RESEARCH & PRACTICE IN ASSESSMENT

Special Issue: Tenth Anniversary Edition

VOLUME ELEVEN | ANNIVERSARY 2017 www.RPAjournal.com ISSN# 2161-4210





RPÅ RESEARCH & PRACTICE IN ASSESSMENT

Associate Editor

Lauren Germain

SUNY Upstate Medical University

Associate Editor

Megan Shaffer

Santa Clara University

Editorial Staff

Editor Katie Busby University of Mississippi

Associate Editor Ciji A. Heiser The University of North Carolina at Chapel Hill

Editoral Assistant

Sarah Andert **Tulane University**

Ex-Officio Members

Virginia Assessment Group President Lee Rakes Viginia Military Institute

Amee Adkins

Illinois State University

Robin D. Anderson

James Madison University

Angela Baldasare

University of Arizona

Brian Bourke

Murray State University

Chris Coleman

University of Alabama

Lindsey Jakiel Diulus

Virginia Assessment Group President-Elect Stephanie Foster George Mason University

Review Board

Brian French Washington State University

Matthew Fuller Sam Houston State University

Megan Moore Gardner University of Akron

Karen Gentemann George Mason University

> Mare E. Gillespie St. John's University

Molly Goldwasser Duke University

Chad Goteh Wahsington State University

> Michele J. Hansen IUPUI

Debra S. Harmening University of Toledo

anthony lising antonio Stanford University

Susan Bosworth College of William & Mary

Jennifer A. Lindholm University of California, Los Angeles

Robin D. Anderson

2006

Joshua Travis Brown

2010-2014

Ghazala Hashmi

J. Sargeant Reynolds

Community College

Editorial Board

Daryl G. Smith Claremont Graduate University

Linda Suskie Assessment & Accreditation Consultant

John T. Willse University of North Carolina at Greensboro

Past Editors

Keston H. Fulcher 2007-2010

Loraine Phillips University of Texas at Arlington

Suzanne L. Pieper Northern Arizona University

> William P. Skorupski University of Kansas

Pamela Steinke University of St. Francis

Matthew S. Swain **HumRRO**

Wendy G. Troxel Illinois State University

Catherine Wehlburg Texas Christian University

Craig S. Wells University of Massachusetts, Amherst

Thomas W. Zane Salt Lake Community College

Carrie L. Zelna North Carolina State University

Nunes Community College

Dorothy C. Doolittle Christopher Newport University

Seth Matthew Fishman Villanova University

Teresa Flateby Georgia Southern University



S. Jeanne Horst James Madison University

> Natasha Jankowski NILOA

San Diego State University

Kimberly A. Kline Buffalo State College

Sean A. McKitrick Middle States Commission

Deborah L. Moore North Carolina State University

> John V. Moore **Community** College of Philadelphia

Ingrid Novodvorsky University of Arizona

Kendra Jeffcoat

Community College

Kathryne Drezek McConnell Association of American **Colleges & Universities**

RESEARCH & PRACTICE IN ASSESSMENT TENTH ANNIVERSARY EDITION

It is with great appreciation that we recognize the former editors of RPA. Your vision and dedication resulted in a publication that positively impacted the higher education assessment community for the past decade and will continue to do so for many years to come.

Robin D. Anderson, Psy.D. James Madison University Harrisonburg, VA *Founding Editor, 2006-2007*

Keston H. Fulcher, Ph.D. James Madison University Harrisonburg, VA *Editor, 2007-2010*

Joshua Travis Brown, Ph.D. University of Virginia Charlottesville, VA *Editor, 2010-2014*

Special thanks also goes to all current and former RPA Board members, reviewers, and editorial staff members. Your contributions over the past decade have shaped RPA and advanced the scholarship of assessment. Thank you to the members and the Board of the Virginia Assessment Group. Your support of RPA has made a lasting impression on the practice and scholarship of assessment in the Commonwealth of Virginia, within the United States, and across the world.

TABLE OF CONTENTS

FROM THE EDITOR

Cause for Celebration and a Call to Action - Katie Busby

5 <u>ARTICLES</u>

4

Tracking the Evolution of Research and Practice in Assessment through the Pages of RPA

- Robin D. Anderson & Nicholas A. Curtis

10 Return of the Pig: Standards for Learning Improvement

- Keston H. Fulcher, Kristen L. Smith, Elizabeth R. H. Sanchez, Allison J. Ames & Cara Meixner

41 The Seven Silos of Accountability in Higher Education: Systematizing Multiple Logics and Fields

- Joshua Travis Brown

CALL FOR PAPERS

Research & Practice in Assessment is currently soliciting articles and reviews for its Winter 2017 issue. 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, but submissions received by August 1 will receive consideration for the winter issue. Manuscripts must comply with the RPA Submission Guidelines and be sent electronically to: editor@rpajournal.com

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 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.

SAVE THE DATE

The Virginia Assessment Group will hold its 30th annual meeting at the Crowne Plaza Hotel in downtown Richmond, VA November 15-17, 2017.

FROM THE EDITOR

A Cause for Celebration and a Call to Action

Research & Practice in Assessment has cause for celebration - the journal commemorated its tenth anniversary! The board of the Virginia Assessment Group, under the leadership of RPA's founding editor, identified the need for a journal dedicated to publishing the work of assessment scholars and practitioners, and in Fall 2006 the inaugural issue of RPA was published online. The journal grew from a newsletter of the Virginia Assessment Group to an online journal appealing to an international audience and the scholarship featured in RPA evolved as did higher education assessment.

Milestone occasions such as this provide an opportunity to reflect on the contributions made by RPA authors. I encourage you to take this opportunity to reacquaint yourself with the articles that have been published and consider what they mean to your work.

To commemorate the tenth year of RPA, I am proud to introduce this special anniversary issue of *Research & Practice in Assessment*. I invited the three former editors of RPA to submit articles, contributing their thoughts and observations on the scholarship and practice of assessment with as few restrictions as possible. As one of these authors stated, I gave an inch and they took a mile! These special contributors have been active assessment researchers, scholars, and practitioners for over a decade and each left an indelible mark on RPA through their leadership and the contributions featured in this volume.

In this special anniversary issue, Anderson and Curtis provide a retrospective of the assessment literature from early psychometric studies to the more recent focus on improving student learning. The authors also include a call to involve students in student learning assessment practices. Fulcher, Smith, Sanchez, Ames, and Meixner advance the process of improving student learning through the "weigh pig, feed pig, weigh pig" model. To demonstrate this work they provide a practical example for improving student learning. Brown examines assessment and its relationship to seven accountability fields through his conceptual model. He challenges readers to close the gaps between the silos through innovative practice and research.

Not only does the tenth anniversary of RPA give us a chance to celebrate and reflect, it also serves as a reminder that our work continues. I hope RPA has informed your assessment efforts, challenged your perspectives, and inspired you as an assessment professional over the years. *Research & Practice in Assessment* will continue to encourage and publish the scholarship of assessment professionals in the decade ahead and I encourage you to imagine and shape what the assessment scholarship and practice will be.

Regards,

Katie Busby

University of Mississippi



AUTHORS

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

Nicolas A.Curtis, Ed.S. James Madison University

Tracking the Evolution of Research & Practice in Assessment Through the Pages of RPA

Ten years ago, Research & Practice in Assessment (RPA) was born, providing an outlet for assessment-related research. Since that first winter issue, assessment research and practice has evolved. Like with many evolutions, the assessment practice evolution is best described as a change of emphasis as opposed to a radical revolution. Previous areas of exploration and challenge have given way to more complex areas of study as the technical skills of assessment coordinators have advanced and stakeholders have become more willing to incorporate assessment into their practices. However, the promise of learning improvement has been slow to follow. In addition, stakeholders are asking new questions regarding the impact of a college education across one's lifespan.

In the early years, researchers and practitioners at several institutions across the Commonwealth of Virginia were engaged in research to advance assessment practice. Some of this research involved investigating the impact of assessment strategies on the validity of assessment findings. For example, conducting research on issues such as how proctors, instructions, and testing conditions affected student motivation and efforts on institutions' student learning assessments became a regular part of the assessment practitioner's role on many college campuses. Those practitioners with measurement training were also conducting extensive studies on the psychometric properties of testing instruments and on the impact of rater characteristics on the reliability of ratings for performance assessments. Researchers and students were routinely conducting these scholarly studies, master's theses, and doctoral dissertations related to assessment practice. However, the authors were often challenged to find a journal that would publish such assessment studies, even if these studies were steeped in traditional research methodology. Research practitioners could present at the Virginia Assessment Group conference each fall and some also published articles in assessmentfocused publications such as Assessment Update. However, the Virginia Assessment Group did not publish conference proceedings, and Assessment Update, while providing an open and professional venue for brief assessment pieces, did not feature full-length journal articles.

CORRESPONDENCE

Email ander2rd@jmu.edu

Volume Eleven | Anniversary Issue 2017



In 2005, the forward thinking Virginia Assessment Group members addressed this scholarly gap. They converted the online newsletter to an online peer-reviewed journal that would disseminate evolving research coming out of institutions across not only Virginia but also the Mid-Atlantic and beyond. We knew that assessment practice was accelerating and here, ten years later, we see this is particularly true in the areas of measurement and "meta-assessment."

Like with many evolutions, the assessment practice evolution is best described as a change of emphasis as opposed to a radical revolution. In the winter of 2006, when RPA first hit the web, many stakeholders were focused on the quality of our institutions' assessment practices. An unwillingness to make programmatic changes based on assessment findings was often couched in concerns regarding the quality of assessment instruments and the soundness of assessment procedures. Practitioners, under increasing pressure from unfunded mandates, were having to build or gear up their assessment initiatives at a mindblowing pace, requiring a heavy reliance on the adoption of pre-existing instruments or the hurried production of local assessments. In essence, we were too often building the plane while also trying to fly it. Practitioners understood that if stakeholders were to invest in the use of assessment findings to improve student learning, stakeholders would need to know that they could make valid inferences from the assessment results. For that to happen, practitioners would have to answer tough questions regarding the reliability of our measures and the validity of our inferences.

Accordingly, a significant portion of the assessment-related scholarship appearing in the early issues of RPA centered on issues of reliability and validity (Ridley & Smith, 2006; Pieper, Fulcher, Sundre, & Erwin, 2008). Assessment researchers needed to fill the gaps within the psychometric literature. These research practitioners wanted to ensure we could document appropriate forms of reliability including test-retest reliability, parallel forms reliability, internal consistency, and interrater reliability. We worked to understand differential item functioning (Thelk, 2008) and its impact on the interpretation of assessment findings. While at least limited reliability evidence was often available for measures of student learning, validity evidence was somewhat nonexistent. Over time, additional research was conducted to document validity evidence. Many of these studies took the form of expected group differences studies, concurrent validity studies, and studies examining the factor structure of new and existing assessment instruments (Smiley & Anderson, 2011). Others were working to encourage practitioners and stakeholders alike to think differently about how to interpret findings using effect sizes (Springer, 2006). Today, much of this work would be considered preliminary psychometric work. For example, we regularly engage in equivalence studies before we compare groups based on the findings from a given instrument. We do this to ensure that the scores do not simply demonstrate reliability and that our inferences are supported but also that these things hold true in a similar way for each of the groups involved.

As further evidence of the advancement in practice, one needs simply to look at how the training offerings for assessment practitioners have changed. When RPA first launched, trainings and workshops in structural equation modeling (SEM) and item response theory (IRT) were considered "advanced." Today, there are graduate courses and professional workshops in *advanced* SEM, *advanced* IRT, hierarchical linear modeling (HLM), and Bayesian analyses. We have evolved past the days where basic psychometric work is what we do; instead it is a given, a necessity, a starting point from which we now build our research agendas. Finding the "right methodology" is still at the center of what we do as assessment practicioners; however, we now do so much more.

As we have moved beyond basic psychometric work, so too have we moved beyond the days of convincing programs that they must engage in student learning outcomes assessment. In 2006 when RPA was launched, many assessment practitioners were still working to convince academic programs on campus to even minimally engage in program assessment. As Marilee Bresciani (2011) stated "there are many others who do not believe the process [outcomes-based assessment] adds value to their day-to-day teaching or administrative duties" (p. 5). Thus it is not surprising that much of the scholarship conducted by assessment practitioners was aimed at convincing faculty that assessment was worth doing and could benefit faculty members' efforts to improve student learning. Even those in areas with a long history of program evaluation found themselves in the position of arguing for a shift toward a greater focus on student learning outcomes. In his piece focused on assessment in academic libraries, Eric Ackermann (2007), pointed out that despite academic libraries having a nearly 100 year history of conducting assessment, they were still using "metrics and reporting protocols that are meaningful primarily to other librarians" (p.23). Ackermann called for new processes that would examine the relationships between academic libraries and student learning, stating "such information could only help the efficacy of libraries across the country" (p. 23). Assessment experts from academic and students affairs alike were producing such argumentive manuscripts in an effort to make the case for assessment.

A decade later, at many institutions, it is a given that programs are engaged in the ongoing assessment of student learning outcomes. Today many of the efforts of assessment offices focus on supporting these program initiatives and assessing the assessment. An increasing number of institutions have systematized a process to gather, review, and evaluate the assessment processes being used in the programs. Many of these assessment-of-the-assessment processes, or meta-assessments, include a step by which reviewers or raters provide feedback to the programs for the purpose of improving programs' assessment practices and hopefully also student learning outcomes.

While we have made huge strides in addressing many of the major challenges that face assessment researchers and practitioners at the time RPA was launched, we remain challenged to help assessment research and practice to evolve to the next level. While certainly no one can predict the future, we are seeing evidence that points to some of the major areas of focus in assessment practice for the coming years. These foci include learning improvement, lifelong impact of college, and the engagement of new groups in the assessment process. One of the criticisms we still hear regarding assessment is that institutions struggle to identify specific examples of where student learning has improved as a result of lessons learned through assessment. However, this is not overly surprising. As many freshmen psychology students would be able to tell you, basic needs need to be met before higher order needs can be addressed (thank you, Maslow). We argue that before widespread learning improvement might occur, we needed to address the basics. We needed to strengthen our assessment processes to ensure that those who are in the position to use our findings have confidence in the results. This means all those studies on the quality of our instruments (Barnes & Burchard, 2011) and on the impact of our systematic assessment procedures (Barry & Finney, 2009) on the validity of our inferences needed to become common place. Now, with a long record of disseminated research and research-based assessment practice becoming common, we are standing at yet another launching point. We believe that this next great advance will be the widespread institutionalization of learning improvement initiatives based on student learning outcomes assessment findings.

Keston Fulcher talks extensively regarding the learning improvement initiatives taking place at James Madison University (JMU) (Fulcher, Smith, Sanchez, Ames, & Meixner, 2017). These initiatives are grounded in a partnership with the University's Center for Faculty Innovation, the teaching and learning center on campus. Such partnerships are proving to be the cornerstone of many learning improvement initiatives. In addition to partnerships like the one at JMU between the Center for Assessment and Research Studies and the Center for Faculty Innovation, we are observing an increase in partnerships between assessment practitioners and those faculty responsible for assessment within the disciplines. Much of the authors' own research is done in partnership with longtime colleagues in our engineering department. It is not unusual for assessment practitioners, with their skills in measurement and evaluation, to serve as evaluators on grants. However, within our partnerships we often serve, not as evaluators, but as full co-principal investigators on major initiatives. We do not merely have a seat at the table or serve to represent the university as the "assessment person" on the committee, but we serve as full partners contributing at every stage of the project. The evolution of these partnerships between discipline faculty and assessment practitioners helps to merge the roles of content experts and measurement experts increasing the likelihood that findings will lead to improvements in student learning and an increase in interprofessional practice.

We believe that this next great advance will be the widespread institutionalization of learning improvement initiatives based on student learning outcomes assessment findings. A review of more recent editions of RPA reveals an increasing number of improvement focused manuscripts. Scholarly pieces calling for a greater focus on learning improvements appeared in RPA as early as 2008. In the 2008 winter edition, Pieper et al. discuss four analytic strategies for addressing educational research questions that could help insistutions "reap the maximum benefits from their assessment efforts" (p.4). However, starting in 2013, articles focusing on using data for improvement began to appear with increasing frequency in RPA (e.g., Gerstner & Finney, 2013). By winter 2015, at least three of the articles appearing in RPA focused on the use of assessment results for improvement purposes. As reflected in the pages of RPA, we see using assessment to inform changes that lead to learning improvement as an idea whose time has come.

In our opinion, research and assessment practitioners should be pleased with our progress and excited by our future.

A second area emerging in our field is an interest in assessing the lifelong impact of the college experience. Efforts by organizations such as Gallup to gather data on lifelong satisfaction and the role college plays in how we view ourselves and the quality of our lives have expanded our view beyond the initial experience of college. Now the horizon is not the senior year, or even graduation, but life five, ten, or forty years beyond the graduation ceremony. With partners such as Gallup, institutions hope to be able to show that the truly meaningful outcomes of college expand well beyond the senior (or fifth) year. One such partnership between Gallup, the State Council of Higher Education for Virginia (SCHEV), and the Virginia Assessment Group resulted in a Summer 2015 Summit on Quality & Value in Virginia's System of Higher Education. As a part of the summit, representatives from the University of Virginia, Virginia Tech, and George Mason University spoke about the importance of preparing students for both work and life. Ray and Kafka (2014) summarize the inaugural findings of the Gallup-Purdue Index stating "when it comes to being engaged at work and experiencing high well-being after graduation...the type of institution they [students] attended matters less than what they experienced there" (para. 1). Such findings open up an entirely new area of assessment practice and research, redefining alumni assessment. The Gallup-Purdue Index, and the importance the initial results place on college experiences, ties to our final thought on emerging assessment trends, student engagement.

As we discussed, in the "dark ages" (i.e. 10 years ago), many faculty were resistant to program assessment. While many things contributed to our current, more assessmentfriendly, climate, one of the biggest contributors was the push to engage faculty at the beginning of the assessment process. Not only did this change the climate, but it also improved the assessment process itself. Now, we believe that a similar engagement evolution is on the horizon for a different audience, the students that we assess. If successful, this shift could again radically alter the value of assessment in higher education. If our goal is student learning improvement, why would we not partner with students at all levels of the assessment process? If our goal is student learning improvement, why would we try to interpret results and make decisions without students' input. If our goal is student learning improvement, why should we not leverage students' obvious stake in the outcomes of our assessments? This is not a radical idea. Charlie Blaich and Kathy Wise, of the Center for Inquiry at Wabash College, have been working with undergraduate students to train them to conduct peer-led groups focused on assessment results. These students have been able to help get at the "why and how" of the assessment data. The students are often able to help frame assessment results to make them more useful and compelling. We are optimistic and excited about the impact partnering with students may make on assessment practice. Research on such topics are just beginning to appear in publication. We believe that as RPA covers the next 10 years of assessment practice these will be among the topics that appear in its pages.

In our opinion, research and assessment practitioners should be pleased with our progress and excited by our future. For those of us who have been at this a while, it is important to occasionally remind ourselves just how far we have come. Given that the ultimate goal of assessment is to improve student learning, we believe we are just now positioned to see the fruits of our labors. Yet, those just coming into the field can rest assured that there are still discoveries to be made and challenges to be overcome.

