence structure
   3. Delivery- Demeanor, eye contact, and expressivity while communicating
   4. Supporting Material- Use of explanations, examples, or other information and analysis to support ideas
   5. Central Message- Offering a clear, easily identifiable central message
   6. Don't recall/ Not sure
   7. Other(s):

Assessment

7. Did your course include any **formal assessment of oral communication**? This includes any situation where students are required to communicate orally and are graded on it. (*Require response)*

   1. Yes
   2. No **[SKIP TO Q12]**
   3. Don't recall/not sure **[SKIP TO Q12]**

Assessment Follow-Up

8. In what way(s) did you assess oral communication in this course? (Select all that apply)
   1. Single-student oral presentations
   2. Group oral presentations
   3. Student facilitation of class discussion
   4. Debate
   5. Other(s):

9. When you assessed oral communication skills in your course, how did you determine student grades? (Select all that apply)
    1. A general sense of student performance
    2. A rubric I created
    3. A rubric created by another source
    4. It was pass/fail based on certain requirements
    5. Peer evaluation (students evaluated each other)
    6. Don't recall/Not sure
    7. Other(s):

10. What form(s) of feedback, if any, did you provide to your students around their oral communication skills? (Select all that apply)
    1. I did not provide personalized feedback
    2. A letter or number grade
    3. A completed rubric
    4. Verbal feedback
    5. Written feedback over email, embedded in Canvas or other software, or on paper
    6. Don't recall/Not sure
    7. Other(s):

11. When grading your students on oral communication, which of the following elements did you consider? (Select all that apply)
    1. <u>Organization</u>- Grouping and sequencing of ideas
    2. <u>Language</u>- Vocabulary, terminology, and sentence structure
    3. <u>Delivery</u>- Demeanor, eye contact, and expressivity while communicating
    4. <u>Supporting Material</u>- Use of explanations, examples, or other information and analysis to support ideas
    5. <u>Central Message</u>- Offering a clear, easily identifiable central message
    6. Don't recall/Not sure
    7. Other(s):

Class Participation

12. Was class participation a percentage of the final grade for students in your class? *(Require response)*

    1. Yes
    2. No **[SKIP TO Q16]**
    3. Don't recall/not sure **[SKIP TO Q16]**

Participation Follow-Up

13. Were students required to verbally participate in class to receive full credit for participation?
    1. Yes
    2. No
    3. Don't recall/not sure

14. Did you utilize any of the following methods in determining participation grades for the course? (Select all that apply)
    1. Tallying counts of class contributions for each student
    2. Assigning a daily participation grade for each class session
    3. Referring to alternative forms of participation that counted toward the grade

    4. A rubric I created
    5. A rubric created by another source
    6. None of the above
    7. Other(s):

15. What forms of participation other than verbal contribution, if any, could count toward students' participation grades? (Select all that apply)
    1. None
    2. Written participation in online discussion boards
    3. Written participation via emails with the professor
    4. Participation in-class in small group work/discussions
    5. Appearing visibly engaged
    6. Other(s):

Enrollment

16. Approximately how many students were enrolled in this course?
    1. 20 or fewer
    2. 21-30
    3. 31-40
    4. 41-50
    5. 51-60
    6. 61-70
    7. 71-80
    8. More than 80 students
    9. Don't recall/not sure

Perception of Student Skill Level

17. How would you rate the students' oral communication skills in your class overall? This refers to things like delivering clear messages, relying on supporting ideas, and speaking clearly and with confidence.

    1. Excellent- Skillful and excellent for some students and wide proficiency for others
    2. Proficient- Consistently proficient across a large portion of the class
    3. Inconsistent- Partially proficient or wide class inconsistency in proficiency
    4. Developing- Low level of proficiency for many students in the class
    5. Don't recall/not sure
    6. I did not see enough student oral communication to answer this question

18. Would you like to add any comments on your overall impression of GenEd students' oral communication skills in this class?

Continue or Loop

19. Before moving to the few remaining questions of the survey, would you like to answer these questions again for another **in-person** GenEd course you have taught in the past two years? *(Require response)*
    1. Yes **[LOOP QUESTIONS 2-18]**
    2. No **[CONTINUE TO Q20]**

Online Courses

20. Have you ever taught an online GenEd course? *(Require response)*
    1. Yes
    2. No **[SKIP TO Q26]**

21. In which online formats have you taught GenEd courses? (Select all that apply)
    1. Fully online asynchronous
    2. Hybrid: online with periodic in-person meetings
    3. Virtual: online with regular virtual meeting times

22. In general, do you require students to use VoiceThread or another online software to record themselves in audio and/or video and post it?
    1. Yes
    2. No
    3. Depends on the course

23. Please explain why or why not.
24. How, if at all, did you incorporate oral communication into your online GenEd course(s)?
25. Please explain any challenges you faced in incorporating oral communication into online GenEd course(s).

