

# INTERSECTION

FALL 2017

## Contents

<i>President's Letter</i> by Monica Stitt-Bergh .....	2
<i>Note From The Issue Editor</i> by Gray Scott .....	3
<i>A Guide For The Perplexed</i> by David Eubanks.....	4
<i>Beyond The Theatre Of Compliance?</i> by Madeline Murphy .....	14
<i>The Crucial Role Of Faculty Development In Promoting Quality Assessment Practices</i> by Glenn Waterbury, Madison Holzman, Beth Perkins, And Allison Ames .....	17
<i>The Journey Of Improvement: How An Institution Enhanced The Quality Of Their Assessment Process</i> by Michelle Rogers .....	22
<i>Assessing In The Red: How To Keep Quality-Of-Assessment Prospering</i> by Yen M. To & Kipton D. Smilie .....	26
<i>Using Factor Analysis To Test A Measure Of Student Metacognitive Ability Related To Critical Thinking And Intellectual Humility</i> by Jeff Roberts, David E. Wright, And Glenn M. Sanford .....	31
<i>How Do Pros At Large, Decentralized Institutions Manage Outcomes Data?</i> by Josie Welsh, Interviewer .....	38
<i>Interview With David Chase, Associate Vice President - Educational Programming, Accrediting Commission For Schools, Western Association For Schools And Colleges</i> by Jonathan Keiser, Interviewer .....	42

# PRESIDENT'S LETTER

By Monica Stitt-Bergh

Quality of assessment—the theme of this issue—starts with a clear vision, a decision about what's at stake, and the resources required to achieve the vision given those stakes.

I take this to heart in my own work on a college campus and also in my role as President of the AALHE. I want AALHE to carry out the quality assessment practices that we promote on our campuses.

With that in mind, the AALHE Board of Directors clarified the organization's mission and goals last June to reinforce a collectively-shared vision. Afterwards, the Board members, as highly-experienced assessment practitioners themselves, developed SMART outcomes and aligned activities of their committees and task forces to the outcomes. We expect to be ready to report on progress and quality by year's end.

Next year we will move from self-assessment to using external evaluations gathered from members who complete the annual member survey and from conference attendees who submit the conference evaluation form. We will use our findings to grow and provide services to assessment practitioners in an intentional, well-founded manner. In this way, we are reflective organizational leaders, committed to using and willing to use evidence in decision-making.

Here's a sample of what AALHE committees and task forces have planned as part of our collective vision and what we will report on next year:

- Publish assessment-related literature in a variety of formats, such as the *Conference Proceedings*, *Intersection*, and *Emerging Dialogues*
- Provide professional development opportunities (via webinars, through online discussion groups, and in person), promoting these opportunities on various social media platforms
- Maintain our vibrant community of assessment practitioners and recruit new AALHE members
- Form new partnerships and collaborations with other assessment organizations
- Write and distribute a white paper on perceptions of higher education assessment held by people involved with assessment on their campuses
- Analyze the last four decades of assessment literature to document advances in the field

The committees and task forces are run by AALHE members who volunteer their time and expertise to bring you professional development opportunities, stay in touch with you, and create publications such as this one. Thank you for opening this issue of *Intersection*. I trust you will find it useful.

Aloha.

---



MONICA STITT-BERGH IS PRESIDENT OF THE AALHE AND AN EDUCATIONAL PSYCHOLOGIST IN THE ASSESSMENT OFFICE, UNIVERSITY OF HAWAII AT MĀNOA. SHE CAN BE REACHED AT [MSTITT-BERGH@AALHE.ORG](mailto:MSTITT-BERGH@AALHE.ORG)

# NOTE FROM THE ISSUE EDITOR

By Gray Scott

The FALL 2017 call for submissions asked contributors to think about *quality of assessment*. The meaning of that phrase was left deliberately open, with writers encouraged to think of it along fairly obvious lines (like validity and reliability), but also to think about the question in ways we might not have anticipated.

The resulting--and outstanding--submissions featured a range of interpretations. We ultimately went with several pairings of articles, each pair representing a slightly different kind of response, and in our organization of this issue, we have grouped those pairs together. Two pieces are thought-provoking challenges to, and critiques of, the status quo. Two address the matter of faculty buy-in, a necessary (though not sufficient) condition for excellence in assessment quality. Two address assessment in terms of research methodology.

In connection with our quality-of-assessment theme, Josie Welsh interviewed two veterans of assessment at large institutions well-regarded for their programs: Stephen Hundley, senior advisor to the chancellor for planning and institutional improvement at Indiana University—Purdue University Indianapolis, and Keston Fulcher, executive director of the Center for Assessment and Research Studies at James Madison University.

In addition, we have a new entry in our ongoing *Conversations with Accreditors* series, this time featuring David Chase, associate vice president of educational programming at the Accrediting Commission for Schools, Western Association for Schools and Colleges. Chase is interviewed by Jonathan Keiser.

In other news, our editorial board has expanded significantly. We would like to welcome to the team Jana M. Hanson, director of institutional improvement at South Dakota State University; Alison Witherspoon, director of assessment and accreditation at American College of Education; Michelle Rogers, assessment specialist at Des Moines University; Elizabeth Smith, planning director at the Yale National Initiative & Teachers Institute for Tulsa, at the University of Tulsa; Jeff Barbee, assessment and evaluation specialist at Indiana University School of Medicine; and Steven J. Michels, assistant provost for teaching and learning at Sacred Heart University.

Comments and suggestions from readers related to the current, past or future issues are welcome. Please feel free to send your ideas to [publications@aalhe.org](mailto:publications@aalhe.org).

## **Editorial Board**

Jane Marie Souza, Editor-in-Chief, University of Rochester  
David Eubanks, Furman University  
Jacob Amidon, Finger Lakes Community College  
George Klemic, Lewis University  
Gray Scott, Texas Woman's University  
Josephine Welsh, Missouri Southern State University  
Jana M. Hanson, South Dakota State University

Alison Witherspoon, American College of Education  
Michelle Rogers, Des Moines University  
Elizabeth Smith, University of Tulsa  
Jeff Barbee, Indiana University School of Medicine  
Steven J. Michels, Sacred Heart University



GRAY SCOTT, ASSISTANT  
DIRECTOR OF ACADEMIC  
ASSESSMENT AND ASSOCIATE  
PROFESSOR OF ENGLISH AT  
TEXAS WOMAN'S UNIVERSITY

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

# A GUIDE FOR THE PERPLEXED

By David Eubanks

*On two occasions I have been asked, "Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?" ... I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.*

—Charles Babbage

The difficulty in using assessment results to improve academic programs is a recurring theme at assessment conferences. This topic also puzzled me greatly for years. I read books and attended conference sessions that described how to define learning outcomes, create rubrics, map a curriculum, and so on. The theory was beautifully simple: set a goal, measure the goal, then use the data to make adjustments. It was that last part where everything seemed to fall apart.

If you have wondered why it is so difficult to use the outputs of standard assessment practices into a credible understanding of learning, you may have at times felt—like I have—that you were simply not following the script closely enough. *Surely all those other people are making it work. Why can't I?* The intent of this “guide for the perplexed” (with apologies to Maimonides) is to show why it is not your fault.

My conclusion, after seeing hundreds of real assessment reports from many different institutions, supervising assessment programs at four institutions myself, and talking to many other assessment practitioners, is that *it is difficult to use assessment results because the methods of gathering and analyzing data are very poor.*

Academic assessment is like any other data-driven enterprise: effectiveness stems from careful attention to the details and bearing in mind the prime directive of any research activity. Richard Feynman said it best in his 1974 commencement address: “The first principle is that you must not fool yourself—and you are the easiest person to fool.”

## **The Rules of Assessment**

The functions of university-wide assessment programs are driven by regulatory requirements as described by regional accreditors. I will use Middle States standards as a template for regulator-prescribed practice, because they are particularly detailed. The requirements below are condensed versions of the criteria in Standard V, which you can find (in video form) [here](#). Quotes are from the video narration.

1. Inter-related goals between programs and institutions, with emphasis on assessing programs and institution (e.g. not individual courses). Documentation must link the mission to outcomes.
2. Assessment at institution and program level “should be of such quality that they meaningfully evaluate the extent of student achievement. [...] Assessment processes should enable faculty and other qualified professionals to identify strengths and weaknesses with regard to the student learning outcomes [...]. Assessments used should be defensible, meaning that they involve direct observation of the knowledge, skills, and habits of mind or values that students are expected to achieve. [...] In summary, the Commission expects accredited institutions to demonstrate that an organized and systematic assessment has prompted meaningful and useful discussions about strengths and weaknesses with regard to student learning outcomes [...]”
3. The assessment results must be used for the improvement of educational effectiveness.

This list describes a research program that depends on good data and well-reasoned analysis. In the second item we find the requirement that assessment data should “meaningfully evaluate the extent of student achievement.” This is a measurement task, in other words. The research program is not allowed to fail, for example by yielding inconclusive results, but must be used to guide decision-making.

Trudy Banta and Charlie Blaich described the corresponding reasoning behind the assessment movement in “Closing the Assessment Loop” (2011).

An internally driven, formative approach to assessment is based on the belief that a key factor inhibiting improvements in student learning or allowing students to graduate without learning enough is that faculty and staff who deal with students lack high-quality information about the experiences and conditions that help students learn. If they had information about how much their students were or were not learning and the practices and conditions that helped them learn, practitioners would put this knowledge to work, and improvement would naturally follow. (pg. 27)

The belief is that the barrier to improving programs is a lack of good information. In this light, the Middle States requirement to measure learning is logical: it is intended to provide this essential ingredient.

However, if the data *do not* “meaningfully evaluate the extent of student achievement,” then the requirement is Kafkaesque, requiring institutions to legitimize the use of bad data, and punishing them when they cannot.

Fortunately, there is a mature body of work on how not to fool oneself with educational measurement (Brennen, 2006). Unfortunately, that accumulated knowledge can almost never be applied because of the large number of concurrent assessment projects and consequent lack of attention each can get.

### **Explaining Failure**

Program assessment requirements like the ones quoted above apply to most institutions of higher education in the United States: thousands of institutions and their respective academic programs, each with a handful of outcomes to be assessed, which must number in the hundreds of thousands when taken together. This program has been in place for more than a decade, comprising a huge number of mini-research projects. It is reasonable to expect that if this program of data gathering and analysis were successful that it would have produced a great volume of useful findings about pedagogy, curriculum, and student development.

In the same article quoted above, Banta & Blaich (2011) looked for such examples of successful assessment efforts. What they found surprised them.

We scoured current literature, consulted experienced colleagues, and reviewed our own experiences, but we could identify only a handful of examples of the use of assessment findings in stimulating improvements. (pg. 22)

As Fulcher, Good, Coleman & Smith (2014) point out, the 6% of submissions that Banta & Blaich found to identify improvements is bound to be an overestimate of the actual case, since these submissions were chosen presumably on their merits, and not at random.

There are two possible conclusions. One is that the faculty are generating good data, but are not using it. This way of thinking extends the diminishment of faculty expertise that began with telling them that grades do not measure learning: we’ve replaced grades with something better—assessments that *do measure learning*—but they still are not producing the intended results. Therefore (the argument goes) we just need

to work more on our processes, so that when the faculty finally do fully adopt these changes a fountain of good educational research will spring forth.

There is another possible conclusion from the Banta & Blaich article, one that is confirmed by my decade of experience: it is not that the faculty are not trying, but the data and methods in general use are very poor at measuring learning.

### **Common Sense Isn't Enough**

In 1989, Patrick Terenzini published “Assessment with Open Eyes: Pitfalls in Studying Student Outcomes” in the *Journal of Higher Education*, an article that anticipates our current situation. On measurement Terenzini wrote:

[...] though locally developed measures may be more carefully tailored to local purposes and educational objectives [than standardized instruments], they are also likely to be untested (at least in the short run) and, consequently, of unknown reliability and validity. [...] Many faculty members will have neither the time, commitment, nor competence to develop local measures. (pg. 657)

This problem is exacerbated because “The certainty implied by statistical testing can mask problems that may lead to the serious misinterpretation of results.” A specific problem is that

Common sense suggests that if one wishes to know whether something changes over time, one should measure it at Time 1 and again at Time 2. The difference between the pre- and post-test scores, the “change” score, presumably reflects the effects of some process. [...] In this instance, however, common sense may harm more than help. (pg. 660)

He describes problems with the common-sense approach of naively comparing one set of numbers to another, including unreliability of difference scores, ceiling effects, and regression to the mean. See Bonate (2000) for a book-length treatment on how to analyze pre/post scores. The common-sense oversimplification of measurement is a general symptom, I believe, of the trade-off that favors breadth of assessment efforts over meaningful depth.

For readers looking for an excuse, Terenzini provides one, writing that “Methodological standards for research publishable in scholarly and professional journals can probably be relaxed in the interests of institutional utility and advancement,” concluding that “Although the methodological issues reviewed here cannot and should not be ignored, neither should one's concern about them stifle action.”