References

- Ackermann, E. (2007). Program assessment in academic libraries: An introduction for assessment practitioners. *Research & Practice in Assessment, 2*, 18-23.
- Barnes, S.K. & Burchard, M.S. (2011). Quality and utility of the Multi-tiered Instruction Self-Efficacy Scale. *Research & Practice in Assessment*, 6(2), 22-42.
- Barry, C.L. & Finney, S.J. (2009). Does it matter how data is collected? A comparison of testing conditions and the implications of validity. *Research & Practice in Assessment*, *4*, 17-26.
- Bresciani, M.J. (2011). Identifying barriers in implmenting outcomes-based assessments program review: A grounded theory analysis. *Research & Practice in Assessment*, 6(1), 5-16.
- Fulcher, K. H., Smith, K. L., Sanchez, E. R. H., Ames, A. J., & Meixner, C. (2017). Return of the pig: Standards for learning improvement. Research & Practice in Assessment, 11(2), 10-27.
- Gerstner, J.J. & Finney, S.J. (2013). Measuring the implementation fidelity of student affairs programs: A critical component of the outcomes assessment cycle. *Research & Practice in Assessment*, 8(2),15-28.
- Pieper, S.L., Fulcher, K.H., Sundre, D.L, & Erwin, T.D. (2008). "What do I do with the data now?": Analyzing a comparison of testing conditions and the implications for validity. *Research & Practice in Assessment*, *3*, 4-10.
- Ray, J. & Kafka, S. (2014). Life in college matters for life after college. Retrieved from: http://www.gallup.com/poll/168848/life-college-matters-life-college.aspx
- Ridley, D.R. & Smith, E.D. (2006). Writing across the curriculum works: The impact of writing emphasis upon senior exit writing samples. *Research & Practice in Assessment*, *1*, 14-17.
- Smiley, W. & Anderson, R. (2011). Measuring students' cognitive engagement on assessment tests: A confirmatory factor analysis of the short form of the cognitive engagement scale. *Research & Practice in Assessment*, 6(1), 17-28.

Springer, R. (2006). Using effect size to assist in NSSE reporting. Research & Practice in Assessment, 1, 18-22.

Thelk, A. (2008). Detecting items that function differently for two- and four- year college students. *Research & Practice in Assessment*, *3*, 23-27.

RESEARCH & PRACTICE IN ASSESSMENT ••••••



AUTHORS

Keston H. Fulcher, Ph.D.

Kristen L. Smith, M.A.

Elizabeth R.H. Sanchez, M.A.

Allison J. Ames, Ph.D.

Cara Meixner, Ph.D. James Madison University

Abstract

Higher education has made impressive progress concerning student learning outcomes assessment practices. Yet – despite the assumption that better assessment would lead to better student learning - few examples of demonstrable student learning improvement exist at the academic degree or university levels. In 2014 Fulcher, Good, Coleman, and Smith addressed this concern in a NILOA Occasional Paper titled, "A Simple Model for Learning Improvement: Weigh Pig, Feed Pig, Weigh Pig". The "pig paper" elucidated basic steps for improvement: collect baseline data, intervene effectively with curriculum and pedagogy, and then re-assess to examine if learning did indeed improve. The current paper builds upon these steps by offering standards for learning improvement. We articulate these standards via a rubric and a hypothetical learning improvement report. These tools are intended to elevate learning improvement conversations, and increase the number of learning improvement examples across higher education.

Return of the Pig: Standards for Learning Improvement

have long been frustrated with hollow statements about assessment's ability to improve higher education. While I am not as cynical as Erik Gilbert – who penned the 2015 *Chronicle* article, "Does Assessment Make Colleges Better? Who knows?" – I get his point. Colleges across the world spend substantial amounts of time and money on assessment, but scant evidence exists to justify the resources (Suskie, 2010). The good news is I believe assessment's state of affairs can be changed. Academe can do better. Collaborators Kristen Smith, Elizabeth Sanchez, Allison Ames, and Cara Meixner join me in an important step toward shifting higher education's focus away from empty assessment practice to something more edifying. We propose a fundamental pair of resources: a) a rubric detailing standards for learning improvement, and b) a learning improvement report from a hypothetical program annotated according to the rubric.

In 2014, Fulcher and Smith contributed to a National Institute for Learning Outcomes Assessment (NILOA) Occasional Paper titled "A Simple Model for Learning Improvement: Weigh Pig, Feed Pig, Weigh Pig" (Fulcher, Good, Coleman, & Smith, 2014). The pig analogy is an extension of the old farm saying: a pig never fattened because it was weighed. Our consensus is that the same logic applies to higher education: merely assessing, repeatedly "weighing" students, will not improve their learning. The Weigh Pig, Feed Pig, Weigh Pig paper elucidated the relationship between assessment and learning, each step essential to a simple model of improvement:

Email fulchekh@jmu.edu

CORRESPONDENCE

- collection of baseline data,
- integration of faculty training and development,
- use of evidence-based teaching strategies,
- effective modifications to the learning environment, and
- re-assessment to affirm efficacy and evidence improvement.

This contribution to RPA expands the simple model by providing more explicit learning improvement standards. While we are under no presumption that higher education will use the exact standards we provide or the very rubric we've created, eventually, universal standards should be adopted; a universal rubric should be used.

We've organized this paper as six parts. First, we provide context for how standards, including rubrics, have articulated best practices in assessment. Second, we examine the current and future state of affairs regarding learning improvement. Third and fourth, we provide learning improvement standards via a learning improvement rubric that more carefully articulates and elaborates the standards in behavioral terms. Fifth, we provide an annotated hypothetical example of a learning improvement report. Finally, we provide suggestions for how academe could use these learning improvement resources for maximum effect.

Before Improvement: Standards and Meta-Assessment Rubrics Advance Assessment Best Practice

Many books provide fine-grained details of an assessment cycle (e.g., Bresciani, Gardner, & Hickmott, 2009; Erwin, 1991; Palomba & Banta, 1999; Suskie, 2009; Walvoord, 2010). But on the conceptual end, the most noteworthy guidance for best practices in assessment comes from the now defunct American Association for Higher Education (AAHE). Despite being over two decades old, Hutchings, Ewell, and Banta (2012) observed that AAHE's nine Principles of Good Practice for Assessing Student Learning have aged incredibly well. We agree. Indeed, many of the principles highlight an underlying improvement emphasis within assessment.

While we have included the first few lines of the nine principles, they are worth reading (or re-reading) in their entirety:

- 1. The assessment of student learning begins with educational values.
- 2. Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time.
- 3. Assessment works best when the programs it seeks to improve have clear, explicitly stated purposes. Assessment is a goal-oriented process.
- 4. Assessment requires attention to outcomes but also and equally to the experiences that lead to those outcomes.
- 5. Assessment works best when it is ongoing not episodic.
- 6. Assessment fosters wider improvement when representatives from across the educational community are involved.
- 7. Assessment makes a difference when it begins with issues of use and illuminates questions that people really care about.
- 8. Assessment is most likely to lead to improvement when it is part of a larger set of conditions that promote change.
- 9. Through assessment, educators meet responsibilities to students and to the public. (Hutchings et al., 2012, para. 6)

Note, we added post-hoc emphasis to principles 3, 4, 6, 7, and 8, which emphasize the key points of this article.

Since AAHE's nine principles were published, the emerging practice of meta-assessment has helped further clarify what constitutes good assessment (Fulcher, Coleman, & Sundre, 2016). Meta-assessment commonly involves using a rubric to evaluate the quality of a report by providing detailed feedback on assessment processes and characteristics such as student learning outcomes (SLOs), methodology, results, and use of results. Each characteristic (or rubric criterion) is described at various quality levels (e.g., beginning, developing, good, and excellent) in the rubric (Popham, 1997).

We propose a fundamental pair of resources: a) a rubric detailing standards for learning improvement, and b) a learning improvement report from a hypothetical program annotated according to the rubric. Prominent organizations such as the New Leadership Alliance (NLA), Voluntary System of Accountability (VSA), NILOA, the Association of American Colleges & Universities (AAC&U), American Association of State Colleges and Universities (AASCU), and the Association of Public & Land-Grant Universities (APLU) have designed rubrics to evaluate the quality of institution-level assessment. While the NLA initiative was short-lived and the VSA, NILOA, AAC&U, AASCU, and APLU initiative is in its infancy, they represent the first national-level efforts to use a rubric to articulate assessment quality and standards of practice. These institution-level rubrics heavily emphasize issues such as communication of results and administrative and structural supports for assessment.

Unfortunately, most meta-assessment at the institutional or organizational level inadequately gauges student learning improvement. In other words, improvement is more than just a byproduct of good assessment processes. Quality assessment results, in our experience, rarely prompt change that improves learning. We agree with those who believe that assessment should be subsumed within a larger learning improvement system (Lumina Foundation, 2016).

Current State of Affairs for Improvement: More Assessment ≠ More Learning Improvement

Again, advances in assessment practice differ from improvements in student learning. While books, standards, and rubrics have refined assessment methodologies, today's practice still fails to capture the learning improvement spirit underlying AAHE's 20-year-old principles.

It comes as no surprise that we have witnessed and demonstrated few improvements in student learning at the academic program level of our institution. Only one of 14 criterion in our own award-winning (Willard, 2015) meta-assessment rubric mentions using assessment results for learning improvement. Unfortunately, faculty at our university may not be aware of what a successful improvement initiative may look like or the specific steps they can take to evidence learning — this information is not provided in the meta-assessment rubric we use.

Over the past few years, a handful of influential scholars have voiced the noted lack of evidenced learning improvement in higher education contexts (Banta, Jones, & Black, 2009; Blaich & Wise, 2011). Banta, Jones, and Black (2009) found that only six percent of the best assessment reports across the nation demonstrated student learning improvement. When Banta and Blaich (2011) were asked by *Change Magazine* to provide current examples of learning improvement, they could not find a sufficient number to write the article. The authors proceeded instead to write an article about obstacles to learning improvement (Banta & Blaich, 2011).

To some readers, the current state of affairs might seem surprising given higher education's intentions of assessment. Indeed, in a survey conducted by NILOA in 2013, provosts reported that "commitment to institutional improvement" was ranked third of the 13 most important reasons to conduct assessment, falling just behind regional accreditation and program accreditation. In a more pointed question, provosts were asked how assessment results were used; "curriculum modification" and "institutional improvement" were reported as uses "quite a bit." Further, curriculum modification and institutional improvement were endorsed as a use of assessment results fifth and seventh, respectively, out of 16 total possible uses (e.g., accreditation, program review, institutional benchmarking, etc.). On average, provosts reported that they used assessment results for curriculum modification "quite a bit" (Kuh, Jankowski, Ikenberry, & Kinzie, 2014).

One reason for the apparent discrepancy between NILOA's survey results and the national lack of demonstrable gains in student learning is the inconsistent and vague definition of *improvement*. Many rubrics (including our institution's meta-assessment rubric), assessment-related books, and assessment measures use the term improvement in an imprecise way, as a synonym for change or perhaps as any use of results (Smith, Good, Sanchez, & Fulcher, 2015).

Quality assessment results, in our experience, rarely prompt change that improves learning.



During the creation of our new learning improvement rubric, we followed the Weigh Pig, Feed Pig, Weigh Pig article's definition of evidencing learning improvement: "making a change to a program and then re-assessing to determine that the change positively influenced student learning" (Fulcher, Good, Coleman, & Smith, 2014, p. 4). Using this definition, we believe that program and/or institutional learning improvement occurs much less frequently than "quite a bit."

Pivoting Higher Education toward Learning Improvement

At the national level, organizations such as the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC) have taken a stance on improving student learning. According to SACSCOC, "the concept of quality enhancement is at the heart of the Commission's philosophy of accreditation" (SACSCOC, 2016, para. 3). More specifically, SACSCOC requires every institution seeking reaffirmation of accreditation to engage in a campus-wide initiative to enhance student learning (i.e., a Quality Enhancement Plan or QEP). Each QEP includes processes for identifying issues or needs that emerge from institutional assessment efforts. Then, the institution must create a plan to enhance student learning and/ or the environments supporting student learning, which includes determining specific goals and assessment strategies.

Another national organization, NILOA, offers advice regarding building strategies to intervene through assignment design (Hutchings, Jankowski, & Ewell, 2014). Other organizations, such as the State Higher Education Executive Officers Association (SHEEO) and AAC&U, developed the Multi-State Collaborative, in which colleges use rubrics to assess various learning areas and encourage processes that change curriculum and pedagogy in intentional ways (http://www.sheeo.org/projects/msc-multi-state-collaborativeadvance-learning-outcomes-assessment). Charles Blaich and Kathy Wise, of the Center of Inquiry (http://www.liberalarts.wabash.edu/), consistently provide great contributions through forward-thought and institutional support and their efforts exemplify how faculty, administrators, and students can work together to use assessment data to influence improvement.

At the state level, institutions are preparing to incorporate models of learning improvement. In Virginia, the Virginia Assessment Group (VAG) and the State Council of Higher Education for Virginia (SCHEV) are encouraging such initiatives. Virginia Commonwealth University and Longwood University are beginning the piloting process. Meanwhile, in Georgia, Kennesaw State University has recently required improvement reports in addition to traditional assessment reports; programs must now re-assess to determine the efficacy of their changes. At James Madison University, we have piloted several learning improvement initiatives under the guide of the simple model (Fulcher et al., 2014) but using a more complex definition:

Strong evidence, from direct measures, supporting substantive learning improvement due to program modifications. This program responded to previous assessment results, made curricular and/or pedagogical modifications, RE-assessed, and found that student learning improved. The rationale and explanation of the modifications leading to the change are clearly laid out. The methodology is of sufficient strength that most reasonable alternative hypotheses can be ruled out (e.g., sampling concerns, validity issues with instrument or student motivation). In essence, the improvement interpretation can withstand reasonable critique from faculty, curriculum experts, assessment experts, and external stakeholders (Fulcher, Sundre, Russell, Good, & Smith, 2015, p. 3).

One academic program, Computer Information Systems, has already demonstrated large gains at the program level. A campus-wide project, The Madison Collaborative, has shown university-wide improvements in students' ethical reasoning skills. These success stories are attributed to partnerships between faculty leaders, assessment experts, faculty developers, and administrators who collectively worked to implement the simple model.

Although isolated success stories can be identified at both the national and institutional levels, our aim is to further cultivate learning improvement examples and more generally to elevate the learning improvement conversation. Although isolated success stories can be identified at both the national and institutional levels, our aim is to further cultivate learning improvement examples and more generally to elevate the learning improvement conversation. To do so, we provide learning improvement standards and the rationale for each in the spirit of the well-conceptualized AAHE principles. These standards for learning improvement are embodied and elaborated via a learning improvement rubric. This rubric can be used to guide and evaluate learning improvement initiatives. To show how the learning improvement rubric can be applied, we provide an annotated example of a report. We conclude with suggestions for how practitioners can use these two resources (i.e., the learning improvement rubric and learning improvement report) such that examples of demonstrable learning improvement become the norm for higher education, not the exception.

The Learning Improvement Rubric: Six Standards of Successful Learning Improvement

Our learning improvement rubric was crafted by Fulcher, Smith, and Sanchez throughout a semester-long independent course; vetted by assessment practitioners and faculty development experts; and is supported by extensive research and a few pilot initiatives. In our experience, all six standards detailed in the rubric are common in and necessary for successfully demonstrating program-level learning improvement.

Some terminology used in the rubric and example report will be new to most readers. Borrowing from our colleagues in computer information systems at JMU and faculty development, we have adopted the terms *as is* [*curriculum* or *environment*] and *to be* [*curriculum* or *environment*]. In this context, the *as is* curriculum or environment is what a program is or is not doing to meet the selected student learning outcome (SLO) before starting a learning improvement initiative; the *to be* curriculum or environment is the culmination of all proposed changes expected to improve student learning.

A. Faculty Involvement: Faculty are participating throughout the learning improvement initiative.

Faculty contribution, buy-in, and engagement are crucial to the success of program-level improvement processes for several reasons (Shavelson, 2010). First, faculty are responsible for the success of their individual classes, which are part of a broader curriculum and academic program. Similarly, faculty serve as a vital bridge between students and knowledge/skill acquisition (Kuh, Ikenberry, Jankowski, Cain, Ewell, Hutchings, & Kinzie, 2015). They are the frontline facilitators and shapers of students' classroom learning experiences.

Ideally, improved learning at the program level means that all graduating students are better educated. Therefore, changes in individual courses must be connected and aligned by multiple faculty members. Such alignment necessitates buy-in and engagement from a dedicated cadre. For clarification, we don't mean that all faculty in a program have to work directly on the learning improvement project for it to be effective. However, the majority of those whose classes cover the selected SLO should be active, dedicated participants. When faculty involvement is maximized and championed, faculty can more effectively:

- create powerful teaching and learning strategies;
- determine how, when, and where changes to the program should be implemented;
- · deliver new curricula to all students as intended to be delivered; and
- appropriately connect new learning experiences across different courses within the program in a way that facilitates learning.

Consider students in a hypothetical communications program at XYZ University who struggle to effectively deliver oral presentations with an engaging introduction, logical and fluid body, and smooth conclusion that reinforces the main ideas of the presentation (i.e., the selected SLO).

Ideally, improved learning at the program level means that all graduating students are better educated. Therefore, changes in individual courses must be connected and aligned by multiple faculty members. To remedy this learning deficit, faculty from the communications program could create an improvement initiative in hopes of bolstering their students' abilities to effectively deliver a presentation. If only two or three faculty in the program of 20 decide to get involved, planning program-level changes would be nearly impossible with so many faculty non-contributors. The chances are good that the oral presentation student learning outcome (SLO) is or would be covered in courses that are taught by faculty not participating in the improvement initiative. When the communications faculty conducts program-level assessments, any learning improvements of the few students who experienced a modified course would likely be washed out by the remaining students who were taught by faculty members who *did not* change their courses in efforts to improve oral presentation. Ultimately, program-level learning improvement cannot be achieved without a "program-level's worth" of faculty participation. Figure 1 depicts the faculty involvement criterion of the learning improvement rubric.

Any learning improvement initiative is a substantial undertaking. For this reason, we highly recommend focusing on one student learning outcome (SLO) at a time.

A. Faculty Involvement: Faculty are participating throughout the learning improvement initiative.					
	0 Absent	1 Beginning	2 Developing	3 Good	4 Exemplary
1. Relevant Faculty Involvement	No evidence of faculty involvement <i>or</i> degree of faculty involvement is unclear.	Evidence of ~50% of faculty involvement through <i>most</i> aspects of the learning improvement initiative.*	Evidence of 51- 74% of faculty involvement through <i>most</i> aspects of the learning improvement initiative.*	Evidence of 75- 89% of faculty involvement through <i>every</i> aspect of the learning improvement initiative.*	Clear evidence (e.g., faculty surveys, progress checks, etc.) of 90-100% of faculty involvement through <i>every</i> aspect of the learning improvement initiative.*
*Aspects of the learning improvement initiative include: student learning outcome (SLO) selection, elaboration, and assessment alignment/match; baseline data collection; investigation of original program; program intervention; and re-assessment.					

Figure 1. Faculty involvement criterion of the learning improvement rubric.

B. Readiness: The program can provide an explanation of why the SLO was selected for improvement, what the outcome means, and how learning and improvement is and will be measured.