Instructor Characteristics

26. How much do you feel you know about the GenEd competency areas, or the eight categories of skills and abilities that the program aims to teach to students overall?
    1. I am very familiar with them
    2. I am moderately familiar with them
    3. I know a little about them
    4. I have heard of them
    5. I do not know what the GenEd competencies refers to

27. Were you aware that one of the GenEd program competencies includes oral communication before taking this survey?
    1. Yes
    2. No
    3. Don't know/not sure

28. How many times have you taught a GenEd course at Temple? Please include summer courses.
    1. Once or twice
    2. 3-5 times
    3. 6-9 times
    4. 10-14 times
    5. 15 or more times

29. Is there anything else you would like to add?
    This might include ideas for improvement, resources that would help you, or any specific obstacles you face in teaching or assessing oral communication skills.

AUTHORS

Nicholas A. Curtis, Ph.D.
Marquette University

Robin D. Anderson, Psy.D.
James Madison University

Ray Van Dyke, Ed.D.
Weave Education

CORRESPONDENCE

Email
nicholas.curtis
@marquette.edu

From the Editor

The education, training and development space for assessment practitioners has not been clearly mapped. This has resulted in an unclear training path for new assessment practitioners, many of whom have advanced degrees in other disciplines. It remains a common occurrence that individuals new to the field of assessment are directed to conferences for their training and development. Conferences, usually designed for development, likely fail to provide the breadth and depth of training needed for most new assessment practitioners. Fortunately, there are a number of other methods designed to provide training such as non-graduate and graduate credit courses, certificates, and graduate degree programs. Unfortunately, without a clear map of how assessment education and training contribute to the assessment practitioner, many new assessment practitioners may experience a mismatch between their learning goals and the method of education, training or development that they experience. This article focuses on "how" assessment practitioners are trained, the available methods for training and development, and the appropriate placement of each within the space of assessment practitioner development.

A Field Without A Discipline? Mapping the Uncertain and Often Chaotic Route to Becoming an Assessment Practitioner

*L*et us face facts; no one grows up saying they want to be a student learning outcomes assessment practitioner. A few people "find" assessment, but most happen, or are encouraged, into it. A few people move purposefully into the field of student learning outcomes assessment (henceforth referred to simply as 'assessment') after finding they enjoy statistics and measurement. Some move purposefully into assessment after developing a love for higher education and wanting to "make a difference." Sometimes it is both. But most assessment practitioners trip into the assessment field. We have met assessment practitioners who were previously in institutional research and "inherited" outcome assessments as part of their expanded role. We have encountered faculty members who were "voluntold" to assume responsibility for their institution's assessment initiatives after having served as the chair of the institution's assessment committee for as little as a year. In a recent Association for the Assessment of Learning in Higher Education (AALHE) poll of practitioners, approximately 71% of respondents indicated less than 5 years of professional work in their current assessment role (Ariovich, et al., 2018). Those new to assessment fill a majority of the senior assessment positions in higher education. The truth is, assessment practitioners come from a wide variety of disciplines and from nearly every corner of the academy (Leaderman & Polychronopoulos, 2019; Polychronopoulos & Leaderman, 2019). Our paths to the profession are nearly as numerous as we are. This diversity strengthens our profession in many ways and it also creates a significant challenge. With so many of us

coming to the profession with terminal degrees in hand, the training of most assessment practitioners does not follow a traditional disciplinary path. There are exceptions of course, but for those who did not find a PhD, PsyD, or EdD in assessment along the way, the path to getting up to speed on student assessment is often a self-directed one. It is not our intent to discount the diligent work of previous and current assessment professionals who arrived at the profession through a winding path. In fact, all of the authors on this paper started outside of assessment. We are suggesting that the discipline has matured to the point of needing a clearer path to becoming an assessment professional.

With the exception of a few doctoral-level programs that have emerged over the last 20 years, most training and development has come in the form of conference sessions, workshops, webinars, and online events. In fact, we have yet to attend an assessment conference without welcoming a new colleague who has been sent by a college administrator to a conference "to learn how to do assessment." However, for that individual, the experience is akin to being 'thrown into the deep end' before being taught how to swim. Such professional events seldom scaffold knowledge, nor do they deliver a purposeful curriculum. This means that some of us who attended the conference, have some knowledge and skills, while others who attended other conferences, may have different skills and knowledge. More recently non-graduate and graduate credit courses and certificate programs have emerged. However, there is no accepted core knowledge for the field and no intentional method for delivering such a core even if there were. While others have begun to address the first issue (Horst & Prendergast, 2020), in what follows, we discuss the second issue, by what method or methods should those new to assessment practice rely on to gain knowledge about assessment? How does the field map the training, development and educational space for assessment professionals? As part of our discussions, we will examine the appropriate placement of each method within the space of assessment professional development and the potential impact the current system has on the professional development of the field and the development of advanced practice. Of course, we are not the only assessment practitioners to consider this issue in one form or another. There have been calls to consider the reasons we engage in assessment (Jankowski, 2017); examinations of established roles and tasks of assessment professionals (Jankowski & Slotnick, 2015; Nicholas & Slotnick, 2018); and encouraging efforts to guide current professionals in examining their own development and how they approach their work with other professionals (Leaderman & Polychronopoulos, 2019; Polychronopoulos & Leaderman, 2020). Building on these previous calls, our hope is that the following discussions bring further attention and spur more widespread consideration of these issues.