This sentiment may seem palliative, as in: *of course, we can't hold ourselves to a high standard of intellectual rigor. We don't have the resources for that.* The whole assessment process would fall apart if we had to test for reliability and validity and carefully model interactions before making conclusions about cause and effect.

How would we feel if the airline industry took that approach to building, flying, and maintaining aircraft? Should we also revert to a pre-scientific era of medical research because randomized trials are difficult and expensive? Are student outcomes valued so much less than health and safety that we should abandon all but the pretense of rigor for the majority of our work?

The disregard for measurement quality combined with the perils of common-sense inference create problems for innumerable assessment projects. There is a sense in the assessment community that once all the proper processes have been followed, then the data produced are inherently meaningful. As such, all manner of comparisons within the data are called forth to illustrate possible uses. One outcome may have

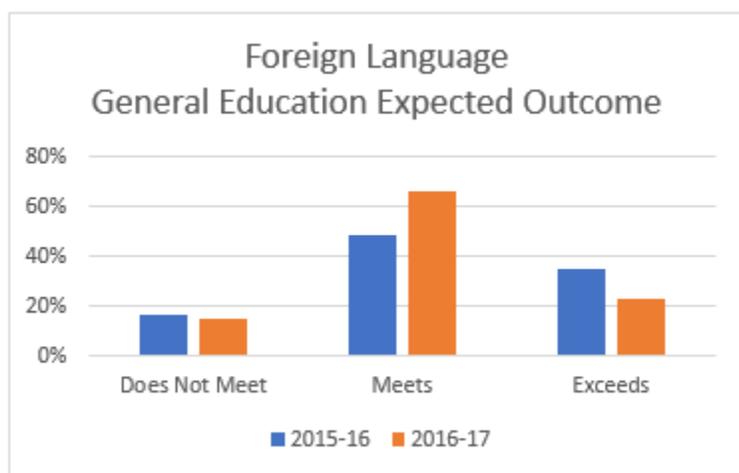
larger averages (or different distributions) than another, or vary from one year to the next, and meaning is read into these differences with only common sense as a guide. Because of the limitations in time and expertise, measurement and statistical considerations are waived. It becomes a pernicious enthymeme: *we used the proper process, ergo the results are guaranteed to be meaningful and amenable to common sense understanding*. Under these conditions the use of data is akin to a Rorschach test.

### **Using and Misusing Data**

The ability to infer meaning from data requires good data and good models of inference. It may be helpful to illustrate this with a real example.

The common-sense method of using assessment data goes like this: (1) find a number that looks too low in assessment results, and (2) imagine some change that might raise the number. Popular changes include adding a new subject to the syllabus, changing reading or assignments to emphasize some aspect of learning, changing a textbook, or sometimes adding a new course to the curriculum. What is being *replaced* in the process is rarely addressed.

The example below shows real data from assessing basic foreign language proficiency over two academic years. Course instructors rate student performance in the language courses for general education using a three-point scale (does not meet expectations, meets expectations, or exceeds expectations), using a rubric established by the faculty when the general education curriculum was created at my institution.



*Figure 1. Summary Graphs of Language Proficiency*

We can see that about 16% of students are not meeting the expectation. The graph gives us some information, but not much. The requirement to use these results to make improvements typically leads to conclusions such as the following:

Over the last two years, 16% of students have not met expectations in the general education language proficiency. An analysis of end-of-term papers shows that lack of proficiency is related to problems with basic vocabulary and grammar. Consequently, we will spend an extra week at the beginning of the term reviewing this material.

On the surface, this “solution” addresses the problem, but it has the hidden cost of reducing content for 84% of the students who do not need the review. Worse, it has been known for year that the common-sense solution of remediation can backfire (Hillocks, 1986).

We now take another look at the same data with a more inclusive model of cause and effect. One useful approach is the Astin model (1991), which can be used to categorize interactions as shown in Figure 2.

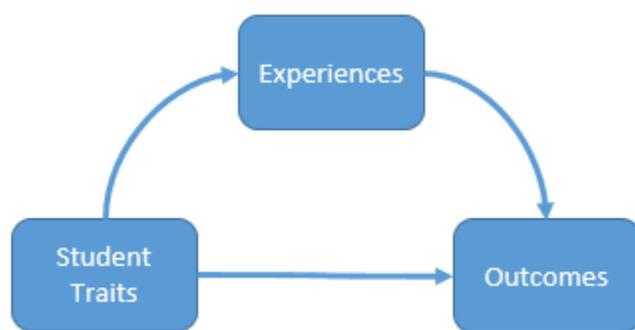


Figure 2. Astin's Input-Environment-Output model

In this case we have:

- **Traits:** Students arrive with varying degrees of academic preparation, which can be partially assessed via their high school transcripts. In particular, a recalculated high school grade average (HSGPA) predicts college GPA, and can be considered a measure of a student's academic preparation, talent, and work habits.
- **Experiences:** Students may wait zero, one, two, or three years before enrolling in the foreign language courses required by the general education requirement. These constitute different learning experiences, since they will begin to forget what they learned in high school.
- **Outcome:** The same ratings found in Figure 1.

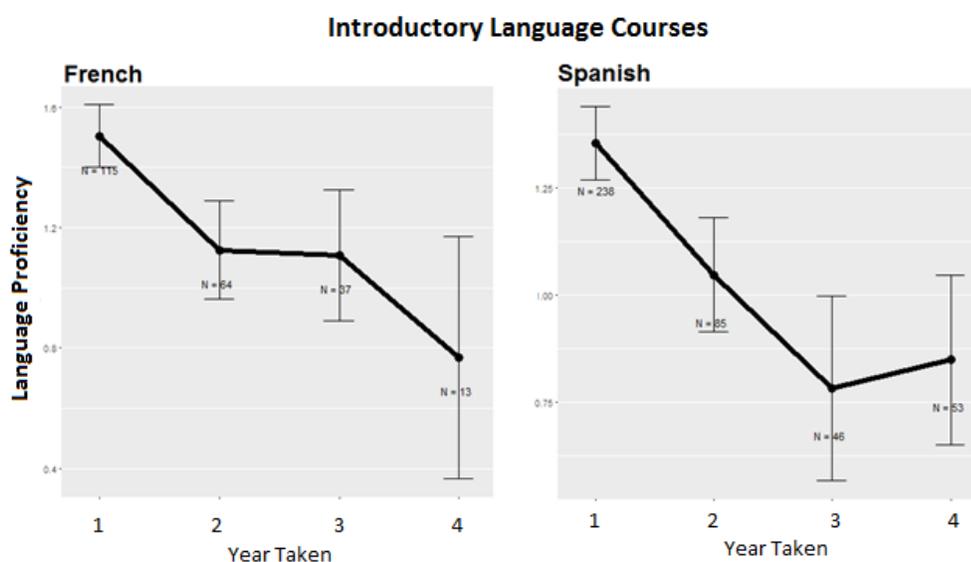


Figure 3. Language Proficiency by Year the First Course Was Taken

The graphs in Figure 3 relate the experience (year taken) to the outcome (rubric rating) and show that for both French and Spanish, students who wait even one year show significantly decreased outcomes on average. This suggests two possibilities. Maybe students who wait to take language courses simply forget what they learned in high school, and their learning suffers. Or it could be that students who are weaker academically avoid the course as long as they can, and it doesn't matter when they take it. To resolve this, we introduced the student characteristic variable (HSGPA), and a regression analysis finds that *both* the wait *and* HSGPA contribute to the decline in scores. Discussions with the language faculty confirmed that this finding was reasonable. The solution is different advising, to prevent students from waiting to take the required language course.

The language program faculty take assessment very seriously, and several of them are certified as instructors for teaching the rating system adopted by the American Council of the Teaching of Foreign Languages (ACTFL). For many years they have been pulling a selection of student essays and rating them according to the ACTFL rubrics. They had gotten accustomed to the 16% unsatisfactory rate, and eventually just assumed that this was the best the program could do. In fact, there was nothing in those essays that would have told them what the actual problem was. They were not surprised by the graphs in Figure 3, and even pointed me to published research that confirmed the finding. With the language faculty's support, the graphs in Figure 3 are powerful in communicating to advisors the danger of letting students wait to take these courses.

The effect in Figure 3 is also detectable using course grades, and a scan of all the 100- and 200-level courses taught at my institution identified other introductory courses (especially mathematics), where it is detrimental for a student to wait a year.

The common-sense use of assessment data illustrated in the discussion of Figure 1, is not complex enough to account for real educational processes, *even when the assessment data are meaningful*.

Unfortunately, data that result from usual academic program assessment activities are inadequate to use a model like the schematic in Figure 2, even if someone has the time to do it. The mandate to use the results leads to a random shuffling of educational practices, or *post-hoc* justification of a change that is desired for other reasons.

### **Data Problems**

Because assessment data must be mass-produced, we typically create dozens or hundreds of shallow pools of data, with small decontextualized samples. There is no time to diagnose, let alone fix, the data problems. This creates insurmountable problems for analysis.

**Samples of student work or observation are small** (e.g. <100), making it likely that even if measurements are good, we will still get the wrong answer to many of our questions. Small samples also make it impossible to assess reliability and validity. The graphs in Figure 3 are based on hundreds of observations; with small sample sizes (e.g. 30) it would not be possible to detect the effect shown there.

**The data are decontextualized**, for example by not considering student characteristics. Omitting context leaves out the most powerful means of discovering cause and effect, as in the example above with foreign language proficiency. It is also essential for assessing change. See (Ewell, 1991), "student learning and development is a complex, multifaceted phenomenon unusually resistant to single-factor explanation" (pg. 95). As Ewell notes, this leads to longitudinal studies. See Singer & Willett (2003) for a comprehensive

statistical treatment on using longitudinal data to estimate change. See Kilgo, Sheets, & Pascarella (2015) for an example assessing high impact practices.

Recall that the mandated objective of assessment work is to place “emphasis on assessing programs and institution (e.g. not individual courses).” But “a program” is almost certainly a different experience for every student (refer back the Astin model in Figure 2), with different instructors, different courses in a different sequence, different starting preparations, and many other important variables that can affect “program outcomes.” Additionally, each student is different, with unique academic strengths and interests, and so on. Since none of these variables is usually accounted for, only large effects could possibly be detected, and even then we may fool ourselves as to the cause. This is a hopeless situation given the state of the actual data and inferential methods used. *If an effect is large enough to show up under these conditions, the faculty almost certainly already know about it from their experiences with students.*

Even in cases where a curriculum is highly structured (e.g. cohort-based with a fixed course sequence), it is necessary to take into account student traits when trying to understand the cumulative effect of the curriculum.

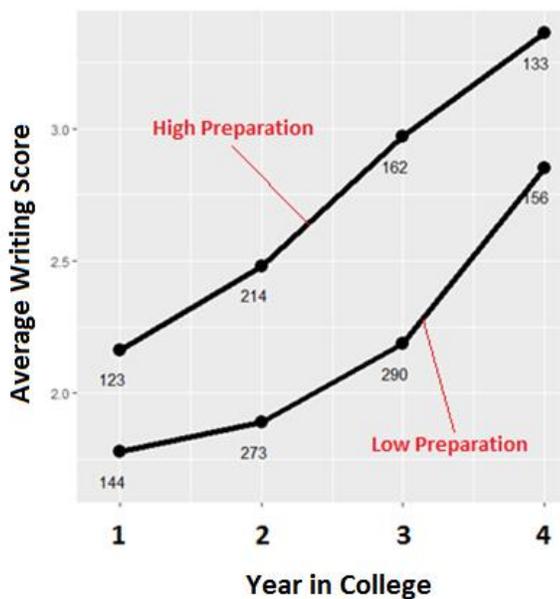


Figure 4. Writing Scores by Academic Preparation, with Numbers of Students

The developmental paths in Figure 4 show two different average trajectories for writing scores at my institution (based on two years of data). The top line shows students who were in the upper half of the standardized high school grade average for their college entering class. The bottom line is the lower half, by high school grades. It is well known that high school grades predict college grades reasonably well, so it is not surprising that writing ratings would show a similar effect. Notice that the lower group appears to lag the upper group by a year or two (survivorship bias is strongest in the lower group, which slightly inflates those scores).

The point of this illustration is that if an analysis does not incorporate levels of student preparation, even if the data are good and the program completely standardized, the results may be driven by varying student qualities and not program effects. Imagine a program so terrible that only the most determined and talented students can survive it. The assessment results will glow with the accomplishments of these talented few,

and since there are no results for all the ones who dropped out immediately, they are invisible to assessments. In other cases, a change in assessment measures may be attributed to program characteristics rather than changing student traits. It is important to understand the difference when trying to make improvements (e.g. more rigorous curriculum versus more tutoring versus better advising).

### **Language as Camouflage**

So why is it not generally accepted that poor data and common-sense inference invalidate the majority of assessment projects? On the contrary, judging from the rhetoric within the assessment community and from accreditors, there is great confidence in the processes that are in place.