Any learning improvement initiative is a substantial undertaking. For this reason, we highly recommend focusing on one student learning outcome (SLO) at a time. Attempting to demonstrably improve more than one SLO in a given year may quickly become overwhelming and discouraging to faculty who are investing their time and expertise to the learning improvement initiative. Selecting an SLO that students are not achieving satisfactorily makes learning improvement manageable and will hopefully focus faculty contributions.

Before making any pedagogical or curricular changes, it is imperative that faculty elaborate the selected (*or targeted*) SLO; doing so will help faculty:

- gain a common, specific understanding of what needs improvement;
- tightly connect and align new pedagogy and curricula with the targeted SLO and assessment instrumentation; and
- provide a framework by which to evaluate the *as is* learning environment and the *to be* learning environment.

Without a clearly detailed or elaborated SLO, faculty will have only a very loose idea of the specific skills, knowledge, or abilities that they want to measure and improve. It may be easy for faculty to want to skip SLO elaboration in favor of immediately implementing a novel curricular change or modifying existing pedagogies. However, a nebulous or vague SLO is detrimental to the success of a student learning initiative because faculty might have different conceptualizations of what the SLO means, how it is best taught, and how it is appropriately measured. Figure 2 depicts the readiness criterion of the learning improvement rubric.



RESEARCH & PRACTICE IN ASSESSMENT ••••••

b. Readiness: 11	ins, and <u>how lear</u>	provide an explana	nent is and will be	e measured.	mprovement, what
	0 Absent	1 Beginning	2 Developing	3 Good	4 Exemplary
1. <i>Program-level</i> SLO Selection: Rationale with Evidence	No rationale or evidence is provided for why the SLO was selected for improvement.	Rationale provided with <i>anecdotal</i> evidence <i>or</i> evidence doesn't clearly support the need for SLO improvement initiatives.	Rationale provided with evidence from one source* and clearly supports the need for SLO improvement initiatives.	Rationale provided with evidence from <i>two</i> sources* and clearly supports the need for SLO improvement initiatives.	Meets criteria for Good and cites literature suggesting a need to achieve the selected SLO that extends beyond the university. SLO must be a higher- order thinking skil (e.g., critical thinking).
*Sources may	include: Direct	measures (e.g., pre	vious test scores);	a survey of current	students; alumni
surveys; progr	am reviews; acc	reditation reports;	insights from expe	erts in the field; or e	mployer feedback.
2. SLO Elaboration: Ascription of Meaning	The SLO is not elaborated (e.g., Students will perform critical thinking.) <i>or</i> is not stated in student terms (e.g., Program faculty).	The SLO is elaborated to a minor degree. (e.g., Students will perform critical thinking, which includes: analysis & evaluation).	Ihe SLO is elaborated. Nevertheless, no explanation of the process is provided. (e.g., Students will perform critical thinking, which includes: analysis & evaluation. Analysis involvesand evaluation means).	The SLO is elaborated and an explanation of the process is provided. (e.g., faculty were polled to determine how "analysis" should be defined and then; the meaning of "evaluation" was ascribed through the process of).	Meets criteria for Good and the elaboration is supported by external experts (e.g., definitions were vetted in job analysis interviews; meaning was supported in an extensive literature review, etc.).
3. Alignment/ Match: Elaborated SLO and Assessment Measures	No relationship between SLO and assessment is provided.	Vaguely references SLO and assessment alignment/match but it is unclear to what degree. (e.g., Faculty wrote items to match the definition of "analysis;" this test was chosen because of its match to our selected "evaluation"	Evidence of SLO and assessment alignment/ match. (e.g., a rubric or set of test items are provided that match elaborated SLO).	Meets criteria for Developing and is affirmed via backward design (i.e., provides a sample of course assignments that were created based on test items) or behavioral anchoring (qualitative examples of performance standards for	Meets criteria for Good and is affirmed by external experts through similar processes.

Figure 2. Readiness criterion of the learning improvement rubric.

Consider, again, faculty members from a hypothetical communications program at XYZ University who engage in a learning improvement initiative to help their students effectively deliver an oral presentation (in addition to another four learning outcomes that have little to do with oral presentation delivery skills). Moreover, they did not detail the meaning of any of the SLOs that they targeted for their learning improvement initiative.

As communications faculty attempt to discuss current learning interventions in the as is curriculum, they may disagree and lack consensus. Indeed, the sheer amount of time it takes to discuss all of the SLOs becomes antithetical to advancing the learning improvement initiative. Furthermore, some faculty may think the SLOs have one meaning, while other faculty members conceptualize the SLOs in a completely different way. The discussions may eventually become circular because no one has defined or detailed what the SLOs really mean. All said, it may prove impossible to develop student knowledge, skills, and abilities when faculty have not elaborated the precise student knowledge, skills, and abilities they were trying to improve in the first place. They subsequently realize that the amount of time and resources

Like with many evolutions, the assessment practice evolution is best described as a change of emphasis as opposed to a radical revolution. it takes to define and elaborate one SLO makes improving multiple SLOs at once unfeasible.

C. Baseline data collection and measurement quality: Student performance is measured before program-level changes are made; high quality measurement is ensured.

When making claims related to improvement or growth in educational contexts, rigorous methodology is a necessity. Perhaps the most often forgotten or neglected part of evidencing learning improvement is that faculty must collect baseline data *before* any pedagogy, curriculum, or course sequencing changes are made in the program. Collecting baseline data allows for a more meaningful comparison with post-learning improvement initiative data.

Evidencing improvement requires data on students' knowledge, skills, or abilities before and after changes are made. Faculty should ensure that they are measuring the targeted SLO in a way that yields reliable and valid scores. Note, the need for baseline data and adequate instrumentation are widely accepted characteristics of best practices in student learning outcomes assessment contexts. Having strong methodology will help faculty:

- · demonstrate that students' assessment scores are reliable; and
- support claims of learning improvement (i.e., defend validity of improvement inference).

Without baseline data collection, programs and their partners in assessment and faculty development cannot empirically demonstrate that student learning has improved. Further, if the team does not use assessment tools that produce valid and reliable scores to measure the targeted SLO, they cannot capture requisite data on the specific and intended knowledge, skills, and abilities. Without using instruments that produce psychometrically sound data, the conclusions or inferences made from those assessment scores lack trustworthiness.

Imagine that faculty members from the hypothetical communications program at XYZ University started implementing new teaching strategies and targeted assignments before they collected data related to students' oral presentation abilities. Any data collected after the changes were made will not be at a true baseline – student learning may have already improved given the new course experiences. In other words, faculty will have a lack of data regarding how well their students were achieving the targeted SLO before they made changes to the curriculum; further, when collecting data after all new implementations are in place, the program cannot ascertain whether or to what degree students' oral presentation skills actually improved as a result of the learning improvement initiative.

Additionally, if the program measured students' oral presentation skills, but used a poorly designed instrument (e.g., an instrument that does not have desirable psychometric properties) there is no validity evidence to suggest that the assessment scores are meaningful. The inferences or conclusions of the initiative that communications program faculty attempt to make based on assessment scores will be severely compromised. Figure 3 depicts the baseline data collection and measurement section of the learning improvement rubric.

Certainly, universities with established assessment mechanics may have an easier time launching improvement initiatives. However, programs with fewer supports and resources should not feel discouraged. In fact, we recommend that institutions focus on one or two programs to pilot first. Having success at a small level will beget more successful learning improvement initiatives in the future.

D. Investigate curriculum and diagnose issues: An explanation of hypotheses for why what was originally being taught in the curriculum relative to the targeted SLO was ineffective.

A key component of demonstrable student learning is changing pedagogies and the curriculum—a learning *intervention* (Fulcher, Good, Coleman, & Smith, 2014). Before faculty can create meaningful, effective learning interventions, they must first investigate and

RESEARCH & PRACTICE IN ASSESSMENT •••••••

C. Baseline data changes are mad	C. Baseline data collection and measurement quality: Student performance is measured before program-level changes are made; high quality measurement is ensured.				
	0 Absent	1 Beginning	2 Developing	3 Good	4 Exemplary
1. Baseline Data Collection: Timing and Sampling	No baseline data collected prior to intervention <i>or</i> assessment measures were indirect (e.g., course grades).	Baseline data collected using direct measures. However, the sample is unrepresentative (i.e., only a few students were assessed, students had low motivation to perform well, little effort was given on assessment, etc.) or not enough information is provided to determine veracity of the data collection process (i.e., the number or level of motivation of assessed students is unknown).	Baseline data collected using direct measures; However, the sample is somewhat unrepresentative (i.e., only some students were assessed, student motivation/effort varied when being assessed) or the data provided are not a true baseline (i.e., some intervention was already implemented at the time of assessment). However, steps are taken to address these issues (e.g., motivation filtering: students who showed little effort were not included in data sample; evidence provided that shows data sample is representative although sampling technique wasn't the best, etc.).	Baseline data collected using direct measures; sample is fairly representative and includes details (e.g., number of students assessed, motivation analysis, etc.). However, minor issues still exist (e.g., only 80% of students assessed).	Baseline data collected using direct measures before any intervention; The sample is representative; details provide evidence of good motivation (high- stakes testing) and defensible testing conditions (e.g., student questions about the assessment are answered, etc.).
2. Psychometrics: Reliability and Validity	of reliability or the activities associated with reliability (i.e., using multiple raters for rubrics, training raters to use a rubric, etc.).	Evidence of activities associated with good reliability provided but reliability coefficients are absent or low <i>or</i> reliability analyses lack rigor (e.g., rater agreement is not exact).	Evidence of appropriate reliability that controls for chance provided. Marginal reliability rates are revealed.	Evidence of appropriate reliability that controls for chance provided. Passable reliability rates at the group level are revealed.	Evidence of appropriate reliability and validity (e.g., scores from assessment behave in predicted ways according to theories such as group studies or modeling).

Figure 3. Baseline data collection and measurement quality criterion of the learning improvement rubric.

understand the program as is – how and what information is being taught throughout the curriculum before any changes are made (Chaplot, Booth, & Johnstone, 2010).

Notions of "understanding about what is happening and what needs to happen to advance student success" and "defining problems and solutions" are emphasized in the Applied Inquiry Framework (Chaplot, Booth, & Johnstone, 2010, p. 10). Indeed, programs may find it difficult to make intentional, informed changes to the curriculum if faculty are unaware of why students are struggling to achieve an SLO.

There could be multiple reasons students are not achieving intended learning outcomes: A simple (and obvious) explanation is often that no faculty are covering the SLO material. More complexly, faculty may be covering the SLO material using ineffective pedagogies or the assignments and learning experiences are misaligned to the learning outcome's level (e.g., faculty are teaching students to *recall* facts when the SLO intends students to be able to *synthesize* the material). Indeed, sometimes the explanation for why students are not meeting the learning outcome is more complicated. Perhaps students are taught the SLO material only in a lower-level course; the content is not reinforced or practiced later on. Over time, students' knowledge or skills deteriorate because content is not properly *scaffolded* (strategically covered) across courses within the major. Or, maybe the new content and teaching strategies in the classroom are not well implemented or received—the learning intended to take place never actually does. In such instances, qualitative data from students about their educational experience in the program can be invaluable to helping faculty make informed, meaningful changes. Investigating and diagnosing the curriculum, an often overlooked practice, is important because the process:

- allows faculty to become intimately familiar with the educational experience students are getting throughout the entire program, not just in the isolated courses or sections that they teach;
- helps faculty identify specific reasons students may not be achieving intended learning outcomes so that they can make more informed and intentional modifications;
- creates a space to discuss shortcomings in the delivery of the new curriculum (e.g., course sequencing or scaffolding, etc.);
- facilitates re-alignment of learning outcomes and assessment instruments; and
- encourages and values student involvement when faculty collect and use qualitative data from students regarding their educational experiences in the program.

Consider the outcome if programs do not investigate their *as is* curriculum and diagnose issues that might contribute to students' lack of success. Faculty may propose modifications to a curriculum based on personal hunches or their own perceptions, but student feedback can contribute to a more robust understanding of how to improve learning. Figure 4 depicts the learning improvement rubric criterion related to investigating curriculum and diagnosing issues.

Let us return once more to the faculty members from the hypothetical communications program at XYZ University who engaged in a learning improvement initiative to help their students effectively deliver an oral presentation. The communications faculty decide to use a different textbook to teach the oral presentation material because the textbook they were using in the past was "dated" and "students needed more modern examples of how to effectively deliver an oral presentation with an engaging introduction, logical and fluid body, and smooth conclusion that reinforces the main ideas of the presentation."

Had the faculty conducted a qualitative study or otherwise collected feedback, they would have learned from their students that the way students are taught to deliver a presentation, craft engaging introductions, and reinforce main ideas in the conclusion is mostly effective but it is more practice, in more courses, that students need. Also, students may have explained that they needed more information on how to deliver a presentation in different contexts and settings, as well as more detailed feedback and a clearer explanation of what "high quality" oral presentations entail. Without knowing that students need more demonstrations of high quality speeches, assistance in making speeches in different context, and time to practice and develop their skills throughout the program, crafting new content and teaching strategies to fit the areas of improvement is nearly impossible.

E. Learning Intervention: The program establishes an appropriate timeline for faculty development, instrument development, and intervention; the intervention is implemented with fidelity; necessary adjustments are made.

Through two learning improvement pilot projects, we have found that successful learning improvement initiatives take multiple years and long-term planning. For instance, one of the pilot programs spent approximately two years before they could evidence initial student learning improvements. Faculty from this program must wait an additional year before

Perhaps the most often forgotten or neglected part of evidencing learning improvement is that faculty must collect baseline data *before* any pedagogy, curriculum, or course sequencing changes are made in the program. Collecting baseline data allows for a more meaningful comparison with post-learning improvement initiative data.

RESEARCH & PRACTICE IN ASSESSMENT ••••••

D. Investigate curriculum and diagnose issues: An explanation of hypotheses for why what was originally being taught in the curriculum relative to the targeted SLO was ineffective.					
	0		2	3	4
1.NoInvestigation:providedProgram-levelor to whaCurriculumgeree, thRelative toselected SISLOis coveredthe "as iscurriculur		A program-level curriculum map is provided; the map shows the courses in which the SLO is covered and indicates the theoretical intensity or degree of coverage (i.e., SLO content is primarily taught in course X, time spent covering content is X, assignments related to SLO are).*	Meets criteria for Beginning and provides a study of faculty to determine if they agree, at the program- level, that the theoretical curriculum matches the actual curriculum.*	Meets criteria for Developing and a scaffolding of the curriculum is provided (i.e., the typical student is taught the SLO content in the following courses to the following degrees).*	A survey and/or focus group, we found that).*
*Note: If inv	estigation accur program	rately reveals little or n s should receive a score	o original conter e of 3 (Good) for	nt coverage for the this criterion.	e selected SLO,
2. Investigation: Individual Course-level Coverage of SLO Content	2. Investigation: Individual Coverage of SLO Content Individual Coverage of SLO Content Coverage of Coverage of SLO Content Coverage of Coverage of SLO Content Coverage of Coverage of SLO Content Coverage of Coverage of		Faculty participating in the intervention provide comprehensiv e, qualitative description of the learning experiences** and how they align with the SLO content.	Meets criteria for Developing <i>and</i> includes approximately how much time students spend with the identified learning experiences.*	Meets criteria for Good and provides evidence that faculty talked to a representative sample of students about the effectiveness of their learning experiences regarding the SLO.**
**Course-leve	el learning expe	riences can be identifie styles, syll	d as/through ass labi. etc.	essments, class ac	tivities, teaching
3. Investigation Conclusions: Logical Insights about Why Efforts are not as Effective as Intended	No insights provided.	Insights provided but do not flow logically from the investigation.	Informed, logical insights provided that point to strengths (if there are any) and weaknesses of how SLO is addressed at <i>either</i> the program <i>or</i> the course level, but not both.	Informed, logical insights provided that point to strengths (if there are any) and weaknesses of how SLO is addressed at <i>both</i> the program <i>and</i> the course level.	Meets criteria for Good <i>and</i> the insights have been vetted through students <i>and</i> external experts or stakeholders.

Figure 4. Investigating curriculum and diagnosing issues criterion of the learning improvement rubric. they can fully implement their learning initiative and hopefully demonstrate the learning

they can fully implement their learning initiative and hopefully demonstrate the learning improvements they had planned. A thorough timeline is critical to keeping track of the learning improvement initiative's route across multiple semesters.

Making changes to a curriculum can include modifying the techniques or strategies faculty use to teach SLO material, introducing new material, changing when/where in the program students learn the content (e.g., course scaffolding), and more. To increase the chances of these changes becoming improvements, faculty participating in the initiative should be well-supported and prepared. Although it is certainly possible for faculty to change their own teaching pedagogies, redesign their courses, and better align course content with the targeted SLO, most will need some assistance before doing so (Fink, 2003). Thus, before faculty can be more effective in delivering SLO material, they would benefit from professional development. Understanding different kinds of pedagogical techniques, learning theories, and course designs

may help equip faculty participating in the initiative. Many peers, faculty development centers, conferences, and online workshops are valuable resources. It is important to remember that changes to both the faculty and student experiences in the program are necessary to evidence learning improvement; that baseline data collection from a well-developed assessment instrument(s) needs to occur before changes are made to the program; and that changing an entire curriculum is a time consuming and intricate endeavor. Figure 5 depicts the learning improvement rubric criterion related to learning interventions.

E1. Learning Intervention: The program establishes an appropriate timeline for faculty development, instrument development, and intervention; the intervention is implemented with fidelity; necessary adjustments are made. 0 2 3 4 1 Absent Beginning Developing Good Exemplary No evidence 1. of what % of Percentage students 90-100% of of Students affected 26-49% of students 50-74% of students 75-89% of students students in Program affected. affected. affected. oraffected. only 0-25% Affected by of students Intervention affected. Meets criteria for Some evidence that A revised Developing and intervention is provides compelling No evidence curriculum map is Meets criteria being implemented 2. of provided that shows explanation for for Good and at the intervention where critical scaffolding (e.g., is validated Programprogram-level but Level at the elements of the SLO explains where and by external it is unclear which experts and programare newly covered why elements of SLO Intervention courses are level. and at what are introduced, students. targeted or at what intensity. reinforced, and intensity. mastered). Some evidence that intervention is 51%-74% of faculty Meets criteria being implemented participating in the for Good in individual intervention have Meets criteria for except 90%-100% courses but one or course student Developing except more of the learning outcomes 75%-89% or more of of faculty No evidence following problems (C-SLO) aligned faculty participating participating 3. of exist: the % of with the appropriate in the intervention in the Individual intervention courses that are degree of coverage have course student intervention Course-Level at the affected is unclear. of the selected SLO; learning outcomes have individual Intervention 50% or fewer of learning experiences (C-SLOs) aligned C-SLOs course level targeted courses for each C-SLO in with the appropriate aligned with are affected or every course is level of the selected the provided: SLO appropriate course-level alignments are made level of the implementation clear in syllabi. selected SLO. does not match program-level plan. Description of Meets criteria faculty development Faculty development for Good and processes are clear, processes are clear faculty Vague references vet, the process is and sufficient; faculty No evidence to faculty consulted insufficient relative member strengths are of faculty preparation (e.g., 4. with outside to what is needed to drawn upon Faculty development we discussed experts effectively intervene (e.g., program implementation in related to the Development or (i.e., "faculty spent disseminates for preparation a meeting, SLO as well 1 hour in training teaching/ pedagogy Intervention for materials were as session;" clearly not knowledge; consults intervention. distributed to curriculum/ enough time for with either an SLO faculty). pedagogical most expert or curriculum design program-level /pedagogical expert). experts. initiatives).