Novice versus Intermediate versus Advanced Practitioners

For their taxonomy, Horst and Prendergast (2020), defined three levels of assessment practitioners. Their Assessment Skills Framework (ASF) outlines knowledge, skills, and attitudes for the novice, the intermediate and the advanced practitioner.

> Specifically, someone at the novice level is described as being able to provide basic explanations of assessment concepts and apply that knowledge to hypothetical examples devoid of context and real-world complexity...The intermediate level is described as being able to provide a more thorough explanation of assessment concepts than someone at the novice level. ...People at the intermediate level successfully apply their knowledge to real-life assessment projects. People at the advanced level are described as being able to provide a nuanced explanation of assessment concepts. These individuals use reflective thinking about their assessment practice that results in the generation of new knowledge or useful alternative conceptions about assessment processes. (p.7)

It is important to note that the authors not only distinguish among the levels based on what the practitioners know and can do, but also practitioners' confidence, their ability to generate new knowledge, and their ability to develop new approaches (Horst & Prendergast, 2020).

A few people "find" assessment, but most happen, or are encouraged, into it. A few people move purposefully into the field of student learning outcomes assessment (henceforth referred to simply as 'assessment') after finding they enjoy statistics and measurement. Some move purposefully into assessment after developing a love for higher education and wanting to "make a difference." Sometimes it is both. But most assessment practitioners trip into the assessment field.

Defining Training, Development, and Education

Even the newest assessment practitioner quickly learns to value the importance of operationally defining terms. To begin, we look at how to define and distinguish between training, development, and education. A review of the literature reveals that there are multiple models (for example, see Bhatia, Rao, & Bhatia, 2019) for differentiating training and development. In their article, Horst and Prendergast (2020), focusing more on the skills acquired rather than the methods, referred to all delivery methods as "professional development" opportunities. For our purposes, we believe the following definitions, adapted specifically for assessment practitioners represent a useful nuance within the wide array of opportunities. *Training* is a process by which the assessment novice can learn the *key* skills required to successfully conduct student learning outcomes assessment. *Training* is usually seen as a short-term process that focuses on building basic skills and knowledge for one's current position. *Development* is the process by which those individuals at advanced levels cultivate their skills, preparing for more advanced positions, addressing future challenges, or advancing practice. *Development* is more of an ongoing process that is future focused.

Unlike earlier definitions of *assessment professional* (e.g., Nicholas & Slotnick, 2018), we believe a differentiation between assessment *practitioners* and assessment *professionals* is helpful. We propose that assessment *professionals* require not just more advanced proficiency in the same skills, knowledge, and attitudes as assessment *practitioners*, but also additional areas of expertise such as qualitative, quantitative, and mixed research methods; advanced statistical skills; and/or an understanding of public and educational policy as it pertains to assessment practice. This, of course, leaves the question of what is *education* in terms of developing an assessment professional? This broad *education* can be acquired through formal graduate or doctoral programs or through a vast accumulation of training, development and practice opportunities.

Mapping the Training, Development and Education Space

We propose that assessment professionals require not just more advanced proficiency in the same skills, knowledge, and attitudes as assessment practitioners, but also additional areas of expertise such as qualitative, quantitative, and mixed research methods; advanced statistical skills; and/or an understanding of public and educational policy as it pertains to assessment practice.

Prior to mapping out the available methods for training and development, we must understand where training, development, and education can each be most useful to practitioners looking to advance their proficiency. Figure 1 below demonstrates that training offers opportunities for practitioners to be introduced to new skills and knowledge. The number of training sessions required to assist an individual in moving from novice to intermediate on a Knowledge, Skill, or Ability (KSAs) depends on the complexity of the topic. For example, in Figure 1, if the KSAs are at the novice level of the Assessment Skills Framework (Horst & Prendergast, 2020) it may take practitioners only one (for least complex KSAs), or two (for somewhat complex KSAs) training experiences to move to the intermediate level. For example, a practitioner may quickly provide "basic descriptions of each step of the assessment cycle" (p.9); yet, it may take many more experiences before that practitioner "defines validity threats in the context of research design" (p. 17), a more complex KSA. Once these foundational KSAs are learned, then developing assessment practitioners may wish to engage in professional development activities to advance their skills, prepare for a more advanced or different professional position, address developing and future challenges in assessment practice, or advance practice itself.