Resolving that paradox requires taking a closer look at where confidence is placed, namely in the language and processes of assessment: its bureaucracy. There are rubrics to rate the language/process “correctness” of assessment programs: a list of checkboxes with things like [X] Defined at least three outcomes, [X] Outcomes are measurable, [X] Outcomes relate directly to the next higher level of outcomes at the institution, [X] Outcomes mapped to the curriculum, and so on. There are articles, books, and lectures on how to write outcomes statements for courses, programs, and institutions, how to create curriculum maps, how to create rubrics, and how to organize and evaluate all of this work for each academic program.

The emphasis on form over function extends to the reviews we do of each other’s programs in accreditation work. *Did the program have outcomes? Were they assessed? Were the results used for something?* Everything is checked except whether or not the data are any good and the inferences are reasonably justified.

Most institutions probably have a small number of assessment projects, perhaps in general education, that do get the attention they need to be successful as educational research. But the majority can only pass accreditation reviews through attention blindness induced by a box-checking mentality of correctness.

### **The Future**

In the era of “fake news,” it is imperative that higher education holds itself to a high standard of intellectual honesty. We should follow the lead of academic psychology in a self-examination of our standards of practice. That field is enduring a “reproduction crisis” that calls into question a large amount of peer-reviewed, published research. Relying on small sample sizes is one of the causes (Simmons, Nelson & Simonsohn, 2011).

Imagine if each town and village were required to research and produce its own drugs, and ignore large-scale science-based medical research. That is our current situation with respect to assessment.

By contrast, research in teaching and learning is booming. Look at the proceedings of *Educational Data Mining* for many examples of the creativity and energy being devoted to this research. There is more data available than ever before, computation is cheap, and new methods for visualization and analysis abound.

We can imagine a future where assessment leaders work closely with institutional researchers and scholars to create and share large sets of high-quality data. These might be organized by discipline or at the institutional level to focus on a manageable number of outcomes—not hundreds of them at once. We would work with faculty members to understand and use research findings instead of cajoling them to do paperwork, re-grade papers, and then stare at bar graphs trying to divine meaning. Assessment conferences can be about what we discovered and how faculty are using that information.

One model of that approach is the English composition program at University of South Florida-Tampa, where my colleague Joe Moxley has turned a two-semester writing requirement into a large-scale research program. His work has garnered grant money, attracted dozens of researchers, launched a journal and a conference, and produced a corpus of hundreds of thousands of student papers, peer reviews, rubric ratings, and survey items that is available to researchers. What makes Moxley's program so outstanding is the constant critical attention to the quality of data. The goal is not perfect measurement; the goal is to not fool ourselves.

Other models are possible that would fit different situations and types of institutions. I am continually impressed with intelligence and dedication of people I meet in the assessment field. It is appalling how much of that talent gets wasted filling out checkboxes. If existing assessment resources were redirected to trying to understand student learning, we could revolutionize education in the next ten years.

The assessment profession is now decades old, and it is time that the standards of practice are defined by the community of practitioners who do the job. In collaboration with other stakeholders, the Association for the Assessment of Learning in Higher Education (AALHE) is the logical choice of a body to lead the creation of such standards.

### **Acknowledgements**

I am indebted to Jeff Barbee, Gray Scott, Jane Marie Souza, and Josie Welsh for their excellent advice and feedback while writing this article.

### **References**

- Astin, A.W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Washington, DC: American Council on Education/Oryx Press Series on Higher Education.
- Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning*, 43(1), 22-27.
- Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: American Council on Education/Praeger.
- Bonate, P. L. (2000). *Analysis of pretest-posttest designs*. CRC Press.
- Ewell, P. T. (1991). Chapter 3: To capture the ineffable: New forms of assessment in higher education. *Review of research in education*, 17(1), 75-125.
- 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. (Occasional Paper No. 23). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Hillocks, G. (1986). Research on written composition. *Urbana, IL: National Council of Teachers of English*.
- Kilgo, C. A., Sheets, J. K. E., & Pascarella, E. T. (2015). The link between high-impact practices and student learning: Some longitudinal evidence. *Higher Education*, 69(4), 509-525.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Terenzini, P. T. (1989). Assessment with open eyes: Pitfalls in studying student outcomes. *The Journal of Higher Education*, 60(6), 644-664.

---

David Eubanks is Assistant Vice President for Institutional Effectiveness at Furman University and a member of the AALHE board. He can be reached at [david.eubanks@furman.edu](mailto:david.eubanks@furman.edu).

## BEYOND THE THEATRE OF COMPLIANCE?

By Madeline Murphy

Assessment is often framed, in the words of one *Inside Higher Ed* contributor, as “an act of care.”<sup>1</sup> We certainly *should* care about student learning. Students trust us with their time, money, and effort; the degrees we confer represent a promise that students will take away something of value to them. Maybe I’m lucky, but my experience of almost twenty-five years of teaching English at a suburban community college suggests that teachers really *do* care. My working life has been filled with committed, creative professionals, all of whom are deeply committed to their students and reflect unceasingly on their pedagogy. So why, then, do so many teachers express skepticism about assessment?

One answer, perhaps, is that assessment has become entangled with a vehement, if unfocused, demand for accountability. Ever since the Reagan-era *A Nation At Risk* announced that our high schoolers were headed straight down the drain, and the Japanese would eat our lunch, the public has furrowed its collective brow at educators. A decade ago, the Spellings Commission deemed higher education “increasingly risk-averse, at times self-satisfied, and unduly expensive.”<sup>2</sup> The implication is clear: we are stuck in our ways, too complacent to self-police. But who would deliver the necessary shake-up? This task fell to the accreditation agencies, who are seen, in Washington, less as peer evaluators than as academic watchdogs - the SEC to our Wall Street.

The critics have a point. College costs have risen inexcusably; and we *could* probably use a good overhaul. But when accountability drives assessment practices, institutions tend to go on the defensive, and potentially useful and interesting assessment ideas are set aside in the name of compliance. The result, very often, is an unceasing demand for busywork which faculty both lament and mock. Our accrediting body, for instance, prioritizes quantity of data over quality. Site visitors want to know whether we have assessed all our outcomes, at every level, for everything. The interrogation is absurd. Of course, we’ve assessed student learning in every outcome; after all, every student gets a grade. But warned that course grades don’t count as assessment, teachers scramble to create a carapace of alternative student learning “data” – quizzes, capstone assignments, and surveys – all packaged as rubric scores behind which we can shelter. Meanwhile, administrators purchase expensive programs that permit them to generate four-column reports on all this data. This theatre of compliance provides passing scores for re-accreditation, marketing for the assessment software vendors, and plenty of work for assessment directors, administrators, and faculty. Meanwhile, many of those involved are quietly (or loudly) scratching our heads, thinking, “So, did anybody learn anything?”

Identification of every SLO, PLO, and ILO imaginable is followed by mapping. Curriculum mapping of these student, program, and institutional learning outcomes is another priority for accreditation. Are the outcomes stitched together to form a coherent education? A worthy goal, but, once again, the wrong question. A cohesive education reflects a cohesive academic community, and this begins with the people involved, not the forms they fill out. If we want to promote interdisciplinary coherence, we should make it a matter of professional routine for instructors to sit down with colleagues from other disciplines to work on curriculum, delivery, and collaborative assignments, and, indeed, to argue about different philosophical and pedagogical goals. Instead, we spend hours pontificating which boxes to check. Should we say that Outcome #2 of the Shakespeare class supports Institutional Outcome 3 (communication) or 4 (critical thinking) or 5 (diversity)? Likewise, is that outcome introduced, reinforced, or mastered? Who is monitoring this, and who really cares? A fancy map decorated with credible I, R, M, and X (noting embedded assessment) clears the mapping requirement for reaccreditation, but it rarely does much to strengthen the college as a community.

Following the maps are the artifacts assumed to have captured that which we call learning. Now, whether the content of those artifacts reflects learning attained via the pedagogy of the college instructor, the tutelage of a high school teacher, or random responses of the student the honest professor never really knows. And these data points aren't objective things, like rates of infection or crop yields. Each of these artifacts represents a performance, in which one person fashions a special task for another, with the goal of finding out what that second person knows. So, assessment data must either be drawn from a standardized task (which is pretty narrow) or from whatever tasks instructors have set (which makes generalizations almost meaningless). Nevertheless, some repository, portfolio, or overpriced software package dutifully maintains said artifacts awaiting the next phase of the assessment process: judgment.

Some outcomes can be judged objectively, such as licensure pass or fail, but not all. At our last department retreat, for instance, we all agreed that one paper showed competent use of appositives. But another paper triggered a good deal of argument. As requested by the essay prompt, the writer had summarized Claude Steele's argument in *Whistling Vivaldi* accurately and clearly; but she hadn't commented on it. Had she demonstrated critical thinking? Some teachers felt strongly that she hadn't. Others felt, equally strongly, that a student who had demonstrated a clear understanding of someone else's argument deserved a round of applause, and that maybe we should spend more effort getting students to absorb other people's ideas, and less time pressing them to take some poorly considered and probably insincere position for the sake of looking analytical. In short: What exactly *is* critical thinking?

The last stages of the assessment process combine results and use of results into the sacred closing of the loop. Faculty tend to resist demands for evidence of cut-and-dried "improvements" in student learning. They are not wrong. The language of the typical industrial quality control cycle (analyze data, identify problem, propose solution, re-evaluate data to see if it worked, close the loop) applies very unevenly to academic work. I am not deriding evidence-based, outcome-oriented decisions. A law school or cosmetology program whose students never got their licenses would need to take a long, hard look at what it was doing. Likewise, controlled studies comparing competing pedagogies inform teaching in academic disciplines, including the humanities. In English composition, for instance, the research strongly suggests that what is called "teaching grammar" (i.e., getting students to memorize usage rules or complete exercises) does little to improve student writing. So, we don't do it. By contrast, much research suggests that getting students to combine sentences *does* work, especially in the context of an authentic inquiry-based writing task, so we do a lot of that. If we want to see if an initiative is successful, we set a goal, then look at the data. And yes, data sometimes reveals the unexpected.

But a lot of student learning doesn't fit the "close the loop" model of improvement. The biggest difficulty with teaching is not gathering or analyzing the "data." It is making sure that the data, that is, the actual moments of student performance that do or do not reveal something, are meaningful. And here, one can ask a thousand questions that make the loop model look like a convenient fiction.

For instance, what peer-reviewed evidence justifies current demands for prescriptive statements about what a student is supposed to learn? If a student takes away something we didn't anticipate, have we captured that? And does it matter? Can we agree what critical thinking looks like? Indeed, *should* we? Have we failed if students recognize and grapple with contradictions and disagreements inherent in a Western education? How do we capture learning that occurs as a result of our teaching long after the 17-week course has ended? What if learning presents itself in counterintuitive ways that quality-control models and KPIs of industry fail to detect? When "you have learned something," Mr. Undershaft says in Shaw's play *Major Barbara*, "that always feels, at first, as if you had lost something." Uncertainty, defeat, irresolution: aren't these experiences also something we should equip students to deal with? Doesn't this matter too?

Underpinning so many of these unhelpful assessment practices are assumptions about teaching and learning – assumptions that come to us not from academia, but from business quality control models, and

from a political culture hungry for “results.” Perhaps the best work we can do, in assessment, is to insist on carving out and discussing, out loud, our own ideas about what education is for. That too would be an act of care.

---

Madeleine Murphy (M.Litt, Edinburgh University) is an English teacher at College of San Mateo and can be reached at [murphym@smccd.edu](mailto:murphym@smccd.edu).

# THE CRUCIAL ROLE OF FACULTY DEVELOPMENT IN PROMOTING QUALITY ASSESSMENT PRACTICES

By Glenn Waterbury, Madison Holzman, Beth Perkins, and Allison Ames

## **Introduction & Background**

Assessment is a tool faculty and student affairs professionals can use to paint a picture of student learning, serving as a conduit to highlight strengths, uncover weaknesses, and glean insight into necessary programmatic changes. To realize these benefits fully, programs must implement high-quality assessment processes. Without high-quality assessment, results may be untrustworthy, and the critical questions stakeholders have regarding students' skills and abilities will go unanswered.

At James Madison University (JMU), academic degree programs have been improving the quality of their assessment processes (Rodgers, Grays, Fulcher, & Jurich, 2013). When Rodgers and colleagues asked programs how they were able to improve their assessment processes, one common theme stood out: professional development opportunities offered by the university. That is, faculty within academic degree programs were able to strategically improve their assessment practices because they were provided opportunities to improve their foundational assessment knowledge and skills. If faculty are expected to develop assessment processes that yield useful and meaningful information for their programs, they must be provided with support to build and transform their assessment skillsets. What follows is a description of the assessment-related professional development opportunities available to faculty members. Table 1 provides a brief overview of JMU's support services. For context, JMU is a mid-sized public university in Harrisonburg, Virginia. JMU is home to roughly 20,000 undergraduate students and employs approximately 900 full-time faculty.

Table 1.