Figure 5. Learning intervention criterion of the learning improvement rubric.

After an appropriate assessment instrument is developed or found (a complex and involved process in-and-of itself) and baseline data are collected, faculty can begin to make changes to the program. In order to systematically evaluate how successfully the changes are being implemented, programs can collect what assessment practitioners call implementation fidelity data. That is, the program can measure the extent to which what is delivered in the classroom differs from what was planned or intended (Fisher, Smith, Finney, & Pinder, 2014; Gerstner & Finney, 2013; O'Donnell, 2008). Implementation fidelity is an advanced technique

Through two learning improvement pilot projects, we have found that successful learning improvement initiatives take multiple years and long-term planning.



that is an incredibly helpful tool yet, grossly underused in higher education. Figure 6 depicts the learning improvement rubric criterion related to interventions.

Creating a timeline, preparing faculty participants, and evaluating progress through implementation fidelity data collection will help programs:

- establish realistic expectations about when and how student learning can improve;
- keep a schedule of data collection and implementation;
- organize efforts to ensure that changes are made at appropriate times and places within the program;
- discover different methods to implementing new pedagogy and curricula;
- make inferences from assessment scores about student learning improvements; and
- identify what changes or modifications can be made to the planned initiative.

	0	1	2	3	4
	Absent	Beginning	Developing	Good	Exemplary
5. Intervention Timeline	No timeline for assessment and intervention is provided.	References assessment and intervention. Not enough details provided to discern sequencing of interventions and assessments <i>or</i> the sequencing departs from assess, intervene, re- asses.	Notes pre- assessment, in what courses the intervention occurs, and post- assessment; the intervention strength (degree of implementation), however, is not laid out.	Notes pre- assessment, in what courses the intervention occurs, and post- assessment; clearly indicates degree of implementation throughout; <i>but</i> time-sequencing may be too ambitious (i.e., does not sufficiently account for instrument construction or faculty development).	Notes pre- assessment, in what courses the intervention occurs, and post- assessment; indicates degree of implementation throughout; clear and well laid out; do-able (i.e., accounts for time to ensure instrument is well-developed; time for faculty development).
6. Intervention Implementa- tion Fidelity: Quality and Adjustments	No context or updates regarding implementa -tion of intervention provided.	Cursory information provided; not enough to make inferences about implementation quality. For example, "the program was implemented according to the timeline" with no further explanation <i>or</i> the implemented intervention differs substantially from the planned	Attempts to investigate the quality of implementation, occasionally bringing up insights of what is changing and where. Nevertheless, the process of checking fidelity is not strong enough to give a full picture of what is actually happening at the program-level.	Systematically investigates program-level fidelity (e.g., surveys of students) throughout intervention implementation. Provides details regarding the match between the planned and actual implementation, yearly. In addition, where issues of implementation arise, program makes or states adjustments for subsequent years.	Systematically investigates course (e.g., auditing) and program-level fidelity throughout intervention implementation. Provides details regarding the match between the planned and actual implementation, yearly. In addition, where issues of implementation or process arise, program makes or states adjustments for subsequent years.

Figure 6. Intervention criterion of the learning improvement rubric.



Again, using the faculty from the communications program at XYZ University who want to improve their students' abilities to effectively deliver an oral presentation as an example, we can demonstrate the importance of allocating time for faculty preparation and instrument development. The hypothetical communications faculty are committed to enhancing their students' oral presentation skills—each member is taking ownership over a well-specified content area and some are researching new ways to teach the material.

However, as the semester begins and the faculty workload becomes more demanding, many of the faculty participants revert back to their original teaching styles and material. Without time allocated for faculty development, few faculty were able to identify and integrate more effective pedagogical techniques for teaching oral presentation skills. As schedules fill up, the faculty meet less frequently. The absence of an implementation timeline contributes to efforts that are inconsistent and uncoordinated, changes that are implemented before baseline data are collected, and progress that is not tracked. As a result, no demonstrable learning improvement exists. The communications faculty have every right to feel disheartened by a lack of success.

F. Re-assess: The impact of the intervention is measured; program-level changes contribute to improvements in student learning.

The final step in the learning improvement process is re-assessment (Fulcher, Good, Coleman, & Smith, 2014; Walvoord, 2010). To demonstrate learning improvement, students must be assessed both before and after they have experienced changes that were made to pedagogy and curricula. Note, in order to demonstrate program-level improvement, re-assessment must take place at the program-level using the same instruments and methodology used to collect baseline data (e.g., standard C). Figure 7 depicts the learning improvement rubric criterion related to re-assessment. Re-assessing students allows faculty to:

- empirically demonstrate that student learning has improved via outcomes assessment data;
- precisely articulate how much or to what extent student learning has improved from pre-intervention (e.g., baseline) to post-intervention;
- integrate outcomes assessment with implementation fidelity data to further tweak or refine any aspects of the learning intervention; and
- use fidelity data to further investigate potential (in)efficiencies of specific features of the learning intervention.

Unfortunately, programs often fail to re-assess student learning after changes are made. Faculty and assessment practitioners alike often mistakenly think that because changes are made to the curriculum or assessment instrument, student learning improves. As Fulcher and colleagues (2014) describe:

They [faculty] make statements like, "We made x, y, and z improvements to the program." But they really mean that they made x, y, and z changes. A change is only an improvement when one can demonstrate its positive effect on student learning. (p. 4).

Imagine the faculty from the hypothetical communications program at XYZ University made changes to their curriculum and pedagogy in an attempt to improve their students' abilities to effectively deliver an oral presentation. Imagine, they also collected baseline data, implemented their agreed upon learning intervention, and now claim students are better at effectively delivering an oral presentation. This could be the case, but because they did not re-assess students' oral presentation abilities after all of the learning modifications or interventions were implemented, faculty have no data or empirical evidence of learning improvements. They have little idea of how well (or poorly) their learning improvement efforts paid off.

We've tried to provide a detailed explanation of our six learning improvement standards. We've also included an example learning improvement report (see Appendix) for the same hypothetical communications program. In the learning improvement report, six

PA 23

F. Re-assess: The impact of the intervention is measured; program-level changes contribute to improvements in student learning.					
	0 Absent	1 Beginning	2 Developing	3 Good	4 Exemplary
1. Re-assess SLO: Same Assessment Measures Used	No pre- intervention data <i>and/or</i> no post- intervention data were collected; therefore, re-assessment is impossible.	Pre-intervention and post- intervention data reported; nevertheless, not enough information is provided to evaluate the veracity of the results (e.g., the methodology from pre- to post- data collection changed to the degree that comparisons are meaningless: different sampling schemes, etc.) or a different assessment measure is used or (test/rubric) items within the same measure change moderately.	Pre-intervention and post- intervention data reported; methodology changed moderately but steps were taken to mitigate those changes (e.g., using SAT scores as a covariate to adjust for group differences). Comparison of pre- and post- data may make some sense but the integrity is compromised.	Pre-intervention and post- intervention data reported. The collection process was reasonable and the method stayed faithful (which could include using psychometrically -verified equivalent test forms).	Post- intervention data reported; evidence that the assessment measures were consistent and data collection processes (i.e., sampling) enable "apples to apples" comparison between pre- (potentially mid-) and post- intervention cohorts.
2. Magnitude of Student Learning Improvement: Statistical Gains	No information regarding student learning improvement is provided.	Student learning improvement (gain) is provided but not in a standardized way or the gain is not statistically significant or the effect size is negligible: Cohen's d of < .15.	Gain is statistically significant and the effect is small: Cohen's d of ~.3.	Gain is statistically significant and the effect size is moderate: Cohen's d of ~.5.	Gain is statistically significant and the effect size is large: Cohen's d of .8 or greater.

Figure 7. Reassess criterion of the learning improvement rubric.

faculty members teaching four courses in the communications program agreed that it was a disservice to allow graduating students to earn a degree without being able to give an effective oral presentation. Furthermore, they decided to use the standards outlined in the learning improvement rubric as a way to guide and evaluate their learning improvement initiative. The example report is annotated with comments that explicitly link the standards included in the learning improvement rubric.

Conclusion: Where does learning improvement go from here?

Learning improvement has been central to the spirt of assessment for decades as evidenced by AAHE's principles of good practice. Twenty years later, there are signs that higher education is slowly pivoting toward demonstrating learning improvement. Calls for evidencing learning at institutions of higher education are being answered with meta-assessment rubrics; experts are vocalizing the need to do better; pilot programs and initiatives are surfacing in several national organizations and independent colleges and universities.

The standards we present via the learning improvement rubric and the example learning improvement report are attempts to advance the conversation of student learning. These resources highlight many of the components that must be in place for learning improvement to succeed.

We have a few suggestions to continue this discussion: some ideas target institutions of higher learning and others are meant for groups representing many states and regions.

To demonstrate learning improvement, students must be assessed both before and after they have experienced changes that were made to pedagogy and curricula. These suggestions are not intended to be an ivory tower wish list. Instead, they are intended to be practical, actionable steps.

At an institutional level:

- Adopt assess, intervene, re-assess (weigh pig, feed pig, weigh pig) as the simple model for improvement;
- Think carefully about and provide resources for academic programs wanting to discuss learning improvement, such resources could include a modified learning improvement rubric and example report that meets internal needs;
- Pilot learning improvement initiatives attempt to find one program that is ready for improvement (i.e., faculty agree on some SLO they want to improve and have buy-in) and try it out using the learning improvement rubric as a guide;
- After an initial success, specify how many programs should show improvement and make it a goal in the university's strategic plan.

For leading higher education groups at a state, regional, and national level:

- Expand many potentially successful initiatives by providing more specifics of what learning improvement is and looks like;
- Pull together resources for actual improvement. (NILOA has a good start but more examples are needed). Again, be more specific regarding advice to practitioners on the ground; advocate for reassessment.
- Give designations for great examples of programmatic learning improvement. Do something similar to what the Excellence in Assessment designation is doing but evaluate related to a more specific definition of learning improvement as opposed to mere quality of assessment practice.
- Provide grants and awards to universities that show improvement at the program and university levels.

We note that this article is published in an assessment journal, RPA, which has extensive reach in assessment circles. While one of our authors (Meixner) is the Executive Director of our Center for Faculty Innovation, this piece is primarily written from the perspective of assessment experts attempting to gather support from fellow assessment experts. Nevertheless, the influence of non-assessment perspectives does not have to stay limited.

We hope other groups pick up and develop program-level learning improvement. Those in faculty development, for example, could provide much more commentary on enhancing pedagogy and curriculum. High-level administrators could provide suggestions for strategically organizing resources. Informal faculty leaders – the ones who make things happen within programs – could provide insight in developing internal support from colleagues. Precocious college students could also contribute. They are, after all, the ones whose learning we are trying to improve. We postulate that the movement's progress will accelerate markedly faster with collaboration from all of these groups and we need to strategize about how we get them to the table.

On a fun note – and bringing us full circle to the title's pig reference – Bellarmine University started what we hope becomes a trend in higher education. They liked the concept of improvement so much that they presented the ideas to their faculty and gave each of them a squeezable pig as a reminder of the weigh pig, feed pig, weigh pig model. We at JMU quickly followed suit; buying and distributing our own squeezable pigs. In fact, JMU and Bellarmine swapped pigs at the December 2015 SACSCOC conference as a gesture of porcine unity. Shortly thereafter a Berea College student leader brought the idea back to his school. Administrators promptly bought and distributed pigs.

Twenty years later, there are signs that higher education is slowly pivoting toward demonstrating learning improvement. Of course, following the assessment cycle and distributing squeezable pigs won't magically foster student learning improvement. Pigs, however, do serve as a fun way to start the conversation by reminding us that learning doesn't improve by merely assessing it. If you would like to be part of this initiative, an easy first step is joining in on this discussion. Come look for us at an assessment conference near you. We'll happily swap pigs with you...and strategies for improvement!

AUTHOR'S NOTE:

As we mentioned, the bulk of the development of the learning improvement rubric was conducted as an independent study with students Smith and Sanchez based on past research on meta-assessment and applied work with learning improvement projects piloted at our home institution. After the independent study concluded and a draft of the rubric was created, Megan Good, Natasha Jankowski, Carol Hurney, and Meixner provided excellent suggestions, especially regarding rubric framing. We acknowledge and extend our gratitude to these colleagues for their contributions to this work.

References

- Banta, T. W., & Blaich, C. F. (2011). Closing the assessment loop. Change: The Magazine of Higher Learning, 43(1), 22-27.
- Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice*. San Francisco, CA: Jossey-Bass.
- Blaich, C. F., & Wise, K. S. (2011, January). From gathering to using assessment results: Lessons from the Wabash National Study (NILOA Occasional Paper No. 8). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Bresciani, M. J., Gardner, M. M., & Hickmott, J. (2009). Demonstrating student success: A practical guide to outcomesbased assessment of learning and development in student affairs. Sterling, VA: Stylus Publishing.
- Chaplot, P., Booth, K., & Johnstone, R. (2010). Building a culture of inquiry: Using a cycle of exploring research and data to improve student success.
- Erwin. T. D. (1991). Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcome. San Francisco, CA: Jossey-Bass.
- Fink, L. D. (2003). Creating significant learning experiences: An integrated approach to designing college courses. San Francisco, CA: Jossey-Bass.
- Fisher, R., Smith, K. L., Finney, S. J., & Pinder, K. (2014). The importance of implementation fidelity data for evaluating program effectiveness. *About Campus*, *19*, 28-32.
- Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014, December). A Simple Model for learning improvement: Weigh pig, feed pig, weigh pig. (NILOA Occasional Paper No. 23). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Fulcher, K. H., Coleman, C. M., & Sundre, D. L., (2016). Twelve tips: Building high-quality assessment through peer review, Assessment Update, 28(4), 1-2, 14-16.
- Fulcher, K. H., Sundre, D. L., Russell, J. A., Good, M. R., Smith, K. L. (2015). Assessment progress report template (APT): Revised rubric draft.
- Gerstner, J. J., & Finney, S. J. (2013). Measuring the implementation fidelity of student affairs programs: A critical component of the outcomes assessment cycle. *Research & Practice in Assessment*, *8*, 15-28.
- Hutchings, P., Ewell, P. T., and Banta, T. W. (2012) AAHE principles of good practice: Aging nicely.
- Hutchings, P., Jankowski, N. A., & Ewell, P. T. (2014, November). Catalyzing assessment deign activity on your campus: Lessons from NILOA's assignment library initiative. Urbana, IL. National Institute for Learning Outcomes Assessment and Institute for Evidence-Based Changes.

- Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Ewell, P. T., Hutchings, P., & Kinzie, J. (2015). Using evidence of student learning to improve higher education. San Francisco, CA: Jossey-Bass.
- Kuh, G. D., Jankowski, N. A., Ikenberry, S. O, & Kinzie, J. (2014, January). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities.*
- Lumina Foundation. (2016). The emerging learning system: Report on the recent convening and new directions for action.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84.
- Palomba, C. A., & Banta, T. W. (1999). Assessment essentials: Planning, implementing, and improving assessment in higher education. San Francisco, CA: Jossey-Bass.
- Popham, W. J. (1997). What's wrong and what's right with rubries. Educational Leadership 55(2), 72-75.
- Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.
- Smith, K. L., Good, M. R., Sanchez, E. H., & Fulcher, K. H. (2015). Communication is key: Unpacking "use of assessment results to improve student learning." *Research & Practice in Assessment*, 10, 15-29.
- Southern Association of Colleges and Schools Commission on Colleges. 2016. *General information on the reaffirmation process.*
- Suskie, L. (2009). Assessing student learning: A common sense guide (2nd ed.). San Francisco, CA: Jossey-Bass Publishers.
- Suskie, L. (2010). Why are we assessing? Inside Higher Education.
- Walvoord, B. E. (2010). Assessment clear and simple: A practical guide for institutions, departments, and general education (2nd ed.). San Francisco, CA: Jossey-Bass.
- Willard, T. (2015). CHEA news release: Winners of the 2015 CHEA award for outstanding institutional practice in student learning outcomes announced.

Appendix

Learning Improvement Report Annotated to Learning Improvement Rubric

Selecting the Targeted SLO

The Communications program endeavors to improve students' oral presentation skills; this goal is articulated through program Objective 4:

Students graduating from the BA program in Communications will (A) effectively deliver a presentation with an (B) engaging introduction, (C) logical and fluid body, and (D) smooth conclusion that reinforces the main ideas of the presentation.

Why are these skills important? According to our alumni survey results, our students often pursue marketing jobs in which oral presentation skills are critical. Additionally, the *Journal of Effective Communications Education* cited oral presentation as the second most important skill for graduate students in the field.

Evidence to Support the Learning Improvement Initiative

While students are performing well on most objectives, they have struggled with oral presentation. Graduating students' skills have fallen below faculty standards in areas A, B, and D listed previously (i.e., effective delivery skills, introduction, and conclusion, respectively) of our oral presentation rubric for the last several years (see Table 1A). Additionally, students self-report their lowest gains in oral presentation (see Table 1B for a summary of these results).

Commented [A1]: Recall the **definition of learning improvement** as rationale for these sections and guiding framework for the learning improvement program presented here. The definition is provided again, here, so that the reader can see clearly the link between the definition, rubric, and example.

Strong evidence, from direct measures, supporting substantive learning improvement due to program modifications.

"This program responded to previous assessment results, made curricular and/or pedagogical modifications, REassessed, and found that student learning improved.

The rationale and explanation of the modifications leading to the change are clearly laid out.

The methodology is of sufficient strength that most reasonable alternative hypotheses can be ruled out (e.g., sampling concerns, validity issues with instrument or student motivation).

In essence, the improvement interpretation can withstand reasonable critique from faculty, curriculum experts, assessment experts, and external stakeholders."

Commented [A2]: Provided rationale for why program selected oral presentation skills as its learning improvement project. This corresponds to **Element/Standard B.1** of the Learning Improvement Rubric: *Program-level* SLO Selection: Rationale with Evidence.

Commented [A3]: Through collaboration with the institution's Assessment Center, the oral presentation rubric was found to be an effective measure at capturing Communication skills. It is important to ensure the measurement instrument is valid and reliable such that gains being seen, or not seen, are captured effectively by the instrument. This corresponds to **Element/Standard C.2** of the Learning Improvement Rubric: Psychometrics: Reliability and Validity.



Table 1A. Objective 4: Oral Presentation	Capstone Assessment	Results of	Three Cohorts, (Oral
presentation Rubric				

Oral presentation Rubric	2011 Results Mean	2012 Results Mean	2013 Results Mean (SD) Desired Mean = 3	Score Difference*** 2012 – 2013	
A. Delivery Skills	2.8	2.5	2.6 (.42)*	No	
B. Introduction	2.7	2.9	2.8 (.55)*	No	
C. Body	3.1**	2.9	3.0 (.38)**	No	
D. Conclusion	2.9	2.7	2.7 (.49)*	No	
*Note. Oral presentation Rubric ($n = 25$): 1 = unsatisfactory, 2 = emerging, 3 = competent, 4 = highly competent					
*Orange coding indicates the degree to which the observed results were worse than desired.					
**Blue coding indicates	the degree to w	hich the observed	results were better than the des	ired result.	