It is important to note that an assessment practitioner, regardless of their overall assessment competency, may be a novice, intermediate, or advanced practitioner on any number of assessment-related knowledge, skills, or abilities. Advancing from *novice* to *intermediate* to *advanced* on the KSAs should not be equated with overall years of experience nor with institutional role or title (ACPA/NASPA, 2015). For example, an experienced assessment practitioner may be an expert in many areas of assessment, including quantitative data analysis approaches, but be a novice in qualitative approaches. Figure 1 intentionally addresses a single, hypothetical assessment practitioner's level of specific KSAs, not their overall competence in assessment practice. The figure outlines two possible ways for a practitioner to develop assessment-related KSAs. First, practitioners might structure their own learning and development through a series of trainings, workshops, and applied experiences. Second, an assessment practitioner might enroll in a more structured full- or part-time program in assessment practice, represented by the dotted pattern in Figure 1. Practitioners in these

programs often move from *novice* to at least advanced on many individual KSAs. However, as new processes and practices develop, these professionals may also find they are in need of training and development to move from *novice* to *intermediate* to *advanced* in areas they had not focused on during their formal education. If the practitioner in Figure 1, as an example, had engaged in this hypothetical formal educational program, they would not yet develop to the *professional*-level in either of the first two KSAs. Thus, a future set of professional development experiences would be necessary to develop those KSAs further. Either self-structured or formal educational experiences have the potential to help practitioners looking to further develop their KSAs as long as the experiences are well matched to the level of KSA being developed.
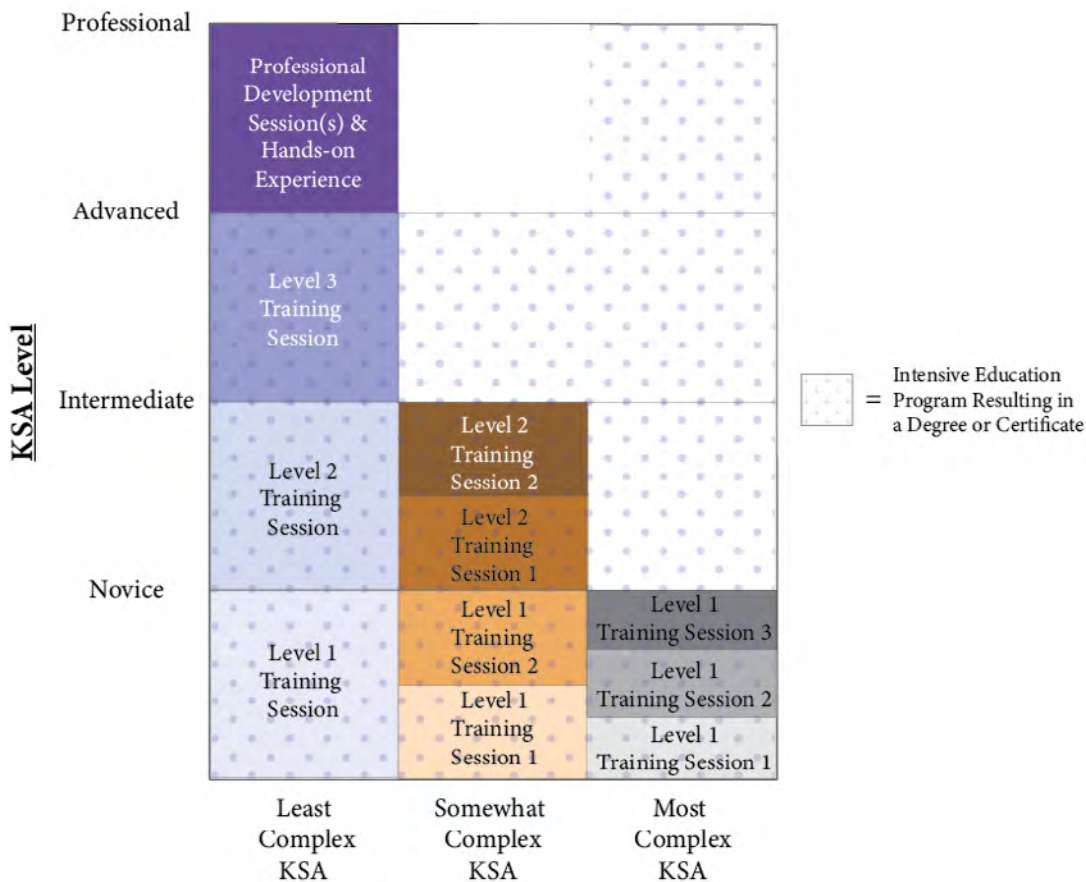
What Practitioners Want

We do not intend this discussion piece as a formal research article. Yet, when considering how to structure a training, development, or education path, it is important to consider what assessment practitioners want from such experiences. Thus, as we considered training and development experiences, we examined results from three related surveys and one market research report to determine what assessment practitioners might want and what they may need from a training, development, and educational standpoint. It is important to note that none of these sources distinguished between training, development, and education. Yet, the results can provide some insight into practitioners' preferences for these kinds of experiences.

In the Association for the Assessment of Learning in Higher Education (AALHE) poll of practitioners, mentioned above (Ariovich, et al., 2018), practitioners expressed a clear preference for professional development opportunities delivered in the form of conferences, webinars, journals, and training sessions. We conducted a similar, informal survey of attendees

at the Virginia Assessment Group annual meetings in November of 2018 and 2019. The responses continued to support demand for an assessment training program and indicated that such a program should ideally have some face-to-face, synchronous components. Some respondents indicated that affiliation with a major college/university was not always necessary.