### *Outline of Professional Development Opportunities*

Opportunity	Time Frame	Yearly Faculty Participation
1 on 1 Consultation	Year-Round	30 to 40
Formal Workshops	Year-Round	10 to 15
APT Lockdown	Late April	30 to 40
APT Rating	Late July	8 to 10
APT Feedback Reports	October 1st	115 to 125
Assessment 101	Mid-June, Mid-July	30 to 40
Learning Improvement Course	Fall Semester	2 to 4
Assessment Certificate	Four Semesters	0 to 2

## **Consultation**

Of the many assessment related professional development opportunities offered at JMU, Good et al. (2013) found that consultation with an assessment expert was the most consistently helpful endeavor. The Center for Assessment and Research Studies (CARS) is an office dedicated to ensuring that quality assessment is ubiquitous across the JMU campus. Faculty members who work in CARS serve a dual role: professors and assessment practitioners. CARS faculty members' commitment to assessment is often manifested through consultation with their colleagues across campus. CARS is also home to a number of doctoral and master's-level graduate students. These students serve multiple roles: student and graduate assistant, with many serving in Program Assessment Support Services (PASS). PASS works to ensure JMU academic programs have high quality student learning outcomes assessment processes through assessment consulting services, the provision of assessment resources, and the dissemination and promotion of quality assessment practice.

Members of the PASS team consult with program assessment coordinators and committees each year. Individual consultation was reported as the single most helpful assessment resource offered at JMU (Good, et al., 2013). Consultation serves as a time for assessment coordinators, both novice and veteran, to work on and discuss any aspect of their program's assessment process. Additionally, this time spent with a PASS consultant serves as a safe environment for coordinators to ask questions, share their concerns, and cultivate their knowledge of assessment. PASS consultants tailor their assistance and guidance to meet the specific needs of the faculty and program with whom they are working, helping to ensure that faculty feel neither overwhelmed, nor unchallenged.

Along with one on one consultation, PASS provides workshops throughout the year. These can be targeted for one academic program, or provide more general training to accommodate many disciplines. As such, workshops can consist of a broad overview of assessment, or discretely target any aspect of the assessment process, such as writing student learning objectives or developing instruments. Faculty members are given the knowledge and skills to meaningfully and thoughtfully engage with any stage of the assessment cycle and critically evaluate their own assessment processes. This can bolster assessment coordinators' confidence and comfort with assessment, which can lead to a self-sustaining assessment process.

Another PASS-led activity intends to alleviate some of the pressure of writing an annual assessment report. PASS hosts two sessions of a "lockdown," an opportunity for assessment coordinators to allot dedicated, uninterrupted time for working on their assessment report. PASS consultants are on hand to answer questions and provide feedback, but many faculty also use the time to discuss assessment with their colleagues and reflect on past feedback. As with the individual consultations and workshops, the emphasis of the lockdown is on building assessment skills.

## **Participation in Assessment Rating**

The Assessment Progress Template (APT), the annual assessment report, summarizes each program's assessment process, from articulating student learning outcomes, to using results for continuous improvement. Two raters provide feedback on each APT in the form of numeric ratings and diagnostic comments. This rating session, which takes place one week every summer, brings together 15 to 20 graduate students and faculty members to rate the APTs. This rating session also serves as a professional development workshop by providing faculty members and graduate assistants the opportunity to take part in the assessment rating process. Specifically, after participating in the APT rating workshop, raters will be able to:

1. Identify the six major areas in the assessment cycle;

2. Differentiate among beginning, developing, good, and exemplary assessment practice in the 14 subareas;
3. Rate similarly and consistently to fellow raters through calibration;
4. Work with a partner to identify and adjudicate rater discrepancies; and
5. Apply or share what they have learned about the assessment process with the program(s) they belong to or consult with.

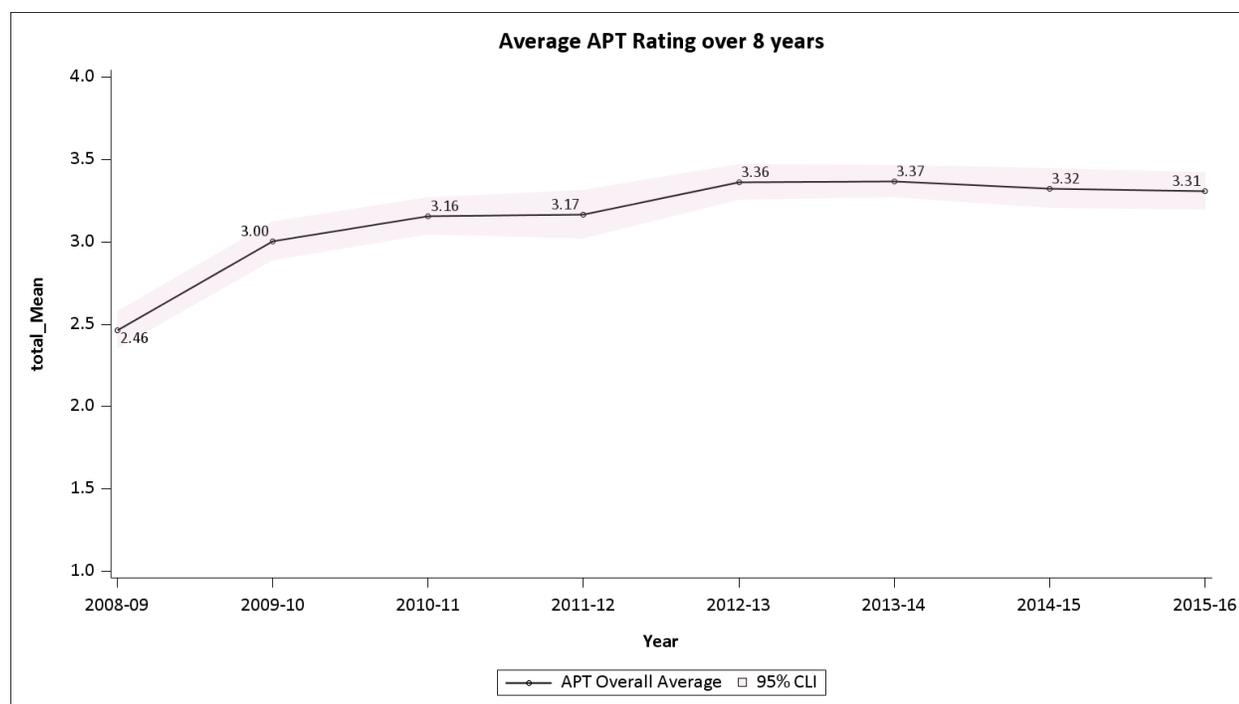
Rater participants are given opportunities to practice providing ratings, write formative comments, and calibrate ratings and comments with their partner. To assist in this process, PASS consultants provide feedback and training throughout the rating and commenting process. This workshop is an opportunity to learn why the quality of assessment matters and how to document assessment practices. Additionally, faculty who participate in APT rater training will have the ability to apply the assessment knowledge they gain to their program's assessment process.

After spending a few days learning about assessment and the rating process, faculty and graduate assistant pairs thoroughly review five to ten APTs and provide ratings for each subcategory of the rubric. Raters also provide comments that are formative in nature, and adjudicate their individual ratings with their rating partner. This process allows faculty raters to review and learn from the assessment practices of many different programs. The ratings and comments provide an overall analysis of the quality of a program's assessment process, highlighting which aspects of assessment are being done well and upon which areas improvement efforts can be focused.

The APT rating process benefits faculty in many ways. Faculty gain a foundation in assessment and learn ways in which to discuss assessment with a wide variety of stakeholders. It is also beneficial for faculty to see how other programs are conducting assessment and making program improvements as a result of assessment. When faculty finish the week of rating, they have made connections with other assessment coordinators, broadening their resources for assistance with assessment.

Figure 1. shows the average APT rating over the past eight years. Of particular note is the rise in scores from the 2008-2009 academic year to the 2009-2010 academic year. This increase in average APT rating may be related to the inclusion of faculty raters in the APT rating process for the first time in the summer of 2009. This finding may suggest that the skills gained by the faculty members who attend this workshop are impacting the assessment processes of academic programs.

Figure 1. Average Assessment Progress Template (APT) Ratings at James Madison University



While only a small percentage of faculty members participate in APT rating each year, every academic program that submitted an APT receives an APT feedback report at the beginning of October. The feedback is intended to be a tool that academic programs can use to strengthen their assessment process. This feedback is specific, targeted, and formative in nature. Faculty can use this feedback to learn about best practices at each step of the assessment process. This feedback can also serve as a springboard for programs to seek additional assessment process support. Moreover, each year, programs take their APT feedback and use it to improve their assessment report and assessment processes. Thus, the feedback has been successful in helping academic programs identify weaknesses and make improvements to their assessment process.

### **Extended Development Opportunities**

The shortest commitment of the extended development opportunities is Assessment 101. Assessment 101 was implemented in Summer 2016 and is a rigorous, week-long workshop designed to transform faculty from “novices” to “intermediates” in assessment. The week is designed to align with each of the critical steps in JMU’s assessment cycle. As such, faculty practice carrying out all steps of the assessment cycle, beginning with articulating student learning objectives, all the way through using results to recommend curricular and pedagogical changes. The key to the week is its interactive nature. By simulating each step of the assessment cycle during Assessment 101, faculty experience the challenges and rewards that accompany assessment. When the week concludes, faculty report that their skillsets are transformed. Moreover, pre- and post- cognitive assessment results suggest that faculty do have improved assessment skills and knowledge after participating in Assessment 101. Faculty also build lasting relationships with a cohort of other Assessment 101 participants whom they can reach out to for questions, ideas, and collaborative projects. A total of 53 individuals representing student affairs, academic affairs, and each academic college have participated since the implementation of Assessment 101.

For individuals who are familiar with JMU’s assessment cycle and have intermediate knowledge of assessment concepts, a graduate-level course in learning improvement is offered. The course is targeted

towards practitioners and allows them to gain an in-depth understanding of how to engage in assessment for learning improvement. Through the course, practitioners learn how to orient their assessment processes toward using results to make curricular and/or programmatic changes, with the end goal of improved student learning. Though the focus of this course is on all aspects of learning improvement, practitioners still gain tips for developing high-quality assessment processes, as high-quality assessment is a stipulation for learning improvement.

Faculty also have the option of the Higher Education Assessment Specialist certificate (<https://www.jmu.edu/outreach/programs/all/assessment/index.shtml>), a long-term professional development commitment. The certificate includes four, three-credit courses taught online and cover topics related to policy, assessment consultation, the foundations of measurement, and instrument design. The wide content of the courses targets a varied audience. For example, college deans can gain the skills and understanding of how to guide and support assessment efforts. Faculty serving as assessment coordinators and department heads will develop professional assessment skills through applied practice and projects.

### **Conclusion**

Put quite simply: faculty are the engine of assessment at JMU, and while there is always room for improvement, many academic programs at JMU are engaging in high-quality assessment practices. We believe that this stems predominantly from the fact that many of our faculty members are knowledgeable about assessment and value its use. The appreciation for and knowledge of assessment has been cultivated over years. For example, our assessment office began working with an assessment coordinator for an arts program several years ago. She was entirely new to assessment and overwhelmed by the process. This was reflected by, among other things, her poor ratings on her assessment reports. Over the past two years, she has participated in several of our faculty development opportunities. She served as an assessment report rater for two consecutive summers. She consulted with assessment experts, both graduate students and faculty, on numerous occasions. She also attended APT lockdown. Through these experiences, and her hard work, she has transformed her knowledge of and comfort with assessment. She has even become a vocal advocate of assessment, as she recently shared her positive experiences at the first annual meeting of assessment coordinators at JMU. She is just one example of the many faculty who have grown as assessment practitioners.

Just as we cannot fairly expect students to acquire knowledge and skills without the requisite opportunities to learn, we cannot expect faculty members to engage in quality assessment without providing them with chances to foster their assessment skills. By providing varied and targeted resources, assessment practitioners at JMU can continue to fine-tune their skills and thoughtfully evaluate their own assessment process.

### **References**

Rodgers, M., Grays, M. P., Fulcher, K., H., & Jurich, D., P. (2013). Improving academic program assessment: A mixed methods study. *Innovative Higher Education*, 38, 383-395.

---

Glenn Waterbury (waterbgt@jmu.edu), Madison Holzman (holzmama@jmu.edu), and Beth Perkins (perkinba@dukes.jmu.edu) are doctoral students at the Center for Assessment and Research Studies at James Madison University. Allison Ames (ames2aj@jmu.edu) is an assessment specialist and assistant professor of psychology at James Madison University.

# THE JOURNEY OF IMPROVEMENT: HOW AN INSTITUTION ENHANCED THE QUALITY OF THEIR ASSESSMENT PROCESS

By Michelle Rogers

Learning outcomes assessment involves gathering information so that evidence-based actions can be taken to support student learning relative to the institutional learning objectives (ILOs). Making these quality improvements requires the intentional collection of meaningful evidence. Sources of evidence might include: clear articulation of the constructs of interest (e.g. ILOs), achievement results from measures that provide insight into student learning relevant to the ILOs, results of indirect measures that provide information about the institutional environment, evidence that all, or at least most programs are engaged in such activities. No single data source allows us to make claims that students are learning. This article describes how our institution made progress in assessment quality, by improving the collection of evidence.