***Based on independent t-tests, using p < .01 as significance level (lower alpha due to multiple comparisons).

Table 1B. Objective 4: Oral Presentation Self-Report, Graduation Survey

Graduation Survey	2011 Results Mean	2012 Results Mean	2013 Results Mean (SD) Desired Mean = 3	Score Difference*** 2012 – 2013
Oral Comm. Skills	2.7	2.6	2.6 (.8)*	No

*Note. Graduation Survey (n = 91): 1 = no gain, 2 = small gain, 3 = moderate gain, 4 = large gain...*Orange coding indicates the degree to which the observed results were worse than desired.

***Based on independent t-tests, using p < .01 as significance level (lower alpha due to multiple comparisons).

Explaining the Course Modification Process: Investigating Oral presentation Skill

Development

In order to begin changing the way we provide oral presentation education,

Communications faculty first investigated how and where we taught these skills across the program curriculum. To do so, we looked at faculty-submitted syllabi and schedules. The curriculum map (see Table 2) lists our required courses along with the degree to which each of our program objectives were *theoretically* covered before this learning improvement project. The

Commented [A5]: This section details how the program investigated the targeted objective BEFORE the modifications were designed and implemented. The purpose of this section is that, too often, programs jump to new modifications without an understanding of current efforts. This corresponds to **Element/Standard D** of the Learning Improvement Rubric: Investigate Curriculum and Diagnose Issues.



Commented [A4]: Here, we see a collection of baseline data. This corresponds to Element/Standard C.1 of the Learning Improvement Rubric: Baseline Data Collection: Timing and Sampling. more instruction time allotted to developing oral presentation skills, the higher the degree of

coverage. Oral presentation skills correspond to program Objective 4.

Course/Learning Experiences	Obj. 1	Obj. 2	Obj. 3	Obj. 4 Oral presentation
COMM201 (Introduction to Communication Theories)	3	0	1	0
COMM301 (Research Methods in Communication Studies)	3	0	1	2
COMM302 (Rhetorical and Scientific Communication)	3	0	1	0
COMM303 (Public Relations)	3	0	0	2
COMM304 (News and Journalism)	3	1	1	0
COMM361(Interpersonal and Organizational Communication)	0	3	1	0
COMM401 (Digital Media and Social Communication)	1	1	3	0
COMM402 (Presentational Speaking)	1	0	1	3
COMM403 (Policy and Campaign Communication)	2	0	0	0
COMM404(Marketing and Corporate Communication)	0	2	3	0
COMM480 (Capstone)	0	2	2	2

Table 2. Curriculum Map of Communications Program (Obj. 4 is Oral presentation)

*Note. Syllabi coverage of Objective 4: 0 = no coverage; 1 = slight coverage; 2 = moderate coverage; 3 = major coverage

As is shown in the curriculum map, four course syllabi addressed the oral presentation learning objective (Objective 4); these courses have been shaded purple in Table 2. In three

courses, oral presentation was theoretically covered at a moderate level (e.g., a few assignments

related to oral presentation skills). One course theoretically covered oral presentation at a major

level (e.g., multiple weeks were dedicated to developing oral presentation skills). On paper, it would seem, students should have ample opportunity to learn how to deliver an effective presentation.

Nevertheless, evidence collected from our oral presentation rubric during the capstone course assessment clearly indicated that students are not as proficient in oral presentation as program faculty expect (e.g., an average of 3 on the communication rubric). Low self-reported gains were also cause for us to change how and to what degree oral presentation skills were taught.

The six faculty members who teach sections in courses with oral presentation objectives (i.e., COMM 301, 303, 402, and 480) met in hopes to discover, as a program, why student scores were falling short of meeting expectations. The meetings were facilitated by our program coordinator and were held three times in the month of March. What follows is a summary of our discussions:

- Indeed, students did verbally present in each of the four courses that had oral presentation course objectives. However, some faculty noticed that students were not taking the assignments seriously. Several faculty members took an informal poll of students in the days after the capstone presentation. Very few students raised their hand when the professor asked if they had practiced the entire presentation at least twice.
- Although the oral presentation rubric was used for COMM480, the capstone course, professors teaching other courses were unaware of the rubric's existence. Many said that the rubric may be helpful in guiding feedback to give to students in their classes.

Commented [A6]: This section corresponds to Element/Standard A of the Learning Improvement Rubric: Faculty Involvement. Across the four courses, oral presentation assignments varied greatly by instructor and section. More often than not, oral presentation assignments and evaluations were more heavily weighted toward course content rather than developing communication skills.
 One professor characterized this trend as follows:

If the presentation was reasonably accurate, the student received an "A," despite lackluster oral presentation skills. I would make comments on the feedback sheet, 'seemed a bit nervous, spoke too quickly...,' but that was about it. I provided more specific feedback regarding the accuracy of the presenter's content. Nevertheless, the presentation quality was far, far away from what would be considered professional or polished.

• Several of the faculty revealed that they did not feel comfortable providing feedback on students' oral presentation skills. Although we acknowledged the importance and necessity of the objective, we had received little or no training regarding how to provide effective feedback.

Course Modifications: Learning Interventions

After meeting with the program coordinator, a course modification plan was devised and supported by all six faculty who teach program courses with an oral presentation component. Given that all four courses are required for all majors, 100% of students in the program will be affected. Note that each of the four courses were modified to some degree; however, the most extensive modifications were in COMM402: Presentational Speaking and COMM480: Capstone. **Commented [A7]:** This section corresponds to **Element/Standard D** of the Learning Improvement Rubric: Hypotheses for why what was originally being taught in the curriculum relative to the targeted SLO was ineffective.

Commented [A8]: Note that ALL SIX of the faculty who teach the oral presentation courses in the program were part of the project. Program-level modifications require coordination across faculty in courses targeted for the modifications. This corresponds to **Element/Standard A** of the Learning Improvement Rubric: Relevant Faculty Involvement.

Commented [A9]: States what percent of program students are affected. Programmatic modifications must affect a large percentage of students. This corresponds to **Element/Standard E2** of the Learning Improvement Rubric: Intervene.

32 **RPA** Volume Eleven | Anniversary Issue 2017

What follows is a short description of each course modification (i.e., learning intervention). An overview of these modifications can be found in Table 3.

Modification 1: Explaining Importance and Clarifying Expectations. One of the first required courses in the major is COMM402 (Presentational Speaking). In this class, students present their final project at the end of the semester. The three faculty who teach COMM402 will explain both the *importance of oral presentation* and the *expectations of program faculty*. Instructors will communicate that this has, in general, been an area of weakness for graduates; furthermore, likely employers and graduate schools desire students who are competent of such skills. Both faculty and students will need to work hard to develop oral presentation skills. Students will watch videos of the three best capstone presentations from previous years. Faculty will describe to their students how each of the presentations were evaluated on the oral presentation rubric.

Modification 2: Using the Oral Presentation Rubric. Oral presentations in each course (COMM 301, 303, 402, and 480) will be evaluated on content (70%) and oral presentation skills (30%). Each faculty member will use the oral presentation rubric to score the 30% of the students' presentations.

Modification 3: Emphasizing Practice. In all courses with an oral presentation component, faculty will urge students to practice their presentations *at least* four times before the in-class performance. Every student will be given a copy of the rubric, provided time in class to practice with other students, and encouraged to tape and review their practice efforts.

Modification 4: Upping the Stakes of Capstone Presentations. For the capstone, the ante will be raised. The final oral presentation will be open to all program faculty and to all

Commented [A10]: Explain program-level modifications in detail. This corresponds to **Element/Standard E1.2** of the Learning Improvement Rubric: Program-Level Intervention. majors; it will also be recorded. The three capstone professors will convey to students that this presentation demonstrates not only what students have learned in the program but also how well-prepared they are for jobs or graduate school.

Note, we (the six faculty members) collectively spent five days in a workshop prior to the first week of Fall classes. The campus Center for Faculty Development, Teaching, and Innovation facilitated discussion and strategy of how to incorporate the listed modifications into our Communications courses.

Table 3. Curriculum Map and Modifications

Faculty will explain the <i>importance of oral presentation</i> and the expectations of program faculty	Experiences	Modifications fied to Course/Learning Experiences
 COMM301 (Research Methods in Communication Studies) Faculty will describe to their students how each of the research presentations were evaluated on the oral presentation rubric an present students with the rubric to clearly articulate the expectations. Oral presentations of final research projects will be evaluated on content (70%) and oral presentation skills (30%) & the oral presentation rubric will be used to score the 30% of the students' presentations. Faculty will urge students to practice their research presentations at <i>least</i> four times before their final presentation to the class. 	COMM301 (Research Methods in fommunication Studies)	 Faculty will explain the <i>importance of oral presentation</i> and the <i>expectations of program faculty</i>. Instructors will communicate that this has, in general, been an area of weakness for graduates and employers and graduate schools desire students who are competent of such skills. Students will watch videos of the three best capstone research presentations from previous years. Faculty will describe to their students how each of the research presentations were evaluated on the oral presentation rubric and present students with the rubric to clearly articulate the expectations. Oral presentations of final research projects will be evaluated on content (70%) and oral presentation skills (30%) & the oral presentations. Faculty will urge students to practice their research presentations <i>at least</i> four times before their final presentation to the class.
 Public relations presentations will be given orally and evaluated on content (70%) and oral presentation skills (30%) & the oral presentations COMM303 (Public Relations) Faculty will urge students to practice their public relations oral presentations <i>at least</i> four times. 	COMM303 (Public Relations)	 Public relations presentations will be given orally and evaluated on content (70%) and oral presentation skills (30%) & the oral presentation rubric will be used to score the 30% of the students' presentations. Faculty will urge students to practice their public relations oral presentations <i>at least</i> four times.
• Oral presentations will be evaluated on content (70%) and oral	COMM402	• Oral presentations will be evaluated on content (70%) and oral
(Presentational presentation skills (30%) & the oral presentation rubric will be used to score the 30% of the students' presentations	(Presentational Sneaking)	presentation skills (30%) & the oral presentation rubric will be used to score the 30% of the students' presentations



	• Faculty will urge students to practice their presentations <i>at least</i> four times.
COMM480 (Capstone)	 Oral presentations of the Capstone project will be evaluated on content (70%) and oral presentation skills (30%) & the oral presentation rubric will be used to score the 30% of the students' presentations. Faculty will urge students to practice their presentations <i>at least</i> four times. The final oral presentation of the Capstone project will be open to all program faculty and to all majors; it will also be recorded. The three capstone professors will convey to students that this presentation demonstrates not only what students have learned in the program but also how well-prepared they are for jobs or graduate school.

Because modifications 1-4 take place in several courses that span multiple semesters of the program, the total effect of the course modifications was not realized or evidenced/captured for several years. In order to provide this final report, we gradually modified courses and evaluated students. In 2014, we evaluated students giving their capstone presentations who had not taken any courses with new oral presentation assignments or instruction. This time point serves as our baseline data point.

In 2015, we again evaluated students giving their capstone presentations using the oral presentation rubric. Because course modifications were made to two courses during both semesters of the 2014-2015 year, students had taken some courses with new oral presentation assignments and instruction. We consider this a "partial modification" time point.

By the time students gave their capstone presentations in 2016, all four courses that we planned to modify were indeed changed. Students graduating in 2016 and 2017, having taken all four modified courses, were evaluated using the oral presentation rubric. This is considered a "full modification." See Table 4 for details.

Planned Course

Communication Studies) COMM303 (Public

(Presentational Speaking) COMM480 (Capstone)*

Modifications COMM301 (Research

Methods in

Relations) COMM402

 Table 4.
 Planned Course Modification and Data Collection Sequencing for Oral presentation in the Communications Program

2015**

2016**

2017**

Commented [A11]: Timeline provided to show macrolevel learning improvement strategy. This corresponds to **Element/Standard E2.5** of the Learning Improvement Rubric: Intervention Timeline.

	This year serves as our baseline data.	This year serves as our partial modification (intervention) data.	This year serves as our full modification (intervention) data.	This year serves as another full modification (intervention) data.
Intervention Implementation Status	Students graduating in 2014 took no courses with oral presentation modifications.	Students graduating in 2015 took two courses with oral presentation modifications.	Students graduating in 2016 took all four courses with oral presentation modifications.	Students graduating in 2017 took all four courses with oral presentation modifications. Also, had added benefit of individual instruction & assignment tweaks.

*Note. Data collected during COMM480; the course is only offered during the Spring semester. **Faculty received training on rubric use before Fall semesters. 0 = No modification; 1 = Explaining importance and clarifying expectations; 2= Using the oral presentation rubric; 3= Emphasizing Practice; 4= Upping the stakes of capstone presentations



Conclusion (2016). Clear evidence, provided by scores from a rubric used to evaluate student presentations in a capstone course, suggested that multiple cohorts of graduates were, in fact, failing to meet faculty expectations. Responses from graduation surveys reinforced the need for better oral presentation education. To determine which courses could be modified, and how, to help students learn, the six faculty in this example met several times. Discussions were insightful and illuminating: some faculty commented that few students took the capstone presentation seriously, many discussed how few opportunities students had to practice their speeches and receive feedback, several faculty teaching lower level courses were unaware of the oral presentation rubric used in the capstone course, and still others noted that they had no training providing feedback on student presentations. A set of course modifications emerged through these meetings. The Communications program did not make changes to all four courses with oral presentation objectives right away. Instead, the program faculty developed a learning improvement timeline. The following example Learning Improvement Report is a complete report that documents *four years* of gradual course modifications and improving student presentation rubric.

Evidence in this report clearly shows (i.e, statistical significance and a large effect size) that the students in this program have improved in giving an oral presentation. Figure 1 illustrates these improvements. Because the cohorts have remained relatively similar throughout the past decade, the Communications faculty can attribute this improvement to the four-course modification student learning improvement initiative. Validity and reliability evidence is provided to support these claims.



Figure 1. Objective 4: Oral presentation Capstone Assessment Results of Three Cohorts, Oral presentation Rubric Average Scores

Conclusion (Updated 2017). After a year of planning and two years of modifying our courses, we are happy to say that students are learning oral presentation skills and meeting faculty standards. As is shown in Table 5, students' oral presentation proficiency in areas A, B, and D of the program objective and oral presentation rubric improved from somewhat below competent (a 3 on the rubric) to midway between 3 and 4 (highly competent); a statistically significant change of over 1 standard deviation (SD) gain (a large effect). See Tables 5A and 5B.

 Table 5A. Objective 4 Update: Oral Presentation Capstone Results of Three Cohorts

Oral presentation Rubric	2014 Results Mean	2015 Results Mean	2016 Results Mean (SD) Desired Mean = 3	Score Difference*** 2014 – 2016
A. Delivery Skills	2.6 (.42)*	3.1	3.5	Yes
B. Introduction	2.8 (.55)*	3.2	3.4	Yes
C. Body	3.0 (.38)**	3.2	3.5	Yes

Commented [A12]: Magnitude of learning improvement provided: The gain is statistically significant and the effect size is at least "moderate" Cohen's D is > 0.5. This corresponds to **Element/Standard F** of the Learning Improvement Rubric: Impact of the Intervention.

Commented [A13]: Direct measure used to evaluate improvement. This corresponds to **Element/Standard F.1**. of the Learning Improvement Rubric: Impact of the Intervention.

Commented [A14]: Note that the assessment results bookend the modifications. Assess – intervene – re-assess. This corresponds to **Element/Standard F** of the Learning Improvement Rubric: Impact of the Intervention

D. Conclusion	2.7 (.49)*	3.3	3.6	Yes

Oral presentation Rubric (n = 25): 1 = unsatisfactory, 2 = emerging, 3 = competent, 4 = highly competent*Orange coding indicates the degree to which the observed results were worse than desired. **Blue coding indicates the degree to which the observed results were better than the desired result.

***Based on independent t-tests, using $p \le .01$ as significance level (lower alpha due to multiple comparisons).

Table 5B. Objective 4 Update: Oral presentation Self-Report, Graduation Survey

Graduation Survey	2014 Results Mean	2015 Results Mean	2016 Results Mean (SD) Desired Mean = 3	Score Difference*** 2014 – 2016
Oral Comm. Skills	2.6 (.8)*	3.3	4.2	Yes

*Note. Graduation Survey (n = 91): 1 = no gain, 2 = small gain, 3 = moderate gain, 4 = large gain, 5 = tremendous gain

*Orange coding indicates the degree to which the observed results were worse than desired.

**Blue coding indicates the degree to which the observed results were better than the desired result.

*** Based on independent t-tests, using $p \le .01$ as significance level (lower alpha due to multiple comparisons).

Please note: In the years before our learning improvement initiative, oral presentation scores on the program's rubric had remained relatively low (e.g., below the desired result of 3, *Competent*). After a thorough investigation, there are no indicators that more recently graduated cohorts of students would be naturally better at giving oral presentations. Our incoming and graduating student cohorts have relatively similar academic demographics (e.g., SAT score averages across the cohort are roughly equivalent). For this reason, we can say that our students improved their oral presentation abilities because they received better instruction, practice, and feedback through a program-level curricular modification.

We took several additional steps in order to ensure that the results documented in this report can be trusted and that learning improvement gains can be linked to the program-level curricular modification:

Commented [A15]: Note that the program summarizes validity evidence of their methodology showing that it is robust. This corresponds to **Element/Standard C.2** of the Learning Improvement Rubric: Psychometrics: Reliability and Validity.

- We carefully selected the oral presentation rubric relative to the oral presentation program objective (content validity);
- we kept the same rubric throughout the entire learning improvement project;
- before non-capstone instructors used the rubric, we had training sessions;
- the Phi Coefficient, an indicator of reliability, ranged from .61 .78 over the years reported (this is an acceptable range for performance assessment);
- over time, the rubric scores correlated with survey scores regarding oral presentation improvement, providing some concurrent validity evidence; and
- more detail regarding the methodology can be found in the program's assessment report (APT).



RESEARCH & PRACTICE IN ASSESSMENT

AUTHOR

Joshua Travis Brown, Ph.D. University of Virginia

Abstract Higher education accountability is a field characterized by complexity. Prior frameworks grounded in psychometrics, economics, and history fall short in explaining the persistence and composition of its complexity. This article employs organizational theory to identify the multiple conflicting approaches of higher education accountability and explain their persistence. The seven identified fields function as specialized silos, each with a unique logic and approach toward accountability, they are: assessment, accreditation, institutional research, institutional effectiveness, educational evaluation, educational measurement, and higher education public policy. The seven accountability silos are systematized into a single conceptual model using an institutional logics framework. This article provides an alternative to the silo-based approach and argues that future accountability efforts must integrate by examining the knowledge domains of other silos to successfully navigate the changing environment of higher education. The implications of an integrated accountability approach are considered for five topic areas: data, the professions, structure, responsibility, and transparency.