In order to build on these surveys, one for-profit assessment software company conducted a series of market-research tests to collect more specific information (Weave, 2020). Findings indicated a strong desire for access to a knowledge base and participation in a professional community of active learners. Respondents also requested participation in on-line training courses and access to expert consultants. Based on this data, the company offered six sessions of two certification courses in assessment from spring 2019 through spring 2020. Upon completion of each course, participants were asked to respond informally to a survey related to their thoughts about the course, with items addressing the course organization, instructors, assignments, and other items. One of the items asked participants to respond to the question "What did you like most about the course?" Responses to this question could be categorized within three topic areas: (a) interactions with peers and experts, (b) acquisition of knowledge and skills, and (c) applicability/relevance to their current work. The interaction with others, peers and experts, was the most often stated reason for appreciating the course.

In summary, the findings from these sources indicate that assessment practitioners have a preference for: (a) training with at least some face-to-face component; (b) participation in a professional community of active learners; (c) access to a knowledge base of highly effective training tools and artifacts based on best practice; and (d) access to expert consultants on a variety of important educational effectiveness topics. As we contemplate how assessment training and professional development is currently structured and how it might be restructured, it is important to consider the factors important to those who would be engaging in these experiences.

Implications for Practitioner Training and Development

So, why is distinguishing between *training* and *development* (and in some cases *education*) important? Assessment in higher education, as a discipline, continues to develop rapidly. Each year, while there are still newcomers to assessment, there are also more and more intermediate and advanced practitioners. This increased diversity in skills, knowledge, and ability presents challenges for the field. Our conferences, once geared entirely for those new to assessment, are having to shift to satisfy multiple audiences. Such a shift is important, because a mismatch between method of training/development/education and the educational goals may result in a lower quality learning experience, a lower feeling of satisfaction from participants and facilitators, and a misuse of valuable time and resources that could have been directed more appropriately. For example, if a practitioner who falls at an intermediate level on a specific skill signs up for a webinar that sounds like it is a form of professional development, but it provides basic training, the practitioner is not going to be highly satisfied nor, perhaps more importantly, will their skill level improve. Conversely, if a novice signs up for a conference session that is marketed as training, yet provides intermediate or advanced professional development, it is unlikely that person will have acquired the foundational concepts required to benefit from that session.

Many assessment conferences, unfortunately, may fall into the role of covering largely introductory topics as their organizers are aware of the current niche they play in training up the novice practitioner. However, this often leads to a shortage of development opportunities for the advanced practitioner. Yet, due to the high turnover of assessment practitioners, we cannot simply shift all conference activities to focus on intermediate, advanced, and professional topics. We are aware that many conferences label their sessions with categories such as beginner, intermediate, and advanced. Yet, what these labels mean varies from presenter to presenter and also from conference to conference. A clear understanding of what is *training* and what is *development* could serve as a guide for selecting and accurately labeling delivery methods (i.e. conference presentation, webinar, graduate course).

With that in mind, we offer the following taxonomy of training, professional development, and education for further discussion. First, we suggest, in Table 1, that training

efforts be aligned with the three levels of the Assessment Skills Framework (Horst & Prendergast, 2020). We further suggest that the training levels might be best considered as relevant to experience with specific KSAs. For example, level one training would be for those with no experience in assessment and designed to introduce *novice-level* KSAs. As practitioners gain additional experience applying what they have learned in their training, training at the higher levels becomes more appropriate. We also propose that professional development occur only after all three levels of training have been accomplished within an area of study.

Table 1
Proposed Taxonomy of Training, Professional Development, and Education

	Target Level of Audience	Purpose and Targeted Level of ASF*
Professional Development	Has demonstrated consistent and correct application of *Advanced*-level of the KSA	To reinforce *Advanced*- level KSAs & provide a platform for advancing the field and developing novel KSAs
Training - Level 3	Has demonstrated consistent and correct application of *Intermediate*-level of the KSA	To reinforce *Intermediate*- level KSAs & introduce *Advanced*- level KSAs
Training - Level 2	Has demonstrated consistent and correct application of *Novice*-Level of the KSA	To reinforce *Novice*- level KSAs & introduce *Intermediate*- level KSAs
Training - Level 1	No Experience with the Assessment-related Knowledge, Skills, and Abilities (KSAs)	To introduce *Novice*- level KSAs (KSAs)

Assessment Skills Framework (Horst & Prendergast, 2020)

Second, building on this taxonomy, we then move to mapping each level of training/development to ideal methods for those seeking to advance their KSAs. Table 2 provides our initial thoughts on some of the methods available to us that might be best equipped for each level. We believe that any method (including many not listed here) might be adapted to meet the needs of practitioners at any level. Yet, we also believe that some methods are more naturally aligned with certain levels and thus, might be best able to help practitioners efficiently and effectively develop the important knowledge, skills, and abilities at those levels. We intentionally present this information divorced from content in an attempt to resist conflating KSAs with methods of delivery.