## **Institutional Context**

Des Moines University is a small graduate health sciences university with five masters' programs (Anatomy, Biomedical Sciences, Healthcare Administration, Physician Assistant, and Public Health) and three doctoral programs (Doctor of Osteopathic Medicine, Physical Therapy, and Podiatric Medicine). We are regionally accredited by the Higher Learning Commission (HLC); six of the programs have their own specialized accrediting agencies.

## **Background**

In 2010, we developed five Institutional Learning Outcomes (ILOs) in an effort to unite our eight programs. The five outcomes were designed to represent areas that all programs should easily be able to address regardless of their disciplinary focus. By 2015, we had undergone multiple reviews and realized several critical barriers to engaging in implementing a quality institutional assessment process. Arguably, the greatest barrier was the vagueness of the outcome statements (shown below):

- Demonstrate a knowledge of the science of human health and well-being.
- Manifest dedication to the highest standards of professionalism.
- Value the human experience with sensitivity to individual and cultural differences.
- Display an ability to work collaboratively/interprofessionally.
- Demonstrate an understanding of research methodology and its relationship to critical thinking.

We lacked clear expectations for student learning and institutional assessment. For example, for the outcome related to scientific knowledge, it was unclear whether this was in reference to both the basic and clinical sciences, or just one of these.

Another barrier to assessing our ILOs was unclear standards for sharing assessment results, which contributed to significant variability in the quality of information shared. Each program has the autonomy to decide how many and which measures will be reported. Some programs included samples of their assignments and evaluation instruments (e.g. rubrics) in the assessment reports to provide insight into their process; other programs reported course grades, completion rates, and pass rates on licensure exams. In addition, while some programs had assessments aligned to the ILOs, others did not. Furthermore, if programs were not meeting their targets based on the measures they reported and how they reported them,

figuring out how to take action was difficult. For example, if a program established a target goal that 100% of students achieve a 3.0 grade point average and only 75% of students met that expectation, there was little to go on. In summary, inconsistencies in assessment quality made it difficult to discern what students were learning, and how to make improvements within a program, and by extension, across the university.

We improved our assessment process by clarifying expectations for student learning, identifying gaps in our assessment, and identifying commonalities in how we were measuring the ILOs across programs. The first phase involved the Student Learning Assessment Committee (SLAC) developing performance indicators using the process shown below. These indicators would serve as examples of evidence to determine if students are achieving the ILOs.

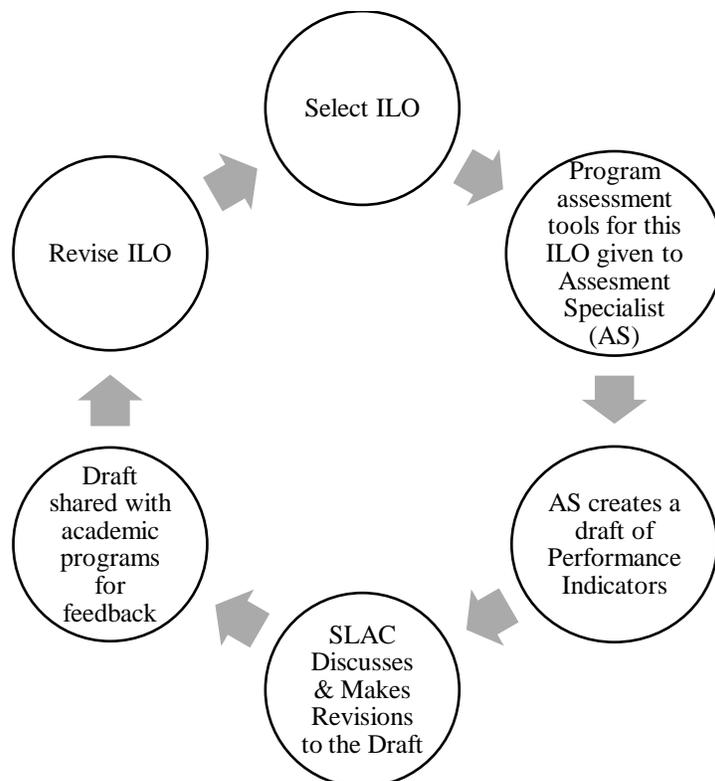


FIGURE 1

It was important that the Performance Indicators (PIs) came from existing assessments, reflecting reasonable expectations of what students should demonstrate. We wanted to avoid “reinventing the wheel” and instead, leveraged what we already had. The indicators were developed by reviewing existing co-curricular and curricular assessment tools and documenting commonalities across programs’ rubrics, checklists, tests, quizzes, surveys, etc. PIs were grouped by similarities, referred to as dimensions. The example provided below demonstrates how we moved from having a very broad professionalism statement, to providing dimensions specific enough to assist us in improving the quality of assessment for the professionalism objective.

Original Statement: “Graduates will manifest dedication to the highest standards of professionalism”

The newly developed framework provides guidance on where to focus assessment; however, it maintains the programs’ autonomy with respect to how to implement assessment of the ILOs and their associated PLOs. (See figure 1).

<b>Revised Wording:</b> Graduates will demonstrate ethical and professional behavior consistent with standards of the profession and the DMU community.	
<i>Dimension 1:</i> Accountability	<ol style="list-style-type: none"> <li>1. Punctual</li> <li>2. Give adequate notice and explanation for absenteeism or if unable to meet deadline</li> <li>3. Make suitable arrangements if unable to fulfill professional duties</li> <li>4. Prepared to perform professional duties</li> <li>5. Follow through on professional commitments</li> <li>6. Accept responsibility for learning successes and failures</li> </ol>
<i>Dimension 2:</i> Communication & Professional Interactions	<ol style="list-style-type: none"> <li>1. Use language and communication styles appropriate to the context and audience</li> <li>2. Establish positive interpersonal relationships with others</li> <li>3. Communicate information in a clear and organized manner</li> </ol>
<i>Dimension 3:</i> Professional Responsibilities	<ol style="list-style-type: none"> <li>1. Maintain appearance suited to work environment</li> <li>2. Promote and build awareness of the profession</li> <li>3. Carry out professional duties in a safe, respectful, and ethical manner</li> <li>4. Follow legal and ethical guidelines applicable to the profession</li> </ol>
<i>Dimension 4:</i> Self-Directed Learning	<ol style="list-style-type: none"> <li>1. Identify areas of self-improvement</li> <li>2. Seek out help when needed</li> <li>3. Set goals independently</li> <li>4. Use information to improve learning</li> </ol>
**All programs are <b>NOT</b> assessing all performance indicators. But each performance indicator is linked back to at least one program's assessment. Most are linked to several programs' assessments.	

After the PIs and dimensions were developed, the committee proposed revisions to the ILO statements to improve clarity. The second phase involved sharing these PIs, dimensions, and new ILO statements across the institution to get feedback and receive the approval of the faculty.

While developing a shared mental model of the meaning of the outcomes was a significant aspect of the project, it was not the only element. We identified other gaps in our assessment process and made improvements. We learned that (1) most assessment tools for interprofessionalism were surveys rather than direct measures of student learning, (2) multiple programs lacked quality measures for the “value the human experience” objective, and (3) some of the measures selected for the objectives did not align with them. Several projects were initiated to improve the assessment quality of the ILOs assessed. Some projects entailed revising an assignment and/or assessment so that they appeared more relevant to the construct of interest. Some examples include the following:

- The Master's in Public Health Program redesigned an assignment and developed a rubric to align interprofessionalism and the “value the human experience” objectives.
- The Master's in Healthcare Administration Program used the interprofessionalism/collaboration performance indicators to inform the redesign of a leadership assignment and assessment to emphasize interprofessional education. Previously, the grading rubric emphasized writing skills.
- The Master's in Anatomy Program aligned one of its assignments to the “value the human experience” objective and redesigned an assessment to better emphasize the collaboration objective.

- Several faculty across programs developed new assessments based on the ILO performance indicators for professionalism.

We also made several major revisions to our assessment report template to improve the consistency of assessment of student learning and context for the institutional activities taking place to support student development.

1. We added a column in the report that asks programs to include what curricular or co-curricular experience is associated with the assessment used and when in the curriculum/co-curriculum do students engage in this experience. (institutional activities)
2. We added a new column in the report that asks for an explanation of how the assessment tools selected link back to the outcome (performance indicators).
3. In the findings section of the report, we provided more direction by requiring programs to provide a brief evidence-based narrative describing student strengths and opportunities for improvement based on the ILO dimensions.
4. We requested that programs include at least one assessment tool in the appendix of the report so that we can see how students are being evaluated. This also allows us to examine its alignment with the learning outcome.

In summary, our work promoted meaningful conversations about expectations for student learning and improvements in teaching, learning and assessment activities. Admittedly, we are far from having a perfect institutional assessment process. However, we now have a better idea of where students are headed and how we expect them to get there through the program curricula. Next steps entail evaluating how changes to the process resulted in improved understanding of how students are learning relative to the ILOs across programs and if the changes enhanced the quality of our conversations about student learning.

---

Michelle Rogers is Assessment Specialist at Des Moines University. She can be reached at [michelle.rogers@dmu.edu](mailto:michelle.rogers@dmu.edu).

## ASSESSING IN THE RED: HOW TO KEEP QUALITY-OF-ASSESSMENT PROSPERING

*By Yen M. To & Kipton D. Smilie*

Assessment in higher education is often a multidimensional approach used by administrators and faculty to identify opportunities for quality improvement of their programs and services. A strength of assessment is its ability to be custom-designed to fit within the framework of any given institution. Interestingly, this adaptive ability is also a limitation. While there is a common consensus about good assessment practices (e.g., ongoing, longitudinal, benchmarked, etc.) (Alstete, 1995; Astin et al., 1992), there are no hard and fast set rules on how to make assessment work perfectly across a diversity of environments or programs. In higher education, assessment is not prescriptive like degree programs nor is it elective like co-curricular programs. What may work in one department may not translate well to another. For example, our Education department uses subject content and performance test scores, course syllabi, and interviews, to demonstrate evidence for meeting their professional standards for teacher accreditation. Nursing and Business, departments seeking discipline-based accreditation, are also responsible for completing lengthy narratives; however, the specific requirements for evidence of student learning vary and the final reports are disparate and incomparable. Compounded upon that, departments not seeking discipline-based accreditation (e.g., Philosophy, Sociology, Art, and Math), formalized assessment is absent or lacks a common structure. At the institutional level, assembling the assessment of student learning across these departments into a uniform approach is extremely challenging, especially for those unfamiliar with assessment. This struggle with multiplicity, and lack of uniformity, also occurs across assessment-focused listservs where experienced assessment professionals frequently request examples of assessment done in areas such as general studies, career development, student affairs, etc. (ASSESS, 2005-2017).

Given the various ways assessment occurs at institutions, administrators and faculty alike are constantly confronted with the challenge of determining the quality of their assessment practices. If assessment leaders can show stakeholders that their assessment efforts have integrity, then stakeholder confidence in the data, subsequent findings, and “closing the loop” recommendations grow to become more meaningful. However, at institutions where resources are lean, assessment staff fall below capacity, limited funds restrict professional development in assessment, or there is a lack of technology-supported assessment products, the challenge of establishing quality assessment standards for each discipline becomes overwhelming.

In theory, institutions are conducting assessment correctly when they routinely produce meaningful results that indicate if goals have been met, and how action can be taken to improve outcomes. This also means that institutions have quality-of-assessment (reliability, validity) established, and both assessment processes and assessment data meet these quality checks. However, often in practice and especially with institutions that are new to assessment, and perhaps also lack resources, quality-of-assessment is a loose post hoc conclusion. At these institutions, the bulk of efforts are focused on establishing a feasible process and identifying appropriate sources of evidence rather than using assessment to determine if expected quality is being met. The appropriate steps to determine true quality of a program has not been reached at these institutions due to the focus being on the process of assessment rather than assessment itself.

## **An Economical Approach**

While establishing validity and reliability are critical to demonstrating quality, institutions waiting to mature in assessment or limited in assessment resources can employ other strategies to inform upon quality. This was the position of Missouri Western State University (MWSU), a small regional open-admissions institution with finite assessment resources and a culture of assessment that has yet to be embraced. Their strategy is to apply the main components of quality-of-data checks (e.g., completeness, uniqueness, timeliness, accuracy, and consistency) to multiple levels of their institutional-wide effort. At MWSU, quality-of-data checks occur during data selection and collection (micro-level) by faculty and are embedded in the process of planning and reporting (macro-level) by administrators. By using quality checks, faculty and administrators can gain confidence that their engagement in assessment is worthwhile. For institutions limited in resources for assessment, MWSU's alternative approach of using quality-of-data to inform quality-of-assessment (e.g., reliability and validity), may be appealing as it minimizes the significant cost of time and money and can still lead to efficient decision-making processes.