The Seven Silos of Accountability in Higher Education: Systematizing Multiple Logics and Fields

Higher education does not lack accountability. Rather it lacks enough of the proper kind, and is burdened with too much of an unproductive kind. (Graham, Lyman, & Trow, 1995, p.7)

ince the late 20th century, colleges and universities have had to respond to persistent calls from multiple social sectors about the expansion of accountability in American higher education. The increased reporting measures are the result of multiple contextual factors that have influenced the system of higher education. In part, the substantial increases in the cost of obtaining a college education have catalyzed the American public to question the value of a postsecondary degree and to call for greater transparency regarding college outcomes (Blumentstyk, 2015; Carey 2015; Webber & Boehmer, 2008). Additionally, many sectors of modern society, such as government, insurance, healthcare, and banking, have been subject to rising levels of standards and standardization as the primary form of regulation, a phenomenon higher education has been unable to avoid (Brunsson & Jacobsson, 2000; Busch, 2011; Lampland & Star, 2009). Finally, lawmakers have increasingly emphasized that resource allocation be awarded based on the performance of the organization, necessitating that the college or university give an account of its educational output (Dougherty et al., CORRESPONDENCE 2016; Doyle, McLendon, & Hearn, 2010). The many drivers of accountability have resulted in a complex system of higher education accountability that is comprised many disparate Email fields and approaches.

jtb8n@virginia.edu

The compounded impact of decades of expanded accountability policies and measures is that universities annually collect multiple types of data at multiple levels in the organization to satisfy multiple regulatory agencies. Administrators and researchers coordinating these efforts within universities have collectively organized into multiple professional fields that advance "best practices" within their respective areas (Banta, Lund, Black, & Oblander,

Volume Eleven | Anniversary Issue 2017

1996). Seven identified fields function as specialized silos, each with a unique rationality and approach toward matters of higher education accountability, they are: *assessment*, *accreditation*, *institutional research*, *institutional effectiveness*, *educational evaluation*, *educational measurement*, *and higher education public policy*. Within the literature, these seven disparate silos lack engagement with one another and possess conflicting definitions of foundational terms. Thus, an important challenge that remains is comprehending both the complex social context and the many disparate approaches to higher education accountability. In this vein, the aim of this article is to develop a conceptual model within which the persistently different accountability approaches may be understood. Moreover, I argue that future accountability efforts must integrate by examining the knowledge domains of other silos in order to successfully navigate the changing environment of higher education.

This article purports that the differential approaches to higher education accountability can be systematically understood through the lens of institutional logics. Institutional logics is a framework from organizational theory used to understand the responses of actorsorganizations and individuals-which operate in complex social environments (Friedland & Alford, 1991). The framework gives consideration to both the internal aspects of organizations, as well as the external forces by which they may be influenced (Thornton, 2004; Thornton & Ocasio, 1999). By employing an institutional logics framework to examine the literature on higher education accountability, I situate the seven accountability silos within the broader context of the market, state, and profession. To achieve this, I first identify the multiple accountability silos by their respective knowledge domains and membership associations. Then, I argue that differences between the multiple accountability silos persist as a result of unique responses to the broader social institutions in which they are embedded (e.g. market, state, and profession). Following this, I map the complex social context of higher education accountability and thereby systematize the disparate silos into a single conceptual model. Finally, I emphasize that effective accountability responses to a changing social context must examine multiple logics and multiple fields.

This article makes three notable contributions to the higher education accountability literature. First, by mapping across the seven fields of higher education accountability, it extends prior research that mapped *within* individual fields of accountability, such as assessment (Ewell, 2009) and institutional research (Volkwein, 1999). Additionally, the mapping of the multiple accountability silos advances the discourse beyond the dominant focus on histories (e.g. how we arrived here) and dichotomies (e.g. research-practice, internal-external, summative-formative, inputs-outputs, and accountability-improvement) that have been traditionally used to describe the complexity of higher education accountability but fall short of explaining its continued persistence and broader composition (Gaston, 2014; Marchand & Stoner, 2012; Reichard, 2012; Suskie, 2015; Upcraft & Schuh, 2002; Zumeta & Kinne, 2011). Finally, employing institutional logics to systematize the field encourages the discourse to focus on integration by giving consideration to the sources of rationality and the disparate responses of the individual accountability silos to broader social institutions.

Overview of the Field

As new policies and regulatory agencies formed over time, universities established offices, practices, and routines within the organization to give an account to various external groups. The university personnel tasked with the oversight of different facets of organizational accountability gradually organized into collective membership associations that provided individuals the opportunity to make sense of their various practices and routines (Volkwein, 2008). Membership associations at the state, regional, and national levels were essential in establishing the boundaries of a field given that they maintained and perpetuated distinct norms, networks, vocabularies, and practices (Reichard, 2012). Membership associations are an important characteristic for distinguishing the disparate fields of accountability given that a field can be identified by locating participants who cohere around a common purpose and carry out exercises that cut across organizations (Thornton, Ocasio, & Lounsbury, 2012). Each association maintains various types of publications in order to communicate with its members, maintain established norms, advance best practices, and sustain the existence of its knowledge domain. Furthermore, each field possesses characteristics that distinguish it from

I argue that future accountability efforts must integrate by examining the knowledge domains of other silos in order to successfully navigate the changing environment of higher education. the other approaches, which include: a unique discourse, scholarly or professional journals, a handbook of research, and a published history of its development.

When the knowledge domain of the membership associations and its scholarly literature are comparatively examined, they may be used to identify the various approaches to higher education accountability. Upon examining these, I identified seven fields of higher education accountability: assessment, accreditation, institutional research, institutional effectiveness, educational measurement, educational evaluation, and higher education public policy. Within a specific college or university, these fields inform the differential organization of individuals, groups, teams, committees, or entire offices. The membership associations affiliated with each accountability field are composed of the many like-minded university employees who have collectively organized around a specific set of shared values, practices, and content (see Table 1). For example, the publications, emphases, and overall accountability approaches by the Association for Education Finance and Policy (AEFP), a policy research organization, are notably different than those of the Southern Associations of Colleges and Schools Commission on Colleges (SACSCOC), a regional accrediting organization. The grouping of the various associations highlight the differences in the seven approaches toward higher education accountability.

 Table 1

 Affiliated Accountability Associations

Accountability Silo	Affiliated Associations		
Assessment	National Institute for Learning Outcomes Assessment (NILOA)		
	Assoc. for the Assessment of Learning in Higher Education (AAHLE)		
	Association of American Colleges & Universities (AAC&U)		
Accreditation*	Regional accreditation: Southern Association of Colleges and Schools (SACS)		
	Disciplinary accreditation: American Bar Association (ABA)		
Institutional Research	Association for Institutional Research (AIR)		
	Society for College and University Planning (SCUP)		
	National Association of College & University Business Officers (NACUBO)		
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1			
Institutional Effectiveness	Association for Higher Education Effectiveness (AHEE)		
	National Council on Macana and in Education (MCME)		
Education Measurement	National Council on Measurement in Education (INCME)		
Evaluation	American Evaluation Association (AEA)		
Evaluation	American Evaluation Association (AEA)		
Higher Education Public Policy	Association for Education Finance and Policy (AFFP)		
Tingher Education Fublic Folicy	Association for Dublic Dolicy Analyzia & Management (ADDAM)		
	Association for Fublic Policy Analysis & Management (APPAM)		
	Council on Public Policy and Higner Education (CPPHE-ASHE)		
Council on Public Policy and Higher Education (CPPHE-ASHE) *For a fuller list of accreditation-related associations, please see <i>Higher education accreditation</i> : How it's changing			

*For a fuller list of accreditation-related associations, please see *Higher education accreditation: How it's changing* why it must, by P.L. Gaston, p. 205–222. Copyright 2014 by Sterling, VA: Stylus.

Not only do differences exist between the seven fields of higher education accountability, but they function as individualistic silos. Within the literature, many of the fields do not engage one another or give consideration to the other forms of addressing accountability in higher education. For example, the field of assessment engages accreditation on matters of student learning outcomes, but takes issue with the use of enrollments and alumni salaries as a meaningful outcome, one that is primarily used in the field of higher education public policy (Baum, May, & Payea, 2013; Gross & Berry, 2016; Schneider, 2016). Furthermore, solutions to ideological tensions or the improvement of practices are predominantly limited to withinsilo perspectives. For example, the field of assessment ardently advocates that members strive to develop and strengthen a "culture of assessment" within individual organizations, rather than give consideration as to how present approaches might be integrated with other accountability silos (Fuller, 2013; Fuller, Skidmore, Bustamante, & Holzweiss, 2016; Ndoye & Parker, 2010). The entrenchment and persistence is due, in part, to individual fields of higher education accountability drawing their logics from the broader social institutions that they engage—the market, state, and profession. These same social institutions are those which have called for the further accountability of higher education. Therefore, to further understand the disparate responses of the multiple accountability silos and their persistent differences, one must examine the broader social institutions in which they are embedded.

The social institutions of the market, state, and profession are those to which higher education must give an "account" of its use of resources and achievement of outcomes.

External Influences on the Field

Prior accountability research has made significant advancements with regard to developing instruments (Shavelson, 2010), articulating histories (Ewell, 2011; Reichard, 2012), establishing best practices (Banta, Lund, Black, & Oblander, 1996), and employing novel methodologies (Doyle, McLendon, & Hearn, 2010; Murnane & Willett, 2010). However, these approaches are unable to examine the social processes and complex social context of the higher education accountability sector. I propose that organizational theory provides scholars with the necessary conceptual tools to understand the complex environment in which various actors—organizations and individuals—function.

More specifically, the institutional logics framework organizes the social embeddedness of actors in order to examine how they are influenced by their social context. The institutional logics perspective assumes that society is not of a singular logic; rather it is comprised of multiple broad social institutions, such as the market, state, profession, family, and religion (Friedland & Alford, 1991; Thornton, 2004). Moreover, each social institution possesses a specific logic and provides the actors embedded within them with unique ways to order their practices, vocabularies, values, and identities (Thornton & Ocasio, 2008). The institutional logics perspective permits researchers to identify the source of differences in practices, discourses, and identities that exist between fields by taking the broader social institutions into consideration. Given that the distinct logic of one social institution (e.g. market) conflicts with the logic of another social institution (e.g. profession), actors must consistently address tensions between the multiple institutions in which they are embedded. For example, College Scorecard emphasizes alumni salaries (e.g. market logic) as valid outcome data for a university, whereas the AAC&U VALUE rubrics emphasize student learning (e.g. professional logic) as valid outcome data for a university. Subject to both of these accountability approaches, a given university must navigate the tensions between the two logics.

Researchers and practitioners of higher education accountability must identify the respective social institutions in which they are embedded if they are to successfully navigate the extant tensions between the various logics. Higher education scholars have continually referred to the same trio of social institutions that influence the postsecondary landscape: the market, state, and profession (Clark, 1983; Slaughter & Rhoades, 2004; Rhoads & Torres, 2006). Joseph Burke (2005) identified this triad of social institutions as those which most influenced the interests, pressures, and priorities of higher education accountability and deemed it the "accountability triangle." The social institutions of the market, state, and profession are those to which higher education must give an "account" of its use of resources and achievement of outcomes. To further understand the role of these social institutions within the higher education accountability context, a typology for each social institution is briefly discussed below. The typologies are a compilation of the select elements and categories of each social institution, as found within the literature.

Market

The institution of the market refers to the social sphere where the exchanges of goods and services occur between buyers and sellers (Scott & Marshall, 2009). The norms of the market center on self-interest and seek to benefit individual actors, whereas its strategy emphasizes the efficiency of transactions (Thornton, 2004). As discussed in the higher education accountability literature, the root metaphor of the market logic emphasizes *performance* (see Table 2). Researchers focus on changes in "performance funding" across states and organizational types. "Performance targets" are monitored via data dashboards using "key performance indicators" (KPIs) such as transfer rates, enrollment data, costper ratios, grant funding, and research output, among many others (Massy, 2016). Broader "institutional performance" is comparatively examined via organizational benchmark data such as graduation rates, alumni salaries, and endowment performance. Market norms are distinct from, and often at odds with, the norms of the profession (Stone, 2002). Scholars have devoted significant attention to examining the impact of the market on the profession of higher education (Slaughter & Rhoades, 2004). The market's expanding influence has brought about substantive changes in the financing of a college degree (Doyle, 2006; Doyle, McLendon,

College Scorecard emphasizes alumni salaries (e.g. market logic) as valid outcome data for a university, whereas the AAC&U VALUE rubrics emphasize student learning (e.g. professional logic) as valid outcome data for a university. Subject to both of these accountability approaches, a given university must navigate the tensions between the two logics. & Hearn, 2010), the affordability of higher education (Archibald & Feldman, 2011), and the acquisition of resources (Berman, 2012).

Table 2	
Institutional Typologies for Market, State & Profession	n

W O		Q	D 0 i
Key Characteristics	Market	State	Profession
Root metaphor	Performance	Compliance	Learning
Basis of norms*	Self interest	Citizenship in nation	Membership in guild
Basis of strategy*	Increase efficiency of transactions	Increase community good	Increase reputation and quality of craft
Organizational form*	Marketplace	Legal bureaucracy	Network organization
Data focus	Data that illustrates outputs, growth, or return on investment	Data that illustrates adherence to policies and standards	Data that illustrates student learning or development
Data treatment	Analysis of efficiency or causality	Presentation of frequency data or narrative argument	Examination of pre/post change or formative/summative

*Denotes a categorical element adapted from *Markets from culture: Institutional logics and organizational decisions in higher education publishing* by P.H. Thornton, p.44–45. Copyright 2004 by Stanford, CA: Stanford University Press.

State

The social institution of the state refers to the collective set of agencies (e.g. armed forces, civil service, judiciary, etc.) that possess the authority to govern a society (Scott & Marshall, 2009). The norms of the state center on citizenship within one's nation, while its form is organized around the concept of legal bureaucracy (Thornton, 2004). As discussed in the higher education accountability literature, the root metaphor of the state logic emphasizes compliance. The state logic focuses on the "disclosure" of information to highlight conformity with "regulations" and "standards" established by the government or their respective monitoring agencies (Brunsson & Jacobsson, 2000). To determine whether the compliance of standards has been achieved, federal agencies rely on two types of approaches. First, agencies oversee the annual collection of quantitative data such as the Integrated Postsecondary Education Data System (IPEDS), Internal Revenue Service (IRS) Form 990, and the Clery Act, to name a few. Second, the agencies coordinate the systematic review of legal narratives, which make the case for compliance or adherence to standards, such as state authorizations for operation (Ewell, Boeke, & Zis, 2010). This dual approach reinforces the legal bureaucratic form that organizes the social sphere as well as the fields and organizations drawing from its logic.

Profession

The social institution of the profession refers to a type of work orientation or work organization for a specific interest group. Entrance into or membership in the group is monitored by a code of conduct, practice, or values (Scott & Marshall, 2009). As discussed in the higher education accountability literature, the root metaphor of the professional logic of higher education emphasizes *learning*. The professional logic focuses on measuring aspects related to the learning of students, including the improvement of learning. Examining learning often takes the form of educational measurement or educational assessment. A measurement approach focuses on employing psychometric techniques to quantify learning, knowledge, or cognitive development (Shavelson, 2010), whereas an assessment approach focuses on examining "student learning outcomes" or "essential learning outcomes" (Arum, Roksa, & Cook, 2016, p. 4). Here, educators are encouraged to utilize "high impact educational practices" and employ "learning outcomes assessments" to examine their influence. Historically, the role of the profession was to establish the quality of craft and safeguard the quality from the influence of the market (Thornton et al., 2012). However, scholars have noted that contemporary professions (education, architecture, accounting, etc.) increasingly must confront the expanding influence of the market logic into the

The term [institutional effectiveness] addresses the systematic examination of the planning and decision making in multiple areas across the university (administrative, educational, etc.) and at multiple levels in order to determine its effectiveness as an organization.



professional domain (Hermanowicz, 2011; Thornton, Jones, & Kury 2005). Therefore, future higher education accountability efforts that use measures of learning must acknowledge their limited application when responding to an expanding market logic.

The influence of the social institutions of the market, state, and profession has garnered the attention of scholars of higher education (Clark, 1983; Slaughter & Rhoades, 2004) and more specifically, higher education accountability (Burke, 2005). This triad of social institutions is the source of logics, order, and rationality for the accountability fields and universities embedded within them. Because the embeddedness of fields and organizations is not mutually exclusive, they are often influenced by multiple logics. Therefore, understanding the differential approaches of higher education accountability is dependent upon identifying the dominant logic or combinations of logics upon which or to which they primarily respond.

Mapping Multiple Logics and Fields

Each of the seven accountability silos is a distinct field supported by a vast literature and a large number of scholars and practitioners in membership associations. Given that limited engagement occurs across the disparate silos, higher education possesses a complex system of accountability that warrants further clarity. This section seeks to converge the seven identified silos and the broader social institutions into a single conceptual model. The differential sources of rationality that ultimately influence the varied approaches to higher education accountability are highlighted by mapping the complex social context.

Assessment

Assessment is the systematic collection, analysis, and translation of evidence on a given topic or outcome (Astin & antonio, 2012; Seclosky & Denison, 2012; Suskie, 2004). Assessment gives priority to student learning, whether the process is led by administrators or faculty, or focused on curricular or co-curricular characteristics of the university. Scholars have specifically noted that assessment is distinct from the other higher education accountability fields of accreditation, measurement, and evaluation (Gaston, 2014; Seclosky & Denison, 2012; Suskie, 2015). The norms of the profession center on collectively established codes such as "Principles of Good Practice for Assessing Student Learning" or "Seven Principles for Good Practice in Undergraduate Education" (Banta, Lund, Black, & Oblander, 1996; Chickering & Gamson, 1987; Upcraft & Schuh, 1996). The field of assessment is predominantly comprised of administrators tasked with the oversight of measuring student learning outcomes across the curriculum of academic disciplines within individual colleges and universities. Recent works encourage faculty to become more involved in the processes of establishing student learning outcomes (Arum, Roksa, & Cook, 2016). Efforts also emphasize the assessment of student learning outcomes across student affairs and co-curricular areas of the university (Bresciani, Gardner, & Hickmont, 2010).

As a field, institutional effectiveness can be described as combining the processes of accreditation (state and profession) with an added emphasis on organizational performance (market).

The field of assessment draws its distinct rationality from the *professional logic* whose root metaphor emphasizes learning (see Figure 1). The broader social institution of the profession guides the responses the field of assessment adopts toward higher education accountability. Data are collected on individual assignments and examined through rubrics in order to determine the extent of student learning. Some choose to showcase student learning through e-portfolios to highlight the array of development or competencies across multiple knowledge domains. Others employ the use of course-embedded techniques in order to strengthen the authenticity of the results. Furthermore, the aim of assessment is often stated to be for purposes of improving teaching and learning, as well as for accountability (Ewell, 2009). The applied and real-world nature of the collected data usually limits its generalizability beyond the context of the specific university. These organized responses by the field of assessment address questions of accountability that are of interest to the profession in an applied manner, and do not address accountability paradigms of interest to the state (e.g. compliance) or market (e.g. performance), as will be shown with some of the remaining silos.

46 Volume Eleven | Anniversary Issue 2017



Figure 1. A Map of the Social Embeddedness of Higher Education Accountability

Accreditation

Accreditation is the peer review process established to examine the educational quality of colleges and universities and to ensure their adherence to standards of practice (Bogue & Hall, 2003; Eaton, 2012; Gaston, 2014). More formally it is, "a process by which an institution of postsecondary education evaluates its educational activities, in whole or in part, and seeks an independent judgement to confirm that it is substantially achieving its objectives and is generally equal in quality to comparable institutions of postsecondary education" (Young, Chambers, & Kells, 1983, p. 21). Three types of accreditors comprise the U.S. system of accreditation: regional accreditors accredit entire colleges, specialized accreditors accredit specific academic programs, and national accreditors accredit entire colleges that are not eligible for regional accreditation (Suskie, 2015). While there are many accrediting agencies in each of the three types, they are all "owned and operated" by the colleges and universities which comprise their membership (Suskie, 2015).