We propose this taxonomy, not for the purposes of trying to propose a definitive structure, but instead we seek to spur discussion across the field of assessment so that we might come to a consensus together; a consensus that boosts our collective efforts to advance the field. We trust that, with the benefit of multiple perspectives across the field, our suggestions can be refined.

We propose this taxonomy, not for the purposes of trying to propose a definitive structure, but instead we seek to spur discussion across the field of assessment so that we might come to a consensus together that boosts our collective efforts to advance the field.

Conclusion

Assessment is a rapidly changing professional field. For more than three decades, many people, enthusiastic about the promise of assessment, have worked to develop and apply increasingly complex and sophisticated methods of assessment. Many, if not most, assessment professionals shifted their professional focus to develop such knowledge, skills, and abilities subsequent to developing a professional identity in another discipline (e.g., psychology, biology, English, etc.). Now, with a sufficient base of knowledge, skills, and abilities that might begin to define the scope of the assessment field, we believe it is time for a paradigm shift: from assessment as a secondary profession to a primary one. As we mentioned at the start of this paper, it is not our intention to discount the diligent work of previous and current assessment professionals who arrived at this point through a winding path. The authors on this paper started their careers as a K-12 educator, a school psychologist, and a mental health counselor. However, like people and organizations, disciplines develop over time. After 30+ years, it is time we move past drafting or

Table 2

Proposed mapping of training/development methods to training and development levels

	Training - Level 1	Training - Level 2	Training - Level 3	Professional Development
Short Conference Presentations				X
Online Single Session Presentations				X
Interactive Single Session Webinars				X
In-Person Workshops (< 2 hours)			X	X
In-Person Workshops (2-6 hours)		X	X	
In-Person or Online Multi-Day Workshops	X	X		
Certificate Programs	X	X	X	
Master's Degree Programs	X	X	X	
Doctoral Degree Programs	X	X	X	X

We hope that this article spurs a profession-wide conversation to better align our collective efforts to produce the best possible assessment professionals.

"voluntelling" people to step into assessment roles and instead move assessment into a professional, legitimate, and defined discipline with a clear path for becoming an assessment professional. The straightest path, like for any discipline in higher education, is through a terminal-level degree. However, there will always be a need for training and professional development. Even assessment professionals with a doctorate in the discipline will need to learn new KSAs as new methods and practices emerge. And as assessment professionals, we work with partners all across campus, including administrators, faculty members, staff, and students. These partners will also continue to want and need training.

In order to achieve this vision, we first need to collectively be more transparent about the goals of our current training and development work. We also need to better map our chosen goals to the wide array of training and development methods available to us. We believe that all training and development methods have strengths and weaknesses that lend themselves best to particular levels of training and development. By being more transparent about the KSAs we are trying to deliver and at what level, we hope to reduce the number of incidences where advanced practitioners are jumping into level one trainings because no advanced professional development activities exist or where novices are sitting in sessions that are being pitched well-above their level of understanding. In addition, if we put a framework of KSAs around the assessment discipline, those KSAs will likely filter into job descriptions for assessment professionals, further increasing the need to ensure that our training and development methods align well with the KSAs in order to help develop people qualified to fill those positions. As we state above, what we present here is a draft, a draft framework of how we may be able to better align the level of training and development with methods of delivery. We hope that this article spurs a profession-wide conversation to better align our collective efforts to produce the best possible assessment professionals.

References

ACPA—College Student Educators International, & NASPA—Student Affairs Administrators in Higher Education. (2015). Professional competency areas for student affairs educators. https://www.naspa.org/images/uploads/main/ACPA_NASPA_Professional_Competencies_FINAL.pdf

Ariovich, L., Bral, C., Gregg, P., Gulliford, M., & Morrow, J. A. (2018). The assessment profession in higher education: Addressing the varied professional development needs of practitioners. *Lexington, KY: Association for the Assessment of Learning in Higher Education and Watermark*. shorturl.at/aoqDO

Bhatia, S., Rao, D.N., & Bhatia, D. (2019). Competency models for training & development: A review. *IITM Journal of Business Studies, 6*(1), 56-64. https://doi.org/10.1111/1468-2419.00069

Clucas Leaderman, E., & Polychronopoulos, G. B. (2019). Humanizing the assessment process: How RARE model informs best practice in educational assessment. Research & Practice in Assessment, 14(Summer/Fall 2019), 30-40. Retrieved from http://www. rpajournal.com/dev/wp-content/uploads/2019/07/A2.pdf.

Horst, J. & Pendergast, C. (2020).The Assessment Skills Framework: A taxonomy of assessment knowledge, skills and attitudes. *Research and Practice in Assessment*, 15, 1-25. shorturl.at/gyFM6

Jankowski, N. A. (2017). Moving toward a philosophy of assessment. *Assessment Update, 29*(3), 10-11. doi:10.1002/au30096

Leaderman, E. C., & Polychronopoulos, G. B. (2019). Humanizing the Assessment Process: How the RARE Model Informs Best Practices. *Research & Practice in Assessment*, 14, 30-40.