## **Quality-of-Data Checks**

In the discipline of data management, quality-of-data refers to the examination of an attribute or feature of data to understand its quality (Askham et al., 2013). In the context of higher education assessment, data quality is often a perception of the data's ability to inform upon assessment outcomes, which can be a meaningful service gauge for institutions. Of the possible quality-of-data attributes in data management, those readily applied to assessment of academic programs and student learning outcomes (SLOs) as quality checks include: 1) Completeness, 2) Uniqueness, 3) Timeliness, 4) Accuracy, and 5) Consistency (Askham et al., 2013).

**Completeness.** Completeness refers to the maximum amount of data that is captured for a given purpose. At MWSU, administrators strive for 100% assessment participation from all academic programs. The completeness rate is determined from the number of programs who submit both their annual assessment plans and reports within a week of the campus-wide deadline. Another method to capture completeness is to determine if all program SLOs are being assessed within a specified time range (e.g., three-year rotating schedule). When making macro-level recommendations to the institution on general competencies or strategies, completeness of data allows administrators the confidence that information from across most, if not all, academic programs were included. For instance, if the institution implemented interventions related to enhancing quantitative literacy, it would be necessary to examine comprehensively this SLO across all programs that teach the content.

At the micro-level, program faculty provide and confirm that data for a specific measure is collected from all students participating in the assignment/activity and that the selected measures are from all core offerings. By striving for completeness, faculty created a comprehensive program map identifying where SLOs are embedded in courses and activities. This quality-of-data check for completeness is important because it ensures a comprehensive examination of the SLO and greater competence in resulting decisions.

**Uniqueness.** The uniqueness of data refers to eliminating redundancy in the data and limiting data-gathering to distinct sources. The process to ensure uniqueness is the same at both the macro- and micro-levels; however, the data sources used vary with the macro- data gathered from programs and the micro-data gathered from students. As a macro-level quality check, administrators request that data gathered on each SLO be captured from multiple distinct sources. Specifically, it is required that the evidence collected comes from a minimum of two measures gathered from different courses or experiential activities. This

requirement for uniqueness at the macro-level allows information on outcomes to be gathered across the program rather than limited to a handful of courses, artifacts, or activities. Faculty attain uniqueness at the micro-level in a similar manner when f measures vary across courses (not just programs) as well as across instructors. No one faculty member should be solely responsible for delivering the content of any one SLO. However, in MWSU's smaller programs, this option is improbable. Given this impracticality, faculty are encouraged to measure the SLOs from different perspectives (i.e., assignment/project formats) to ensure that the SLO is being assessed through multiple modes. For example, the SLO of ethical reasoning can be captured from a written assignment, but it can also be examined as a case study, oral presentation, treatment plan, etc. Likewise, at the macro-level, one program should not be solely responsible for delivery of a specific SLO. Administrators are expected to measure SLOs from multiple perspectives incorporating the academic, co-curricular, student affairs, and support programs.

**Timeliness.** Timeliness refers to data that is collected and reported in a recent manner. At MWSU, this quality check is similar across macro- and micro- levels and is dependent upon the assessment cycle. Timeliness ensures that the data used to form recommendations or inform modern practices is still relevant to the SLO under study. Using recent data to inform decisions safeguards the process and programs from inaccurate and dated assessment applications. For example, when deciding how many employers to invite to a career fair, event staff should not rely on how many students attended three years ago but rather the number who attended last semester. Timeliness applies to data collection and the process. Not only should the data collected be current, but also assessment plans should be routinely updated and reported on a continual basis (i.e., schedule). The updated plans allow changes (flexibility) in both the measures of the SLOs and the SLOs themselves, in order to match program realities and new directions in the program field. Administrators or faculty can determine the appropriate limitations and acceptable time constraints on the timeliness of data/process beforehand, though data collection and brief analysis each semester can help ensure quality. Timeliness as a quality check is also related to the accuracy attribute, as accuracy in almost all cases decays over time (Askham et al., 2013).

**Accuracy.** Accuracy of data is the extent to which the data correctly represents the SLO assessed. Think of this as the appropriateness (e.g., fit) of the selected measure (artifact) to the rubric or the alignment of the rubric to the articulated proficiency levels of the SLO. This is done at the macro-level when administrators look at the fit of specific program SLOs to MWSU's goals or competencies. With faculty, the accuracy of data is determined from the selection of measure type (direct/indirect) for the SLOs. Faculty are content experts and therefore they provide program consensus on what to identify as an appropriate measure for each SLO. Measures that are direct, acquired from the source (i.e., student) regarding a specified demonstrable behavior (e.g., SLO), are stronger and more fitting accounts of the outcome than indirect measures, such as perception surveys, or other third-party accounts. Justification statements for target levels are also more meaningful at this micro-level, as faculty have a clearer threshold (e.g., national benchmarks, baseline data) and rationale for determining SLO success.

**Consistency.** Consistency is roughly defined as the absence of difference in the data when comparing more than two collections (Askham et al., 2013). In the example of MWSU, administrators compare the consistency of data over two or more assessment cycles. The emphasis is on regularly collecting data to provide trend information on SLOs. Recommendations from assessment based on consistent and reoccurring trends in the data are more trustworthy than those made from brief snapshots of the data. At both macro- and micro-levels, consistency in the form of data trend reliance is a quality check that is essential to making data-informed decisions that are more attuned to the institution's and programs' actual

functioning. The comparison of trend data to identified benchmarks is more accurate and appropriate to utilize for decision making.

### **Balancing Data Quality with the Institution's Culture**

In the quest to address quality-of-assessment, assessment leaders should be mindful of the cost to faculty and staff. Extra requirements related to quality checks equates to extra work for those at the micro-level. Quality checks may consist of external vetting of processes, internal and external scoring of student assignments to measure SLOs, identifying external sources for program-level benchmarking, assessment software for data management, or data triangulation. For example, in the Education Department, to assess a student's professional disposition (SLO) faculty must coordinate the collection and integration of data from the student, state exam, instructor, university supervisor, mentor teacher, and principal. The workload for quality checking this single SLO places high demand on faculty and program resources. At MWSU, faculty participate in assessment is a more intimate, hands-on process, in comparison to administrators who engage in assessment from a coordination and management perspective. However, but both parties are responsible for maintaining assessment integrity. Faculty work with assessment more directly, potentially leading to differences in perceptions of data quality and assessment culture as compared to their administrator counterparts. Performing the data quality checks but not communicating them across the macro- and micro-levels may lead to disagreement and "off-boarding" (i.e. resentment) of assessment.

Faculty are often charged with collecting and interpreting assessment data through multiple means and for various entities: program reviews, accreditation applications (state, national, professional), and state and federal reporting requirements. Therefore, faculty are informed on the quality of data collected from a variety of different instruments and measures. Faculty's knowledge of students, courses, and programs can lead to a better sense of completeness of the data. Having the first-hand experience of working with students and understanding their struggles provides them with insight that collected data alone could overlook. These faculty insights can provide context to outliers in the data that may otherwise skew analysis and interpretation of the data.

By implementing quality checks, assessment leaders have provided faculty and administrators a common method of inquiry to inform assessment integrity at MWSU. Assessment equips faculty with the ability to identify contributing factors and underlying causes of data findings, and administrators have the confidence in the data to inform institutional quality initiatives. However, the finite resources allocated to this quality-of-data approach (an assessment practice) should be balanced with those allocated to fostering an assessment culture (assessment spirit). Heavy emphasis on quality checks may lead to concerns regarding faculty intimacy with the data and erosion of stakeholder trust in data-based institutional initiatives. Likewise, the distance from which administrators operate may lead faculty to resist administrator-led initiatives they feel are too remote and detached.

Quality-of-data as indicators of quality-of-assessment can be a useful approach to institutions similar to MWSU as it involves an embedded integrity review of both the assessment process and data collection. Those intending to use this quality-of-data approach should be mindful that quality checks should be performed by both administrator and faculty groups. Similar to promoting a culture of assessment, it is crucial to have both perspectives involved in the assessment process in order for it to be successful. The quality-of-data approach may help to facilitate a cooperative assessment culture where both groups benefit from working together because they have a common understanding of the attributes and methods used to establish integrity, even when resources are limited.

## References

- Alstete, J. W. (1995). *Benchmarking in Higher Education: Adapting best practices to improve quality*. (ASHE-ERIC Higher Education Report No. 5). Washington, DC: Office of Educational Research and Improvement (ED).
- Askham, N., Cook, D., Doyle, M., Fereday, H., Gibson, M., Landbeck, U., ... Schwarzenbach, J. (2013). The six primary dimensions for data quality assessment: Defining data quality dimensions (October Report). Rüdgeaway, Bristol: International Data Management Association UK [DAMA UK]. Retrieved from <http://www.damauk.org/RWFilePub.php?&cat=403&dx=2&ob=3&rpn=catviewleafpublic403&id=106193>
- Assessment in Higher Education [ASSESS]. (ca. 2005-2017). Archives of the ASSESS Listserv, University of Kentucky, Louisville, KY. Retrieved from <http://lsv.uky.edu/scripts/wa.exe?A0=ASSESS>
- Astin, A. W., Banta, T. W., Cross, K. P, El-Khawas, E., Ewell, P. T., Hutchings, P., ... Wright, B. D. (1992). *Principles of good practice for assessing student learning*. American Association for Higher Education (AAHE).

---

Yen M. To is Director of Assessment and Institutional Research at Missouri Western State University. He can be reached at [yto@missouriwestern.edu](mailto:yto@missouriwestern.edu). Kipton D. Smilie is Assistant Professor of Education at Missouri Western State University. He can be reached at [ksmilie@missouriwestern.edu](mailto:ksmilie@missouriwestern.edu).

# USING FACTOR ANALYSIS TO TEST A MEASURE OF STUDENT METACOGNITIVE ABILITY RELATED TO CRITICAL THINKING AND INTELLECTUAL HUMILITY

*By Jeff Roberts, David E. Wright, and Glenn M. Sanford*

Locally-developed measures represent great tools for institutions to use in assessing student outcomes. Such measures can be easy to administer, can be cost-effective, and can provide meaningful data for improving student learning. However, many institutions struggle with questions surrounding the quality of their locally-developed assessments. Are their instruments reliable? Are their instruments valid? Can the data generated from these instruments be trusted to drive change and improvement? The good news for faculty, staff, and assessment professionals is that there are steps they can take to address these concerns and help to ensure the validity and reliability of their processes. This article describes the development and testing of a novel research instrument of students' attitudes and abilities relating to critical thinking, metacognition, and intellectual humility. Using a \$1,000 assessment grant from Sam Houston State University (SHSU), Dr. Glenn Sanford and Dr. David Wright devised the early drafts of the instruments, collaborated with colleagues, and joined with Mr. Jeff Roberts, Director of Assessment at SHSU, to develop and to test this new instrument. What follows is a description of the development of the resulting research instrument, results from the factor analysis and reliability testing of that instrument, and an overview of how those results have been used to make further instrument improvements.

## **Metacognition, Intellectual Humility, and Critical Thinking**

Ensuring that your assessment measures are solidly grounded in theory is important step that can help address many questions regarding your instruments' validity. At the outset of this project, Wright and Sanford ensured that the basis for their new instrument was well-grounded in research around critical thinking, intellectual humility, and metacognition. By grounding the instrument in theory, Wright and Sanford helped to improve the construct validity of their instrument (Johnson & Christensen, 2012). Skill in critical thinking requires a variety of competencies, not least of which includes metacognition about one's knowledge and abilities as a critical thinker. Whether one conceives of critical thinking as a skill (Missimer, 1990; 1995) or some combination of skill and dispositions (Ennis, 2015), adept critical thinking requires students to possess a keen awareness of their cognitive strengths and weaknesses. In short, good critical thinkers know what they know as well as what they do not. Something similar might be said about intellectually humble thinkers: they do not overestimate their knowledge and cognitive skills.

Recent research has explored the connections between the nature of intellectual humility and its relation to metacognition (Church & Samuelson, 2017), but less attention has been given to the intersections of intellectual humility, metacognition, and critical thinking. Kruger and Dunning (1999) discussed how those who were least skilled in various intellectual domains were also unaware of their lack of skill. Moreover, the same research suggests that those who are among the most competent at a particular skill slightly underestimate their competence (Kruger & Dunning, 1999). This effect, commonly known as the Dunning-Kruger effect, has been observed across intellectual domains including logical reasoning, memory, and interviewing skills (Dunning, 2005; Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2007). The Dunning-Kruger effect has not been widely tested with respect to critical thinking skills, although there have been some preliminary efforts in this direction (Bensley et al., 2016; Bensley & Spero, 2014; Ku & Ho, 2010; Magno 2010).

## **Development of the Original Instrument**

A \$1,000 assessment grant from the Office of Academic Planning and Assessment at SHSU was crucial for developing the draft instrument. The grant provided resources to host luncheons where the philosophy faculty critiqued early versions of the questions in order to align them with the concepts tested on the Texas Assessment of Critical Thinking Skills (TACTS). It also funded Wright's travel to an intellectual humility conference to share early thoughts and get feedback from other experts in the field. The questions were then submitted to Kathryn Wright, a composition instructor at SHSU, who offered further suggestions on improving question wording to ensure that students could understand what was being asked without loss to necessary theoretical complexity. Each of these steps helped improve both the construct (Johnson and Christensen, 2012) and content (Banta & Palomba, 2015) validity of the instrument, which serves as an important mark of quality for an assessment measure.