The field of accreditation draws its distinct rationality from a combination of two logics—the *professional logic* (learning) and the *state logic* (compliance). The dual combination of these logics informs the response the field of accreditation maintains toward higher education accountability. The self-regulatory processes of accreditation originated as a form of professional accountability upheld by the ethic of peer review (Gaston, 2014). In constructing arguments for accrediting bodies, universities must provide evidence to argue their compliance with the standards of the accrediting bodies that function as a stand-in for the federal government (Ewell, 2011). The process resembles the practice of peer-review journals whereby the publication of knowledge is governed via the review of evidence by one's peers within the profession. However, declines in public trust across many professions in society have also negatively impacted the perceived legitimacy of the self-regulatory processes of accreditation (Blumenstyk, 2015; Ewell, 2008). Over time, the process has evolved into an organized "federal regulation of academic practices" (Newell, 2012, p. 36).

A persistent challenge for the field of higher education public policy has been to connect market-focused and state-focused variables with learning-focused variables more robust than graduation rates.

Institutional Research

The field of institutional research is comprised of persons and groups whose function within universities is to conduct research in order to "provide information which supports institutional planning, policy formation, and decision making" (Saupe, 1990, p. 1). The "typical" functions of institutional research address four areas within colleges and universities: (a) external and internal reporting, (b) planning and special projects, (c) data management and technical support, and (d) research and development (Volkwein, Liu, & Woodell, 2012). Data to support these four areas are predominantly queried from existing sources of information within the organization. For example, most data reporting requirements for the Integrated Postsecondary Educational Data System (IPEDS) or U.S. News & World Report rankings are fulfilled by institutional research offices or personnel.

The field of institutional research draws its distinct rationality from a combination of two logics—the state logic (compliance) and the market logic (performance). The two logics influence the response the field of institutional research has toward higher education accountability, one that has been characterized as "organizational intelligence" (Terenzini, 1993). The various practices of organizational intelligence aim to provide service and support to faculty members, administrators, and coordinating groups (Stecklein, 1971). Recent advancements in technology and data analysis have enabled institutional researchers to further support university planning and decision making through the use of data mining, predictive analytics, business analytics, and data dashboards (McLaughlin, Howard, & Jones-White, 2012). These approaches permit institutional researchers to examine organizational data in innovative ways that offer the potential of cross-silo efforts toward higher education accountability.

Institutional Effectiveness

Institutional effectiveness is a "multifaceted construct with a myriad of meanings and interpretations" (Alfred, 2011, p. 104). Of the seven silos, it is the most widely misunderstood within the literature given that the term describes a university process, office, and field and is frequently conflated with assessment (Head, 2012). Institutional effectiveness originated in the mid-1980s when the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC) regional accrediting body implemented the term to describe a new emphasis within its stated policies (Ewell, 2012). The notion of assessment had become too contentious and policy makers wanted a "broader and more acceptable" term (Rogers, 1997). Thus, the notion was distinctively different from assessment given that "institutional effectiveness" was to examine all aspects of the university, whereas the field of assessment had limited itself to examining student learning outcomes and development. The term addresses the systematic examination of the planning and decision making in multiple areas across the university (administrative, educational, etc.) and at multiple levels in order to determine its effectiveness as an organization. While different from assessment, the new term strengthened the relationship between assessment and accreditation. By the end of the 1980s, each of the six major accrediting bodies had adopted similar language and practices to evaluate the institutional effectiveness of colleges and universities (Ewell, 2011).

The field of institutional effectiveness draws its unique rationality from all three institutional logics—the professional logic (learning), state logic (compliance), and market logic (performance). Of the seven accountability silos, it is the only one to draw from all three social institutions to inform its response toward higher education accountability. What differentiates institutional effectiveness from the silos of accreditation or institutional research is that it draws its rationality from each of the three social institutions, whereas they draw from two. As a field, institutional effectiveness can be described as combining the processes of accreditation (state and profession) with an added emphasis on organizational performance (market). Although it may be one of the more nebulous silos, institutional effectiveness is identified herein as a field unto itself because over time its meaning has expanded from a policy in a single agency to one that possesses a membership association, professional journal, and published history of its development.

Given that the field of accreditation is organized around the professional system of peer review, it will face increased legitimacy challenges should the societal distrust and decline continue.

Educational Measurement

Measurement focuses on employing psychometric theories and principles to collect student-level educational data using test instruments. The gathering of data based on test items or judgements from examinees enables researchers to make decisions based on inferences from the data (Secolsky & Denison, 2012). Although college entrance exams, such as the ACT or SAT instruments, have received significant scrutiny as of late, they are the most common test instrument used in higher education (Lemann, 2000; Soares, 2011). Other areas of the university that frequently employ the testing approaches of educational measurement are skills assessments for incoming students, major field tests, and career aptitude tests. The majority of these tests are standardized instruments that have been created using an array of psychometric properties that include: classical test theory, item response theory, generalizability theory, reliability, validity, scaling, norming, and statistical modeling to name a few. In Shavelson's (2010) four eras of learning assessment, the origins of standardized testing in the American university (1900-1933) precede the era of external accountability (1979-present). Therefore, it is important to note that the uses of educational measurement and testing have changed over time with the changing social context of higher education.

The field of educational measurement draws its distinct rationality from the professional logic, whose root metaphor emphasizes learning. The broader social institution of the profession guides the response that the field of educational measurement adopts toward higher education accountability. While assessment and education measurement both share the singular professional logic with a focus on learning, the two are distinct fields with different approaches to accountability. Many of the instruments used in the field of education measurement are external learning assessments (e.g. standardized). The instruments are predominantly indirect measures given to students at a single point in time (e.g. SAT, ACT, MFAT, etc.), although some standardized instruments (e.g. College Learning Assessment) have been issued at two points in time or longitudinally in order to address change in student performance (Shavelson, 2010). On the other hand, the field of assessment tends to emphasize internal learning assessments-direct measures to examine the improvement of student learning outcomes.

Educational Evaluation

The Evaluation Thesaurus defines evaluation as, "the process of determining the merit, worth, or value of something, or the product of that process." (Scriven, 1991). Educational evaluation addresses the fidelity of educational programs, or more specifically, asks whether what was proposed and what was delivered were in alignment. Focal data for evaluation exist at the program level, whereas the focal data for the field of educational measurement resides at the student level. Grayson (2012) identifies three types of educational evaluation: formative, summative, and developmental. Formative evaluation focuses on improvement through constructive feedback to program implementers and clients. Summative evaluation emphasizes the measurement of inputs and outcomes in order to determine performance and impact. Developmental evaluation focuses on identifying the effects of innovation and testing hypotheses in situations of uncertainty and complexity (Patton, 2011). An important difference between evaluation and the field of assessment is that evaluation uses information to make an informed judgement regarding performance and efficiency (Suskie, 2004). Process measures provide data that address program quantity (e.g. how much did we do?) and outcome measures provide data that address program impact (e.g. how successfully did we do it?) (Friedman, 2007).

The field of educational evaluation draws its distinct rationality from a combination of two logics-the professional logic (learning) and the market logic (performance). The two logics influence the response the field of educational evaluation has toward higher education accountability. Efficiency is a fundamental aspect of educational evaluation that focuses on the examination of the various costs and resources (e.g. money, facilities, people, etc.) a given program expends in relation to its value or benefits (Boulmetis & Dutwin, 2005). In addition to examining the efficiency of a program, evaluators also address its impact on participants and "return on investment" for the organization (Kirkpatrick, 1994).

While accrediting bodies and individual universities may object to these market- and state-oriented approaches to accountability, the societal distrust in the professions creates a dilemma with regard to maintaining past practices predominantly informed by the professional logic of learning.



The emphasis of evaluators on the relationship between resources and program outcomes highlights the two logics that influence the field—the market logic and professional logic.

Higher Education Public Policy

The field of higher education public policy examines the relationship between resources and education-related policies. Researchers are interested in the ways in which resources are both generated and distributed across systems and how these processes relate in various ways to local, state, and federal policies on education. The field of higher education public policy excels at producing timely research on specific education policies such as performance-based funding (Bouge & Johnson, 2010; Hillman, Tandberg, & Gross, 2014) and federal financial aid (Gross & Berry, 2016). The field also examines the extent to which specific educational policies migrate or diffuse across districts, organizations, states, and regions over time (Dougherty, Natow, Bork, Jones, &Vega, 2013; Doyle, 2006; Doyle, McLendon, & Hearn, 2010). Focal data for the field of higher education public policy are rarely individual case studies of organizations; rather, employed data are more commonly large scale data sets that enable researchers to examine the broader impact of education policies and accountability phenomena.

The field of higher education public policy draws its distinct rationality from a combination of two logics—the market logic (performance) and the state logic (compliance). The two logics influence the response the field of higher education public policy has toward higher education accountability. The dominant paradigm of the field is supported by frameworks from economics and public policy, whereas the dominant paradigm of the field of educational measurement is supported by psychometric theories. Consequently, the variables of interest to researchers in this field also differ from other accountability silos as they examine alumni earnings data, graduation rates, financial aid, resource allocation, resource acquisition, and transfer rates among many others. A persistent challenge for the field of higher education public policy has been to connect market-focused and state-focused variables with learning-focused variables more robust than graduation rates.

Synthesis & Evaluation of the Fields

Applying an integrated paradigm may provide insights as to how fields may respond proactively, rather than reactively, to market and state pressures to govern the profession.

The review of the higher education accountability literature across each of the seven fields highlights three notable characteristics: ambiguity of terms, engagement with other fields, and absent linkages. First, within the literature, scholars repeatedly acknowledged difficulty in consistent definitions of germane terms such as assessment, accountability, and institutional effectiveness (Alfred, 2011; Head, 2012; Wall, Hursch, & Rodgers, 2016; Zumeta, 2011). At times, definitions explicitly conflicted with one another: assessment is comprised of measurement and evaluation (Astin & antonio, 2012) versus assessment is distinct from measurement and evaluation (Suskie, 2004). Some noted that the rapidly changing context of higher education impacts the identity and vocabulary of the field because it is continually responding to new and changing conditions within the regulatory and resource environments (Krist, Jones, & Thompson, 2012). Arguably, the persistent confusion is not necessarily due to a dynamic or changing social context but rather a lack of understanding of the social composition of the context and its influence upon those fields and organizations embedded within it. Mapping the field addresses the ambiguity of terms by establishing typologies for the entire sector—both the social institutions and the individual accountability silos. Furthermore, mapping the field shifts the discourse to focus on sources of rationality and the disparate responses of the individual fields of higher education accountability.

A second notable characteristic within the literature addresses the degree of engagement one field has with that of another. In rare instances, the knowledge domains of fields engage with one another, whereas in most instances there was little engagement at all. In reviewing these occurrences, interaction between silos can be explained based on commonalities or differences in logics. The limited number of fields that engage one another are those which share a root logic or combination of root logics: (a) assessment and education measurement (profession) and (b) institutional research and higher education public policy (state and market). These silos share institutional logics and

thus share a common understanding, vocabulary, and rationality about the context of higher education accountability.

A final notable characteristic pertains to the lack of linkages between social institutions and specific accountability silos. As discussed above, the knowledge domains of some silos possessed common root logics, or "linkages." However, some accountability silos do not share common linkages between social institutions. In part, these missing linkages explain the extant tensions between the accountability silos. The field of accreditation, which is missing a link to the market logic, has recently been challenged with alternative methods of accountability that emphasize performance, return on investment, and alumni salaries (Fain, 2016; Stratford, 2015). The field of assessment, which is missing links to the state and market, must continually defend itself as being a legitimate form of reviewing higher education outcomes. Higher education public policy, which is missing the link to the profession, is efficient with its outcomes but has been unable to provide evidence of learning in ways that appease assessment professionals. A few select scholars have attempted to address topics related to absent linkages (Wall et al., 2014) but further work is necessary to integrate the disparate accountability approaches.

Discussion & Future Implications

The present state of higher education has been one characterized by persistent change. While prior works have discussed how individual silos of accountability have evolved over time as a result of the changing context, the works have not given consideration as to how the broader changes might impact multiple silos (Gaston, 2014; Ewell, 2008; Suskie, 2015; Volkwein, 2008). This final section provides an alternative to the silo-based emphasis toward higher education accountability. It advocates that future efforts must integrate by engaging the knowledge domains of other silos in order to successfully navigate the changing social context. It addresses five important topics: data, the professions, structure, responsibility, and transparency.

Data is an important topic that pertains to each of the seven silos. In this vein, it is worth considering how the various fields of higher education accountability respond as new data types emerge. Recent education research has started to engage the development of analytics, more commonly referred to as "big data." The scope of this type of data addresses learning analytics (Baker & Corbett, 2014), business analytics, predictive analytics (Denley, 2014), action analytics (Milliron, Malcolm, & Kil, 2014), and social analytics (De Laat & Prinsen, 2014), among others. Traditional approaches to assessment have focused on the use of rubrics and e-portfolios as a form of examining the extent of student learning. In contrast, learning analytics provide researchers with new forms of learner data that are both broad and deep, thereby potentially enriching assessment processes (Thille et al., 2014). Applying an integrated paradigm to the data may make it possible to connect learning analytics (logic of learning) with business or predictive analytics (logic of performance) in innovative ways. Furthermore, the broad and deep nature of analytics introduces possibilities to aggregate and disaggregate data in ways that may benefit multiple silos with a single type of data.

A second notable topic to consider is the increasing societal distrust and decline of the professions (Ewell, 2008; Thornton, Jones, & Kury 2005). Given that the field of accreditation is organized around the professional system of peer review, it will face increased legitimacy challenges should the societal distrust and decline continue. More recently, alternatives to hold universities accountable have focused on market- and state-oriented solutions. The required annual reporting of annual alumni salary data to gauge "return on investment" asserts a market based form of accountability. In contrast, the establishment of the United States Education Dashboard by the Department of Education to generate comparative organizational metrics asserts a state-based form of accountability. While accrediting bodies and individual universities may object to these market- and state-oriented approaches to accountability, the societal distrust in the professions creates a dilemma with regard to maintaining past practices predominantly informed by the professional logic of learning. Applying an integrated paradigm may provide insights as to how fields may respond proactively, rather than reactively, to market and state pressures to govern the profession. In an effort to further reduce costs, attempts to structure the system of higher education accountability should occur beyond individual universities and give consideration to redundancies across silos. As we aggregate data, is it also possible to aggregate accountability systems? Whether an organization chooses a centralized, decentralized, or other approach to assigning responsibility for accountability mandates, senior administrators must also give consideration to how the broader seven silos will influence university decisions, as well as the changing context of higher education.

Third, given the multiple approaches to higher education accountability and limited financial resources for universities, the structure of offices and personnel fulfilling accountability requirements within a university should also be considered. Individual universities should consider exploring areas of redundancy and duplication of labor with regard to fulfilling accountability mandates. The alignment of offices, committees, employees, reports and data queries could assist with reducing "administrative bloat" and associated organizational costs (Blumenstyk, 2015; Kirk, 2014). In a recent study, an elite research university calculated that its cost of complying with accountability mandates totaled \$146million annually, or approximately \$11,000 per student (Woodhouse, 2015). Many policy makers vehemently criticized the study, but the university asserted that its broader purpose in conducting the research was to emphasize that matters of compliance costs and efficiencies have seldom been part of the national accountability discourse (Moran, 2015). In an effort to further reduce costs, attempts to structure the system of higher education accountability should occur beyond individual universities and give consideration to redundancies across silos. As we aggregate data, is it also possible to aggregate accountability systems? These broader initiatives to identify redundancy within and across fields could begin to identify commonalities in structures, division of labor, and cost containment.

Similarly, a fourth important topic to consider is that of responsibility. Specifically, with such disparate accountability approaches within the university, who represents the organization with regard to accountability? The notion of accountability and its respective practices impact a variety of employees within individual colleges and universities. A centralized approach might appoint a senior administrator to oversee all accountability efforts for the university in a similar manner that a chief information officer (CIO) represents the university on matters of information technology or a chief financial officer (CFO) represents the university on matters of finance and budgeting. In contrast, a decentralized approach might distribute the responsibility to fulfill accountability mandates to the respective university offices such that the assessment office addresses student learning outcomes, institutional research fulfills data mandates for the Integrated Postsecondary Educational Data System (IPEDS), campus safety addresses Clery Act compliance, and an associate provost handles matters pertaining to regional accreditation, to name a few. Whether an organization chooses a centralized, decentralized, or other approach to assigning responsibility for accountability mandates, senior administrators must also give consideration to how the broader seven silos will influence university decisions, as well as the changing context of higher education.

A final topic of importance is that of transparency. A significant portion of the public scrutiny regarding higher education, and the field of accreditation more specifically, pertains to its "black box" nature (Gaston, 2014). From the public vantage point, there is a degree of uncertainty about exactly what happens inside a college or university regarding its resource allocation, decision making, bureaucratic procedures, and value-added processes, among other operations. Two immediate areas of increased transparency should address financial data and processes. In a market context, protecting one's financial data or delaying its release provides the opportunity to maintain a sustained competitive advantage over other organizations. With many universities competing for financial resources, it benefits a college or university to conceal its data, particularly its financial data. Present federal accountability processes for non-profit universities conceal their financial data for nearly three years before being made publicly available (IRS Form 990). In contrast, present federal accountability processes for for-profit universities require them to make the same financial data publicly available on a quarterly basis (Form 10K). Given that both postsecondary organizational types are funded using public subsidies from federal student financial aid (FAFSA), the transparency time delay for the use of taxpayer monies should be significantly reduced. Equalizing the transparency requirements for these two organizational types will help higher education researchers more effectively examine the behavior of universities in a market context before the data are obsolete. It will also provide a skeptical public with relevant and timely information about how universities utilize the public resources with which they are entrusted.

Just as the availability of financial data should improve, the processes of higher education accountability also require improved transparency. The field of accreditation has recently been scrutinized for maintaining opaque peer review processes as its primary approach to certifying individual colleges and universities (Blumenstyk, 2015; Suskie, 2015). In order to maintain its established legitimacy, the field of accreditation must give consideration as to how it will address the logics of the three broader social institutions (market, state, profession) in a transparent manner. In a comparative case, peer review journals—who share the same professional "peer review" value as the field of accreditation responded to similar scrutiny by refining internal processes and establishing metrics to make such processes more transparent and understood by their host associations and members (ASA, 2016). What metrics can be made available about accrediting agencies that would assuage a skeptical public? Identifying relevant metrics and making those transparent may serve as an initial first step for accrediting agencies or other accountability silos attempting to address the changing context of higher education.

Conclusion

Twenty years ago Graham, Lyman, and Trow (1995) argued that the American system of higher education was not short in approaches to accountability. Rather, they claimed that higher education "lacks enough of the proper kind, and is burdened with too much of an unproductive kind" (p. 7). The multiple approaches, or silos, of higher education accountability have persisted for decades, resulting in increased administrative costs, compounded policies, and redundant practices (Blumenstyk, 2015; Kirk, 2014; Moran, 2015). Prior research grounded in psychometrics, economics, and history has attempted to explain the complexity of higher education accountability (Gaston, 2014; Marchand & Stoner, 2012; Reichard, 2012; Zumeta & Kinne, 2011). However, these frameworks have been unable to explain both the existence of multiple approaches, as well their continued persistence. Rather, I argue that we could more fully understand both the complexity of the environment and the processes of higher education accountability by employing organizational theory. Using an institutional logics framework, the seven accountability silos are systematized into a single conceptual model, which provides actors with a new paradigm for transforming the sector-one that suggests that strategies of change must examine multiple fields and multiple logics.