Polychronopoulos, G. B., & Leaderman, E. C. (2019). Strengths-based Assessment Practice: Constructing Our Professional Identities through Reflection. https://www.learningoutcomesassessment.org/wp-content/uploads/2019/07/Viewpoints-Polychronopoulos-Leaderman.pdf

Weave (2020). Internal document. *www.weaveeducation.com*

Authors

Nicholas A. Curtis is the director of assessment at Marquette University. After working as a school psychologist in K-12 public education, he transitioned to the field of assessment earning a Ph.D. in Assessment and Measurement in 2018.

Robin D. Anderson is a professor and academic unit head at James Madison University where she also coordinates the JMU Higher Education Assessment Specialist Graduate Certificate program. After working in the field of mental health counseling, she transitioned to the field of assessment earning a PsyD degree in Assessment and Measurement in 2001.

Ray Van Dyke is Senior Vice President at Weave and Founder and Executive Director of Weave Academy. Previously, he was the Executive Director of the Office of Assessment and Evaluation at Virginia Tech. He holds a doctorate in Educational Leadership and Policy Studies from Virginia Tech.

Book Review

*Using Evidence of Student Learning
to Improve Higher Education*
George D. Kuh, Stanley O. Ikenberry,
Natasha A. Jankowski, Timothy Reese Cain,
Peter T. Ewell, Pat Hutchings, Jillian Kinzie.
San Francisco, CA: Jossey-Bass, 2015. 304pp.
ISBN-13: 978-1118903391. Hardcover, $36.00.

REVIEWED BY:
Jason P. Browning, MBA
Dixie State University

Few topics have permeated the higher education landscape in the past several years more than academic assessment and the issue of how colleges can better assess student learning. From legislators and politicians, to the general public and the media, there are seemingly endless calls for colleges and universities to share evidence of their effectiveness. George Kuh, the founding director and senior scholar at the National Institute for Learning Outcomes Assessment, joins Ikenberry, Jankowski, Cain, Ewell, Hutchings, and Kenzie (2015) to outline approaches to the assessment process in *Using Evidence of Student Learning to Improve Higher Education*.

The authors describe the catalyst for this work, noting that "[t]his volume grows out of a deep concern that the practical value of otherwise well-conceived efforts to assess student learning in American higher education is often diminished by deeply nested misconceptions" (p. ix). Academic assessment is seen primarily as a compliance task for accreditors or administration, rather than as a tool that can be used to inform or improve student learning. The book endeavors to refute this conventional view, which is examined more thoroughly in Chapter 1. Although compliance requirements are real and must be fulfilled, the authors strive to demonstrate the use of assessment of student learning as a catalyst for continuous improvement and as a barometer of student success and institutional health (p. x).

Chapters 2 through 4 focus on the fundamental task of finding and using evidence of student learning. These chapters establish a foundation for the remainder of the text by providing an overview of what we may think of as more traditional academic assessment, noting succinctly that "evidence is essential to improving student learning and responding to accountability expectations" (p. 27). Throughout these chapters, however, the authors go beyond the traditional view of assessment and evidence gathering.

Chapter 2 focuses on three primary questions about assessment evidence: "1. What are the different sources and properties of assessment evidence now in use and what can be said of their respective strengths and limitations for stimulating improvement? 2. What are the obstacles to the effective use of evidence? 3. What counts as evidence for different audiences and purposes?" (p. 29). The responses to question one will be familiar to most assessment practitioners;

surveys, general knowledge and skills tests, classroom assignments, portfolios, rubrics, and student analytics are discussed as potential sources of evidence.

The authors note a common concern when they write, "Also, a nagging question persists: Does the availability and use of evidence of student learning make a material difference to improving student learning and institutional performance?" (p. 53). Chapters 3 and 4 address in greater detail how to encourage use of assessment results to promote effective practice, and how to make assessment consequential. The book provides several examples of effective use of assessment results in Chapter 3, including practices at Texas A&M International University, Carnegie Mellon University, Augustana College, Richland College of the Dallas County Community College District, Georgia State University, and others. In addition to these illustrative practices, the authors provide tangible guidance, such as linking assessment to both internal and external processes (e.g., academic department and program review) and keeping the ultimate intended use of the assessment in mind as practices are designed. Chapter 4 focuses on the practical challenge of organizing for assessment and how to design an effective assessment function.

As many seasoned practitioners can attest, the more difficult part of an effective assessment regimen can often be the engagement of key stakeholders, which is the primary focus of Chapters 5 through 8. These chapters progress through internal and external stakeholders, focusing first on faculty and students, and then institutional leadership, accreditors, and other external entities. In Chapter 5, the authors note that both faculty and students have a role to play in assessment, and that "[e]xplicitly bringing students into assessment activities strengthens that partnership and underscores the fact that assessment is about learning, not about reporting" (p. 107). As in earlier chapters, the authors provide several examples of successful institutional models that involve both faculty and staff in the assessment process.