Prior to the start of this project all Critical Thinking courses at SHSU were already using pre-post administrations of the TACTS, a 35 question validated critical thinking test (Fair, Miller, Muehsam, & McCoy, 2010), for annual programmatic assessment. To test for the Dunning-Kruger effect, students were asked to estimate the number of TACTS questions they answered correctly (out of 35), give a percentile estimate of how their score compared with their peers who took the TACTS test in that session (1-100), and give a percentile estimate of how their score compared with all students who had ever taken the TACTS test. As with previous research on the Dunning-Kruger effect, this provided a way to compare students' predicted and actual competences related to critical thinking.

The Dunning-Kruger-related metacognitive questions only evaluate students' ability to postdict their scores and make peer comparisons. Though important, metacognition about critical thinking involves the thinker being aware of many other aspects of their thinking including, but not limited to, their ability to recognize informal fallacies, to identify logical relationships between pieces of information, and to distinguish between causal relationships and mere correlations. To create a measure of these abilities, Sanford and Wright analyzed the TACTS in terms of which particular critical thinking skills were involved in correctly answering individual questions. Subsequently, they investigated the alignment between TACTS questions and the skills SHSU professors focused on in their Critical Thinking courses. Ultimately, twelve skills were identified. Sanford and Wright then devised questions aimed at describing the relevant skills in ways that students new to critical thinking could comprehend, since the questions were to be used in conjunction with the existing TACTS instrument. These represented important steps to help ensure the content validity of the instrument (Banta & Palomba, 2015).

Reviewing recent studies of intellectual humility, the two Sanford and Wright found provided the best framework were Roberts and Wood (2007) and Whitcomb, Battaly, Baehr, and Howard-Synder (2015). Roberts and Wood (2007) argued that the intellectually humble person pursues knowledge and wisdom while avoiding vanity and arrogance. Whitcomb et al. (2015) identified intellectual humility as being when someone is aware of their epistemic shortcomings and adopts appropriate attitudes and behaviors regarding these shortcomings (e.g., avoiding anger when someone points out an intellectually humble person's cognitive error). Rather than choosing between these two competing accounts, Sanford and Wright viewed both accounts as containing key insights into the nature of intellectual humility and developed a series of questions reflecting both theories.

## **Methods of Instrument Testing**

Following the development of the original instrument, Sanford and Wright partnered with Roberts to develop a plan for further testing and research. The researchers had designed each of the 21 questions included within the instrument in such a way that they expected to fall within one of those two expected factors; however, as the instrument was locally-developed and unvalidated, further testing was needed to

confirm our expectations. Therefore, a restricted factor analysis was used to confirm the presence of the two hypothetical factors included within the instrument, (a) *student metacognitive ability related to critical thinking*, and (b) *student intellectual humility*, and to determine the relative fit of each of the instruments questions within these two factors. According to Johnson and Christensen (2012), factor analysis is “a useful technique for examining the internal structure of tests” (p. 146). Internal consistency analyses were then calculated to determine the reliability of the identified factors. A total of 259 students, from the fall 2016 and spring 2017 semesters, were included in this analysis.

### **Findings**

A varimax factor analysis was conducted using IBM SPSS Statistics v. 22 to determine whether the two potential factors within the instrument each met the eigenvalue-greater-than-one-rule (Kaiser, 1958). This also allowed for the researchers to determine the relative fit of each question within the two factors by determining whether each met the recommended correlational cutoff of .3 (Lambert & Durand, 1975). Questions 1-13 all met the required cutoff for inclusion within the potential factor of *student metacognitive ability related to critical thinking*, while Questions 15-21 all met the required cutoff for inclusion within the potential factor of *student intellectual humility*. Readers are directed to Table 1 for the factor loadings for each of the test items. Of the 21 questions, only Question 14: “It would not bother me very much if I found out that I was very poor at critical thinking” did not meet the recommended cutoff of .3 for inclusion within either factor (Lambert & Durand, 1975).

Table 1

#### *Factor Analysis of Student Assessment of Metacognitive Ability Relating to Critical Thinking and Intellectual Humility*

Question Number	<i>Student Metacognitive Ability Related to Critical Thinking</i>	<i>Student Intellectual Humility</i>
Question 1	<b>.87</b>	.20
Question 2	<b>.83</b>	.22
Question 3	<b>.85</b>	.18
Question 4	<b>.79</b>	.06
Question 5	<b>.85</b>	.18
Question 6	<b>.66</b>	-.14
Question 7	<b>.83</b>	-.00
Question 8	<b>.81</b>	.08
Question 9	<b>.54</b>	-.11
Question 10	<b>.77</b>	-.03
Question 11	<b>.84</b>	.18
Question 12	<b>.70</b>	.19
Question 13	<b>.78</b>	.20
Question 14	-.17	-.00
Question 15	-.01	<b>.75</b>
Question 16	-.10	<b>.63</b>
Question 17	.28	<b>.60</b>
Question 18	.45	<b>.64</b>
Question 19	.22	<b>.77</b>
Question 20	.03	<b>.72</b>
Question 21	-.07	<b>.72</b>

Following this initial factor analysis, internal consistency analyses were conducted of the thirteen questions (i.e., questions 1-13) identified as being potentially part of the first factor, *student metacognitive ability related to critical thinking* and of the seven questions (i.e., questions 15- 21) identified as being potentially part of the second factor, *student intellectual humility*. The questions included within both factors were determined to be internally consistent (i.e., reliable) using the standard of a .70 or greater Cronbach's coefficient alpha (Nunnally, 1978). The Cronbach's coefficient alpha for *student metacognitive ability related to critical thinking* was .95, indicating excellent internal reliability. The Cronbach's coefficient alpha for *student intellectual humility* was .83, indicating good internal reliability.

For each question, the students were asked to rate their level of agreement using a 5-point Likert scale, with 1 equaling "Strongly Agree" and 5 equaling "Strongly Disagree." Prior to analysis, student responses to questions 1-13 were reverse scored so that a higher question score would indicate greater confidence with that particular critical thinking skill. Questions 15-21 were written in such a way that the higher the student score already indicated a greater level of intellectual humility without the need for reverse scoring. Readers are directed to Tables 2 and 3 for descriptive statistics for the questions comprising both factors.

Table 2

*Descriptive Statistics for Questions Identified Within Student Metacognitive Ability Related to Critical Thinking Factor*

Question Number	<i>m</i>	<i>SD</i>
Question 1	3.46	1.34
Question 2	3.50	1.30
Question 3	3.43	1.34
Question 4	3.35	1.14
Question 5	3.39	1.23
Question 6	3.23	1.11
Question 7	3.31	1.13
Question 8	3.36	1.17
Question 9	3.02	1.04
Question 10	3.14	1.14
Question 11	3.45	1.28
Question 12	3.46	1.27
Question 13	3.49	1.38

*Note.* The number of students who completed this instrument was 242.

Table 3

*Descriptive Statistics for Questions Identified Within Student Intellectual Humility Factor*

Question Number	<i>m</i>	<i>SD</i>
Question 15	3.15	1.37
Question 16	3.18	1.16
Question 17	3.44	1.44
Question 18	3.56	1.50
Question 19	3.20	1.26
Question 20	3.17	1.34
Question 21	3.10	1.27

*Note.* The number of students who completed this instrument was 255.

### **Discussion**

The findings from both the confirmatory factor analysis and internal consistency analyses have proved very useful as we have moved forward with the research using this instrument. First, it confirmed the existence of the two factors, *student metacognitive ability related to critical thinking* and *student intellectual humility*, within the instrument. It also helped identify that question 14, which had originally been designed to fit within the *student intellectual humility* factor, was actually an outlier that needed to be removed from further analysis. Finally, the positive findings from the internal consistency analysis has given us confidence to combine the student scores within each factor to form reliable scale scores that can then be used for further research.

The results have also helped us improve the quality of the instrument. Sanford and Wright were surprised that question 14 did not fit into the *student intellectual humility* factor and debated whether to drop the question, modify the question, or add other questions related to the concept purportedly being tested by the question. Ultimately, it was decided that the question should be modified for the next version of the instrument. Because question 14 seemed to be asking about an important part of intellectual humility, our objective in modifying the question is to determine whether this was an issue with our question, the underlying conception of intellectual humility, or some combination of both. The other major change to the instrument was to strip out all of the questions from the TACTS that were not directly linked to questions in the metacognitive instrument (with permission from its original authors). This allowed for the creation of a single instrument with a much shorter administration time, and should help to remove noise in the pre-post analysis going forward caused by students answering questions that were not related to the course content.

### **Conclusion**

This article has attempted to show how quality of assessment questions can, in part, be answered by examining instrument reliability and validity. Easy steps for ensuring instrument validity include solidly grounding the instrument in theory (i.e., construct validity; Johnson & Christensen, 2012), and making sure that the design and content of the instrument clearly match to your expected student learning outcomes

(i.e., content validity, Banta & Palomba, 2015). Something as simple as a factor analysis can help inform whether your instrument is, in fact, measuring what it is supposed to measure. Furthermore, internal consistency analysis can then tell you if your instruments are providing reliable data. When combined, these reliability and validity efforts can help address many questions surrounding the quality of an institution's assessment measures, particularly when those measures were locally-developed and were previously untested. Faculty, staff, and assessment professionals should all be encouraged to explore such efforts with their own assessment activities.

## References

- Banta, T. W., & Palomba, C. A. (2015). *Assessment essentials: Planning implementing, and improving assessment in higher education* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Bensley, D. A., Rainey, C., Murtagh, M. P., Flinn, J. A., Maschiocchi, C., Bernhardt, P. C., & Kuehne, S. (2016). Closing the assessment loop on critical thinking: The challenges of multidimensional testing and low test-taking motivation. *Thinking Skills and Creativity, 21*, 158-168. doi:10.1016/j.tsc.2016.06.006
- Bensley D. A., Spero R. A. (2014). Improving critical thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity, 12*, 55–68. doi: 10.1016/j.tsc.2014.02.001
- Church, I. M., & Samuelson, P. L. (2017). *Intellectual Humility: An Introduction to the Philosophy and Science*. London, UK: Bloomsbury.
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press. doi:10.4324/9780203337998
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*, 98-121. doi:10.1016/j.obhdp.2007.05.002
- Ennis, R. (2015). Critical thinking: A streamlined conception. In Ma. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 31-48). New York: Palgrave Macmillan. doi:10.1057/9781137378057\_2
- Fair, F., Miller, J., Muehsam, V, McCoy (Elliot), W., (2010). TACTS™: Developing a new critical thinking assessment instrument. *Inquiry: Critical Thinking Across the Disciplines, 25*, 37-41. doi: 10.5840/inquiryctnews20102521
- Johnson, R. B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). Los Angeles, CA: Sage.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200. doi:10.1007/BF02289233
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121-1134. doi: 10.1037/0022-3514.77.6.1121
- Ku, K.Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition Learning, 5*, 251-267. doi: 10.1007/s11409-010-9060-6
- Lambert, Z. V., & Durand, R. M. (1975). Some precautions in using canonical analysis. *Journal of Market Research, XII*, 468-475. doi:10.2307/3151100

- Magno, C. (2010). The role of metacognition in developing critical thinking. *Metacognition and Learning*, 5, 137-156. doi:10.1007/s11409-010-9054-4
- Missimer, C. (1990). Perhaps by skill alone. *Informal Logic*, 12, 145–153. doi:10.22329/il.v12i3.2610
- Missimer, C. (1995). Where's the evidence? *Inquiry: Critical Thinking Across the Disciplines* 14, 1–18. doi:10.5840/inquiryctnews19951441
- Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford University Press. doi:10.1093/acprof:oso/9780199283675.001.0001
- Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2017). Intellectual humility: Owning our limitations. *Philosophy and Phenomenological Research*, 94, 509-539. doi:10.1111/phpr.12228

---

Jeff Roberts is Director of Assessment at Sam Houston State University. He can be reached at jeff.roberts@shsu.edu. David E. Wright is Visiting Assistant Professor of philosophy at Sam Houston State University. He can be reached at dew019@shsu.edu. Glenn M. Sanford is Associate Dean of the College of Humanities and Social Sciences and a professor of philosophy at Sam Houston State University. He can be reached at sanford@shsu.edu.

## HOW DO PROS AT LARGE, DECENTRALIZED INSTITUTIONS MANAGE OUTCOMES DATA?

*Josie Welsh, Interviewer*

How do assessment professionals at large institutions manage outcomes data in a highly decentralized context, and how do leaders at these institutions manage concerns over data quality?

I interviewed two professionals with substantive experience by phone and email. **Stephen Hundley** is Senior Advisor to the Chancellor for Planning and Institutional Improvement and Professor of Organizational Leadership at Indiana University—Purdue University Indianapolis (IUPUI). **Keston Fulcher** is Executive Director of the Center for Assessment and Research Studies at James Madison University (JMU), Harrisonburg, Virginia.