The model presented in this article systematizes the complexity by identifying the multiple silos of higher education accountability and their broader forces. It offers a new paradigm, and therefore new possibilities, for thinking about the future transformation of the sector. Informed by the systematized model, future changes to the sector of higher education accountability should adopt a three-fold focus on engagement, consolidation, and elimination. First, engagement across multiple accountability silos must occur through means such as scholarly discourse, practitioner interaction, and agency coordination, among others. Engagement across multiple silos will be a challenge given that the established norms, values, and cultures of individual silos have existed for decades. Solutions to ideological tensions must move past the within-silo paradigms and consider the accountability discourses occurring in other fields if the sector is to begin to identify a "proper kind" of accountability. This article provided examples of how the silo of assessment might advance beyond strengthening organizational "cultures of assessment" or how the silo of higher education public policy might consider improved measures of learning beyond graduation or GPA. Scholars across multiple fields might take the lead in transforming the sector by establishing strategic collaborations that yield joint professional meetings, policy reports, special journal issues, or "new directions" volumes of research. Without a commitment to initiate engagement and discourse across the multiple silos the transformation of the sector will be severely limited, if not impossible.

While engagement addresses the discourse between actors in different silos, consolidation addresses the unification of content between different silos. Efforts to integrate the processes, policies, and practices of disparate accountability silos will ultimately highlight redundancies in the system. Policymakers and practitioners must give consideration as to how data predominantly used in one field may also be used to further inform questions of accountability in other fields. As greater numbers of postsecondary organizations engage in market-based practices to ensure organizational sustainability,

The model presented in this article systematizes the complexity by identifying the multiple silos of higher education accountability and their broader forces. It offers a new paradigm, and therefore new possibilities, for thinking about the future transformation of the sector.

53

the accountability mechanisms must correspondingly change to sufficiently examine such practices. Present accrediting processes, which are established on peer review norms, must further incorporate market-oriented approaches of evaluation. For example, accreditation teams could be provided with standardized IPEDS data dashboards that highlight changes in key variables since the previous on-site review ten years prior. Consolidating practices in data collection or reporting would further reduce administrative costs at the organizational level. However, given the scope of consolidation that is necessary, redundancies must also be examined at multiple levels beyond individual colleges and universities (e.g. federal, regional, and state levels).

What remains crucial is the principle this article explains—deriving responses to address the changing social context of higher education must examine multiple logics and multiple fields. The silo of accreditation will be unable to maintain its legitimacy within society if it cannot sufficiently engage the logics of the market, state, and profession.

Finally, efforts to integrate the accountability silos need to call attention to characteristics of the sector that warrant elimination, particularly the "unproductive kind" which continue to burden colleges and universities (Graham, Lyman & Trow, 1995). Consolidation focuses on retaining the effective attributes of the accountability system that must remain, whereas elimination focuses on removing the ineffective attributes of the system. Efforts in elimination must predominantly occur at federal, state and regional levels, where many accountability agencies reside. For example, policy makers and researchers should be permitted to examine "accreditation effectiveness" in similar ways that the regional accrediting agencies monitor "institutional effectiveness" of organizations. A group of researchers representing multiple fields (e.g. assessment, evaluation, and higher education public policy) will examine the silo of accreditation in novel ways, particularly if tasked with examining strategies of elimination. These strategies—engagement, consolidation, and elimination—are necessary to transform an accountability system characterized by complexity.

Addressing the future of higher education accountability amidst a changing societal context is not limited to the topics (e.g. data, the professions, structure, responsibility, and transparency) and examples discussed herein. A limitless number of matters will surface, many of which cannot be foreseen. What remains crucial is the principle this article explains—deriving responses to address the changing social context of higher education must examine multiple logics and multiple fields. The silo of accreditation will be unable to maintain its legitimacy within society if it cannot sufficiently engage the logics of the market, state, and profession. Similarly, the silo of assessment will make few advancements in its practice or methodology if it cannot sufficiently engage across other silos to examine the practices of those fields. Is it possible to advance beyond the "best practices" of an individual field to "best practices" of accountability that promote integration? By looking at multiple logics and the multiple practices across fields, scholars and practitioners can address gaps in ideology, apply novel methodologies, improve efficiencies, and establish innovative approaches in a rapidly changing context.

AUTHOR'S NOTE:

I would like to acknowledge the valuable feedback from Josipa Roksa, Kate Drezek McConnell, Megan Juelfs-Swanson, Michele Darling, Betsy Ackerson, Lauren Germain, Tom Butler, Brandon Moore, and George Cornelius. I am indebted to the RPA editor and reviewers for encouraging me to sharpen and extend my arguments. Any remaining errors or omissions are solely my own.

References

Alfred, R. L. (2011). The future of institutional effectiveness. New Directions for Community Colleges, 153, 103–113.

Archibald, R. B., & Feldman, D. H. (2011). Why does college cost so much? New York, NY: Oxford University Press.

- Arum, R., Roksa, J., & Cook, C. (2016). Defining and assessing learning in higher education. In R. Arum, J. Roksa & A. Cook (Eds.), Improving quality in American higher education: Learning outcomes and assessments for the 21st century. San Francisco, CA: Jossey-Bass.
- Astin, A. W., & antonio, a. l. (2012). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education, (2nd ed.). New York, NY: Rowman & Littlefield Publishers, Inc.
- Baker, R. S., & Corbett, A. T. (2014). Assessment of robust learning with educational data mining. Research & Practice in Assessment, 9(2), 38–50.
- Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (1996). Assessment in practice: Putting principles to work on college campuses. San Francisco, CA: Jossey-Bass.
- Baum. S., Ma, J., & Payea, K. (2013). *Education pays 2013: The benefits of higher education for individuals and society*. New York, NY: The College Board.
- Berman, E. P. (2012). *Creating the market university: How academic science became an economic engine*. Princeton, NJ: Princeton University Press.
- Blumenstyk, G. (2015). American higher education in crisis? What everyone needs to know. New York, NY: Oxford University Press.
- Bogue, E. G., & Hall, K. B. (2003). *Quality and accountability in higher education: Improving policy, enhancing performance.* Westport, CT: Praeger.
- Bogue, E. G., & Johnson, B. D. (2010). Performance incentives and public college accountability in the United States: A quarter century policy audit. *Higher Education Management and Policy*, 22(2), 9–30.
- Boulmetis, J., & Dutwin, P. (2005). *The ABCs of evaluation: Timeless techniques for program and project managers* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Bresciani, M. J., Gardner, M. M, & Hickmott, J. (2010). *Demonstrating student success: A practical guide to outcomesbased assessment of learning and development in student affairs*. Sterling, VA: Stylus.
- Brunsson, N., & Jacobsson, B. (2000). A world of standards. New York, NY: Oxford University Press.
- Burke, J. C. (2005). The many faces of accountability. In J. C. Burke (Ed.), *Achieving accountability in higher education:* Balancing public, academic, and market demands. San Francisco, CA: Jossey-Bass.
- Busch, L. (2011). Standards: Recipes for reality. Cambridge, MA: The MIT Press.
- Carey, K. (2015). The end of college: Creating the future of learning and the university of everywhere. New York, NY: Riverhead Books.
- Clark, B. R. (1983). *The higher education system: Academic organization in cross-national perspective*. Berkeley, CA: University of California Press.
- Chickering, A. W., & Gamson, Z. F. (1987, March). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, *39*(7), 3–7.
- De Laat, M., & Prinsen, F. R. (2014). Social learning analytics for higher education. *Research & Practice in Assessment*, 9(2), 51–60.
- Denley, T. (2014). How predictive analytics and choice architecture can improve student success. *Research & Practice in Assessment*, 9(2), 61–69.
- Dougherty, K. J., Natow, R. S., Bork, R. H., Jones, S. M., &Vega, B. E. (2013). Accounting for higher education accountability: Political origins of state performance funding for higher education. *Teachers College Record*, *115*, 1-50.
- Dougherty, K. J., Jones, S. M., Lahr, H., Natow, R. S., Pheatt, L., & Reddy, V. (2016). *Performance funding for higher education*. Baltimore, MD: The Johns Hopkins University Press.
- Doyle, W. R. (2006). The adoption of merit-based student grant programs: An event history analysis. *Educational Evaluation and Policy Analysis*, 28(3), 259-285.

- Doyle, W. R., McLendon, M. K., & Hearn, J.C. (2010). The adoption of prepaid tuition and savings plans in the American states: An event history analysis. *Research in Higher Education*, *51*, 659–686.
- Eaton, J. (2012). An overview of U.S. accreditation. Washington, DC: Council for Higher Education Accreditation.
- Ewell, P. T. (2008). U.S. accreditation and the future of quality assurance. Washington, D.C.: Council for Higher Education Accreditation.
- Ewell P. T. (2009, November). Assessment, accountability & improvement: Revisiting the tension. (NILOA Occasional Paper No. 1). Urbana, IL: University of Illinios and Indiana University, National Institute for Learning Outcomes Assessment.
- Ewell, P. T. (2011). Accountability and institutional effectiveness in the community college. *New Directions for Community Colleges*, 153, 23–36.
- Ewell, P. T. (2012, July/August). From the states: Learning and accreditation: The elites push back. *Assessment Update*, 24(4), 10–11.
- Ewell, P. T., Boeke, M., & Zis, S. (2010). State uses of accreditation: Results of a fifty state inventory. (CHEA Occasional Paper September 2010). Boulder, CO: National Center for Higher Education Management Systems.
- Fain, P. (2016, June 21). Scorecard for accreditors. Inside Higher Ed.
- Friedland, R., & Alford, A. (1991). Bringing society back in: Symbols, practices, and institutional contradictions. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis* (pp. 232–263). Chicago, IL: University of Chicago Press.
- Friedman, M. (2007). *Trying hard is not good enough: How to produce measurable improvements for customers and communities*. Bloomington, IN: Trafford Publishing.
- Fuller, M. B. (2013). An empirical study of cultures of assessment in higher education. *Educational Leadership Review*, 13(1), 21–28.
- Fuller, M. B., Skidmore, S. T., Bustamante, R. M., & Holzweiss, P. C. (2016). Empirically exploring higher education cultures of assessment. *The Review of Higher Education*, 39(3), 395–429.
- Gaston, P. L. (2014). Higher education accreditation: How it's changing, why it must. Stirling, VA: Stylus.
- Graham, P. Lyman, R. W., & Trow, M. (1995). Accountability of colleges and universities: An essay. In *The Accountability Study*. New York: Trustees of Columbia University.
- Grayson, T. E. (2012). Program evaluation in higher education. In C. Secolsky & D. B. Denison (Eds.), *Handbook on measurement, assessment, and evaluation in higher education*. New York, NY: Routledge.
- Gross, J. P. K., & Berry, M. S. (2016). The relationship between state policy levers and student mobility. *Research in Higher Education*, 57, 1–27.
- Head, R. B. (2011). The evolution of institutional effectiveness in the community college. *New Directions for Community Colleges*, *153*, 5–11.
- Hermanowicz, J. C. (2011). Anomie in the American academic profession. In J. C. Hermanowicz (Ed.), *The American academic profession: Transformation in contemporary higher education* (pp. 216–237). Baltimore, MD: The Johns Hopkins University Press.
- Hillman, N. W., Tandberg, D. A., & Gross, J. P. K. (2014). Performance funding in higher education: Do financial incentives impact college completion? *The Journal of Higher Education*, 85(6), 826–857.
- Kirk, A. F. (2014, May 19). Think college costs too much? Thank the government. The Chronicle of Higher Education.
- Kirkpatrick, D. L. (1994). Evaluating training programs. San Francisco, CA: Berrett-Koehler.
- Knight, W.E., Moore, M. E., & Coperthwaite, C. A. (1997). Institutional research: Knowledge, skills, and perceptions of effectiveness. *Research in Higher Education*, 38, 419–433.
- Krist, P. S., Jones, E. A., & Thompson, K. (2012). Accreditation and the changing role of the institutional researcher. In R. D. Howard, G. W. McLaughlin, & W. E. Knight (Eds.), *The handbook of institutional research* (pp. 3–21). San Francisco, CA: Jossey-Bass.
- Lampland, M., & Star, S. L. (2009). *Standards and their stories: How quantifying, classifying, and formalizing practice shape everyday life*. Ithica, NY: Cornell University Press.

Lemann, N. (2000). The big test: The secret history of the American meritocracy. New York, NY: Farrar, Straus and Giroux.

- Marchand, S., & Stoner, J. (2012). A brief history of accountability in higher education. Phi Kappa Phi Forum, 92(1), 16-18.
- Massy, W. F. (2016). *Reengineering the university: How to be mission centered, market smart, and margin conscious.* Baltimore, MD: Johns Hopkins University Press.
- McLaughlin, G., Howard, R., Jones-White. D., & Reichard, D. J. (2012). Analytic approaches to creating planning and decision support information. In R. D. Howard, G. W. McLaughlin, & W. E. Knight (Eds.), *The handbook of institutional research* (pp. 3–21). San Francisco, CA: Jossey-Bass.
- Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. Research & Practice in Assessment, 9(2), 70–9.
- Moran, M. (2015, July 31). Assessment estimates cost of federal regulation compliance at Vanderbilt. Vanderbilt News.
- Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.
- Ndoye, A., & Parker, M. A. (2010). Creating and sustaining a culture of assessment. *Planning for Higher Education*, 38(2), 28–39.
- Newell, M. (2012, November). The pursuit of credibility: Part three. Career College Central, 6(6), 34-36.
- Patton, M. Q. (2011). Developmental evaluation: Applying complexity concepts to enhance innovation and use. New York, NY: Guilford Press.
- Reichard, D. J. (2012). The history of institutional research. In R. D. Howard, G. W. McLaughlin & W. E. Knight (Eds.), *The handbook of institutional research* (pp. 3–21). San Francisco, CA: Jossey-Bass.
- Rhoads, R. A., & Torres, C. A. (2006). *The university, state, and market: The political economy of globalization in the Americas*. Stanford, CA: Stanford University Press.
- Rogers, J. (1997). Assessment in accreditation: Has it made a difference? Assessment Update, 9(4), 1-2.
- Saupe, J. (1990). The functions of institutional research. Tallahassee, FL.: Association of Institutional Research.
- Schneider, C. G. (2016). Policy priorities for accreditation put quality college learning at risk. Liberal Education, 101(4).
- Scott, J., & Marshall, G. (2009). Oxford dictionary of sociology. New York, NY: Oxford University Press.
- Scriven, M. (1991). Evaluation thesaurus. Newbury Park, CA: Sage Publications.
- Secolsky, C., & Denison, D. B. (2012). Improving institutional decision making through educational measurement, assessment, and evaluation. In C. Secolsky & D. B. Denison (Eds.), *Handbook on measurement, assessment, and* evaluation in higher education. New York, NY: Routledge.
- Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.
- Slaughter, S. & Rhoades, G. (2004). Academic capitalism and the new economy: Markets, state, and higher education. Baltimore, MD: The Johns Hopkins University Press.
- Soares, J. A. (2011). SAT wars: The case for test-optional college admissions. New York, NY: Teachers College Press.
- Stecklein, J. E. (1971). Institutional research. In A. S. Knowles (Ed.), Handbook of college and university administration (pp. 4.123¬–4.134). New York, NY: McGraw-Hill.
- Stratford, M. (2015, September 23). A tougher test for colleges. Inside Higher Ed.
- Stone, D. (2002). Policy paradox: The art of political decision making. New York, NY: W.W. Norton & Company.
- Suskie, L. (2015). Five dimensions of quality: A common sense guide to accreditation and accountability. San Francisco, CA: Jossey-Bass.
- Suskie, L. (2004). Assessing student learning: A common sense guide. Bolton, MA: Anker Publishing Company, Inc.
- Terenzini, P. T. (1993). On the nature of institutional research and the knowledge and skills it requires. *Research in Higher Education*, 34, 1–10.
- Thille, C., Schneider, E., Kizilcec, R. F., Piech, C., Halawa, S. A., & Greene, D. K. (2014). The future of data-enriched assessment. *Research & Practice in Assessment*, 9(2), 5–16.

- Thornton, P. H. (2004). *Markets from culture: Institutional logics and organizational decisions in higher education publishing*. Stanford, CA: Stanford University Press.
- Thornton, P. H. & Ocasio, W. (1999). Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958-1990. *American Journal of Sociology*, 105, 801–843.
- Thornton, P. H., & Ocasio, W. (2008). Institutional logics. In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin-Andersson (Eds.), *The sage handbook of organizational institutionalism* (pp. 99–129). Los Angeles: Sage.
- Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure, and process*. New York, NY: Oxford University Press.
- Thornton, P. H., Jones, C., & Kury, K. (2005). Institutional logics and institutional change in organizations: Transformation in accounting, architecture, and publishing. *Research in the Sociology of Organizations*, *23*, 125–170.
- Uperaft, M. L., & Schuh, J. H. (1996). Assessment in student affairs: A guide for practitioners. San Francisco, CA: Jossey-Bass.
- Upcraft, M. L., & Schuh, J. H. (2002). Assessment vs. research: Why we should care about the difference. *About Campus*, 7(1), 16–20.
- Volkwein, J. F. (1999). The four faces of institutional research. In J. F. Volkwein (Ed.), What is institutional research all about? A critical and comprehensive assessment of the profession (pp. 9–19). New Directions for Institutional Research, no. 104. San Francisco, CA: Jossey-Bass.
- Volkwein, J. F. (2008). The foundations and evolution of institutional research. In D. Terkla (Ed.), *Institutional research: More than just data* (pp. 5–20). New Directions for Higher Education, no. 141. San Francisco, CA: Jossey-Bassy.
- Volkwein, J. F., Liu, Y., & Woodell, J. (2012). The structure and functions of institutional research offices. In R. D. Howard, G. W. McLaughlin & W. E. Knight (Eds.), *The handbook of institutional research* (pp. 3–21). San Francisco, CA: Jossey-Bass.
- Wall, A. F., Hursch, D., & Rodgers, J. W. (2014). Assessment for whom: Repositioning higher education assessment as an ethical and value-focused social practice. *Research & Practice in Assessment*, 9, 5–17.
- Webber, K. L., & Boehmer, R. G. (2008). The balancing act: Accountability, affordability, and access in American higher education. *New Directions for Institutional Research*, 2008, 79–91.
- Woodhouse, K. (2015, August 3). Does compliance cost \$11k per student? Inside Higher Ed.
- Young, K., Chambers, C., & Kells, H. (1983). Understanding accreditation: Contemporary perspectives on issues and practices in evaluating educational quality. San Francisco, CA: Jossey-Bass.
- Zumeta, W. M. (2011). What does it mean to be accountable? Dimensions and implications of higher education's public accountability. *The Review of Higher Education*, *35*(1), 131–148.
- Zumeta, W. M., & Kinne, A. (2011). Accountability policies: Directions old and new. In D. E. Heller (Ed.), *The states and public higher education policy: Affordability, access, and accountability* (pp. 173–199). Baltimore, MD: The Johns Hopkins University Press.