Chapter 6 emphasizes the role of the institution's governing board, president or chancellor, provost or chief academic officer, and deans or department chairs in assessment processes. The authors discuss the necessity of consistent and aligned support from leaders throughout the institution to move from a compliance orientation toward a culture that uses assessment to facilitate continuous improvement. Chapter 7 addresses this shift further, noting that "assessment at its best contributes to both accreditation and internal institutional processes" (p. 157). The need to reframe the culture of compliance is again mentioned in Chapter 8, which notes that an effective assessment system that focuses on continuous institutional improvement will largely satisfy accountability demands by virtue of its existence: "External needs for evidence of student learning will not diminish, but they can be met more rationally and efficiently by focusing first on the needs of students and the campus itself" (p. 180).

Part Three of the book focuses on the next steps in the evolution of assessment and reiterates the necessity of maintaining an assessment program that is acutely focused on student learning. The common problem across campuses of assessment and initiative fatigue is explored in Chapter 9. While many assessment tomes focus on implementation of assessment and creation of effective practices, the important issue of acceptance on campus is often skirted or dismissed. In this book, Kuh et al. (2015) address directly the familiar reluctance and cynicism on campus. The authors outline several factors that can contribute to initiative fatigue, and explicitly address the heightened likelihood of assessment to exacerbate this phenomenon. Several strategies are then provided as potential ways to diffuse and diminish initiative fatigue.

Chapter 10 rehashes and further examines the transition from compliance reporting to effective communication, with a particular focus on transparency. Importantly, the chapter distinguishes between the traditional disclosure and making data available as transparency and a more coherent system of transparency. Data and information must be shared with context and interpretation in order to foster true transparency. As the authors note in their summarization of the chapter, "[t]o be transparent about student learning outcomes… institutions need to consider how best to tell the story, to present relevant contextual information, and to help the target audiences grasp the implications" (p. 219).

The constant theme of converting assessment from a compliance burden to a value-added activity permeates the book. It is fitting, then, that the concluding chapter takes a macro-level view of assessment through this theme, asking the central question: "What can be done to help colleges and universities supplant the compliance culture that has dampened the productive use of assessment results?" (pp. 220-221). The authors reflect briefly on the last century of higher education, and what the next several years may hold in academia. As with many other works across the higher education canon, the authors give great credence to technology and technology-enhanced platforms and strategies, as well as learning analytics. Technology-based educational alternatives and economic realities also threaten the higher education landscape. The appropriate response, the authors contend, is to create an assessment environment that can clearly demonstrate the impact of higher education on student learning and enhance institutional effectiveness.

Kuh et al. (2015) provide assessment practitioners and academic administrators with a thorough overview of assessment. The overview of assessment evidence and programs in the opening chapters coupled the discussions of organizing for assessment and assessment leadership summarize effectively both current assessment practices and the challenges that assessment – and more broadly, higher education accountability – face. The book provides an accessible discussion of not only these challenges, but also practical advice on how these challenges might be addressed.

Throughout the book, the central tenant for effective assessment remains the need to shift from a compliance-based assessment mentality to one focused on continuous improvement and evidence that can be used to inform efforts to improve teaching and learning. The authors provide clear and convincing evidence from the research that assessment matters. Indeed, this is an important observation, though by the later chapters it grows a bit tedious. This work is also unique in that it strives to address both the *why* of assessment and the practical realities (and frustrations) that arise on campuses when assessment is discussed.

The book itself promises a major reframing of how to develop and implement strategies to assess student learning, and it is on this point that the text falls short. Although the authors offer tangible examples from universities and colleges across the country of effective assessment programs, there is often not sufficient exploration of how these programs were developed. The authors make a strong case, buffeted by these examples, that assessment is most effective when it is used to inform the forward progress of an institution, but they provide less clarity on how these assessment data and practices might be used to achieve these aims. It is likely, however, that detailed prescriptions would be difficult to provide at scale due to differing political views, appetite for innovation, and levels of assessment maturity across individual campuses.

Regardless of these criticisms, the authors clearly illustrate that the current compliance-based culture of assessment is not working, and that it is alienating faculty, administration, and external constituents. The effective use of assessment activities and findings is critical to move an institution forward, and the continuous improvement environment informed by assessment outlined by Kuh et al. (2015) is a worthy goal for many, if not all, institutions.

This text will resonate strongly with assessment administrators and provosts, many of whom have been making these arguments for years, and may provide useful guideposts and research to engage in these conversations. Faculty at large would also benefit from the discussion provided in this book and from an understanding of the broader necessary purposes of assessment. Higher education administration and assessment programs would also benefit from this work, both as an introduction to the current assessment climate and as a potential catalyst for further research.

References

Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Dwell, P. T., Hutchins, P., & Kinzie, J. (2015). *Using evidence of student learning to improve higher education*. San Francisco, CA: Jossey-Bass.