*How many programs does your institution offer? Approximately how many outcomes get assessed each year?*

**Hundley:** IUPUI offers 350 undergraduate, graduate, and professional programs from Indiana University and Purdue University. We have 18 degree-granting schools and 30,000 students. Some schools or programs choose a cyclical approach of assessing—for example, three program learning outcomes each year for three years or three programs per year for five years. Others assess annually but may address only a few program or strategic priorities in a given year.

**Fulcher:** JMU’s assessment office serves our Quality Enhancement Plan (ethical reasoning), five general education areas and approximately 120 academic degree programs and certificate programs. We also serve several dozen student affairs programs. The number of student learning outcomes vary across these ~150 “units” but is typically about 10. And most of these units assess all student learning outcomes every year.

*What are typical sample sizes for programs?*

**Hundley:** Because of the programmatic diversity and decentralized nature of our culture and structure, assessment-related sample sizes for programs vary considerably.

**Fulcher:** With respect to general education outcomes, we assess all incoming freshmen and then assess these same students again about two years later. Students are randomly assigned to assessment room based on their student id numbers. Each room has a different battery of assessment measures that students take. This process allows us to gather large, representative value-added samples from 200 to 1,000 students on most general education student learning outcomes.

*How many of those outcomes are assessed using measurement standards, versus the more casual standards common in assessment? What's the difference in how they are used?*

**Hundley:** IUPUI faculty use multiple sources of evidence to assess student learning directly (e.g., embedded authentic course assessments, Principles of Undergraduate Learning evaluation, electronic portfolios, discipline-based standardized tests and licensure exams) and indirectly (e.g., through surveys of practicum or clinical supervisors, internship directors, and students themselves). Programs use disciplinary norms in their assessment approaches. Drawing on information from the Program Review and Assessment Committee (PRAC) assessment reports, we produce an annual Assessment of Student Learning at IUPUI Report, which provides a comprehensive look at assessment and improvement efforts institution-wide.

**Fulcher:** Relatively speaking, our general education tests and data collection processes meet high measurement standards. We often spend a year or two developing an instrument before it is used officially. There is more variability in the assessment processes for academic degree and student affairs programs; some processes are very strong from a measurement standards perspective, but others are not as well developed.

*Describe how a reliability/validity effort typically happens. Who leads it? What does that result in?*

**Hundley:** IUPUI has distributed leadership for assessment across campus, and reliability/validity efforts are embedded in assessment-related professional development programs and resources around campus. For example, CTL and IRDS both offer well-attended regular workshops on teaching, learning, assessment, and use of data for decision-making. Furthermore, the adoption of AAC&U's VALUE rubrics for assessing written communication, quantitative reasoning, and critical thinking skills has been encouraged; to date, 15 IUPUI faculty have been trained by AAC&U staff to use the rubrics. We also are involved in a number of other national assessment-related efforts (e.g., DQP; Tuning).

**Fulcher:** At JMU the validation process of test scores starts with content experts (CE) and assessment experts (AE). CEs, in great detail, articulate what a student should know, think, or do as a function of a program. AEs often push the content experts for more detail. Once the student learning outcomes are clearly defined, the AEs often hold item writing workshops. CEs then write items to map back to the student learning outcomes. If the assessment instrument is a rubric, the CEs and AEs work collaboratively to build out the rubric criteria and behavioral anchors so they are consistent with the SLOs. Note that up through this part of the process, there are no fancy analyses. The effort is spent on the collaboration between CEs and AEs, and the alignment between assessment instrumentation and the student learning outcomes. After a prototype instrument has been designed then it is piloted by the AEs. The AEs will then analyze the pilot data and investigate the data's psychometric properties. CEs are brought into examine the "bad" items and help decide which to modify and which to delete. The next version of the instrument is typically deemed worthy enough to go live. We also examine how test scores relate to course exposure, course grades, and other testing data. If the data behave in theoretically predicted ways, this adds additional validity evidence. Note that this resource consuming process is reserved for programs with the widest reach (most often general education).

*When making inferences, does data quality play a role? How?*

**Hundley:** While the collected data themselves are important, the most crucial part of the assessment and improvement process is the conversations about and actions taken in response to the data. Although the department does not aim for research publication-level statistical validity in its assessment data collection, the careful and systematic analysis of the data in light of supplemental evidence—for example, whether the instructor has concerns about student performance on any outcome in his or her class; whether or not those concerns are evident in the collected data; or whether students tend to perform well or poorly in subsequent classes that depend on a particular outcome or performance indicator—provide some confidence that any actions taken in response to assessment findings are motivated by true areas of concern and not just an outlier in the data. When there is a strong mismatch between an instructor's concerns and the data he or she actually collected, the assessment process itself is re-evaluated and refined to help ensure that collected data truly are valid and meaningful.

**Fulcher:** Absolutely. I have never presented results to faculty and not had to field legitimate questions about the trustworthiness of the data. If I can provide reasonable support for the veracity of the assessment evidence, then the conversation can transition to inferences about the program. If not, the conversation devolves into a diatribe about worthless data.

*What do you see as the major similarities and/or differences between assessment and research?*

**Hundley:** Good assessment and good research do share some commonalities, in that each strives to use methods and approaches that are designed to enhance our understanding of a given situation. Assessment tends to be focused on student learning or development taking place in a given program or context. This tends to make assessment more local and practical in nature. For all of the reasons described in other questions, assessment may be viewed as differently rigorous as research studies, and given the purposes of assessment, this is entirely appropriate. Research studies usually have the goal of contributing to the body of knowledge and may be designed to yield generalizable conclusions.

**Fulcher:** I've heard people make distinctions between research and assessment, and I just don't get it. Can good assessment be bad research? In my opinion "no." That said it is difficult or even impossible to achieve true experimental designs in higher education assessment. Random assignment to conditions, for example, is often either unethical, logistically unfeasible, or both. Nevertheless, we can aspire to strong quasi-experimental designs and data analytic processes like propensity-score matching to mitigate some design problems.

*What quality-of-assessment standards should peer reviewers for regional accreditors apply when assessing assessment findings?*

**Hundley:** Accreditors, perhaps rightfully so, tend to be more interested in the processes and what an institution is doing with its results. Hopefully, the standards themselves are being refined periodically to incorporate best practices in assessment, and peer reviewers are being trained to evaluate an institution against those continually-refined standards.

**Fulcher:** Few universities if any have the resources to conduct professional-quality assessment for every student learning outcome for every program. Nevertheless, they should spend the resources to have reasonable quality assessment particularly with large scale or high stakes programs. For example, if all students are expected to learn critical thinking skills at an institution, shouldn't faculty and other stakeholders have access to relatively accurate information regarding students' skills? At a bare minimum, any program should be able to make the argument that they have been thoughtful about the alignment of their student learning outcomes, their curriculum, and their assessment processes.

*What quality-of-assessment standards should peer reviewers for assessment journals apply when assessing assessment findings?*

**Hundley:** The purpose and audience of the assessment journal will largely determine the approach that reviewers use. In *Assessment Update*, we tend to look for timely, well-written, and practical case studies and advice on how assessment in practice occurs and is used to make improvements in various program and institutional types. In more research-oriented publications, reviewers may seek study designs that inform an understanding of measurement, incorporate appropriate disciplinary theories or models, and contribute to the Scholarship of Teaching and Learning.

**Fulcher:** Relatively speaking, they should be higher than the general standards for regional accreditation. In many assessment articles, the point is to investigate the efficacy of a program. And, from that article, readers should be able to make an informed decision whether this program may be effective at his or her own institution.

---

Josie Welsh is Director of Institutional Effectiveness at Missouri Southern State University. She can be reached at [Welsh-J@mssu.edu](mailto:Welsh-J@mssu.edu).

## INTERVIEW WITH DAVID CHASE, ASSOCIATE VICE PRESIDENT - EDUCATIONAL PROGRAMMING, ACCREDITING COMMISSION FOR SCHOOLS, WESTERN ASSOCIATION FOR SCHOOLS AND COLLEGES

*Jonathan Keiser, Interviewer*

***Based on your experiences, what are effective practices for promoting an institutional culture of assessment and learning?***

I think about these kinds of issues every day as they align with WASC's [Western Association of Schools and Colleges Senior College and University Commission] mission and priorities, and much of my work is focused on developing and offering programming that provides institutional assistance for developing a culture of learning and assessment. I have found institutions that approach student learning in an integrated and holistic way tend to have a stronger culture of learning and assessment, because it permeates their institutional practices and policies. WASC offers a variety of programming (<https://www.wscuc.org/educational-programs>) that assists institutions with these efforts. For instance, our workshop, *The Learning Institution: Aligning and Integrating Practices to Support Quality*, helps educators of all kinds within an institution to focus inward through reflective strategies to better understand what the institution is doing and use what they learn to improve experiences for students. I think many institutions that have a mature culture of learning and assessment have figured out how to operationalize a convergence model of leadership where administration, faculty and staff demonstrate a shared responsibility and accountability for assessing and improving student learning. The cycle of collecting, analyzing and utilizing data on student learning is well integrated throughout all parts of the institution and discussing evidence and data is a routine practice.

***From a Commission's perspective, what are common obstacles an institution encounters when designing and implementing assessment practices? How can institutions leverage accreditation agencies to assist with overcoming these obstacles?***

Probably the most common obstacle is approaching assessment of student learning from only a compliance perspective. Focusing on compliance misses the point and can counter the intent of accrediting standards and criteria. I have noticed that institutions that have well defined priorities and design assessment practices that are aligned to these priorities, and spend time considering how their institutional culture drives assessment practices are less likely to encounter obstacles in collectively understanding student learning and development. This is particularly true for institutions that do not couple assessing learning only with accreditation reporting and timelines. Accreditation efforts heat up around the time an institution is preparing for reaffirmation, which can create competition for resources (e.g., time, budget, staff) with other institutional initiatives. If an institution does not have well established and longstanding assessment practices, assessment efforts might be competing for time and energy that is also being directed toward reaffirmation. The best way to avoid this problem is to make sure assessment practices aren't being driven solely by external prompts such as an upcoming reaccreditation visit and use student learning assessment to drive internal improvements as part of an ongoing effort that has momentum independent of external drivers and prompts. This mission of the institution should drive efforts to understand, confirm, and improve student learning and development and serves as the touchstone for prioritizing the work.

***What advice would you offer to people tasked with coordinating and writing the assessment section of a Self-Study report?***

This is a common question that many people ask who find themselves leading their institution's reaffirmation efforts. Building a coalition across the institution that is grounded in collaborative processes that provides enough time for iterative feedback and review helps mobilize the institution as whole, and takes the burden off a single person tasked with writing and coordinating the assessment section of a self-study report. This reminds me of a characterization of leadership from Bolman and Deal (2008) as "...a subtle process of mutual influence fusing thought, feeling, and action. It produces cooperative effort in the services of purposes embraced by both leader and led." The person tasked with writing about assessing student learning wants to establish collaborative processes that bridge faculty insights and disciplinary knowledge with administrative support. These sorts of strategic alliances aren't one time conversations in a single meeting; rather they are collaborative processes requiring ongoing communication that fosters a shared responsibility for assessing, improving and reporting on student learning. There should be frequent conversations centered on data about students and their learning.

***Is there an intrinsic tension between improving student learning and proving students learn at an institution? If so, do you have any advice for institutions struggling with this tension?***

"Improving learning" versus "proving learning" is an interesting way to frame the tension that often exists when there are groups of people with different priorities discussing what, how, and why we assess student learning. Samuel Hope, Executive Director Emeritus of the National Association of Schools of Theatre and Music, is a brilliant thinker who recognized this tension and described large-scale assessment efforts as operating from a rhetoric of permanent accusation. This concept is powerful, and worth considering because assessment results can easily appear to be reductionist and framed by a deficit mindset rather than a means to understand the nature of student learning in an effort to improve teaching and learning. Engaging faculty in nuanced conversations about student learning that posits a shared responsibility for learning among students, teachers, administrators and the institution as whole can move from a rhetoric of accusation to one of understanding and improvement. Approaching learning and assessment conversations from a strengths mindset with a focus on improving learning while recognizing disciplinary expertise can help avoid or at least diminish this tension.

***Can you discuss how your organization trains peer reviewers regarding how to understand and evaluate an institution's assessment efforts?***

I'm glad you asked this question. We are revising and rolling out new training processes that have face-to-face, hybrid, and online components that leverage a learning management system so the structure and interface should be familiar to many faculty and administrators. WASC is investing in programming to assist institutions with improving educational quality and peer reviewer training focused on assessing and improving educational quality. We also sponsor the Assessment Leadership Academy, an 11-month professional development training program that already has almost 300 alumni. These 'graduates' are well trained in a variety of best practices for assessing student learning recognizing the unique mission of many institutions – and they poised to be excellent peer reviewers. We have also convened a Community of Practice with funding from a Lumina Foundation grant that is developing greater institutional capacity for assessment practices and increasing the visibility of student learning outcomes assessment and student achievement in the overall accreditation process. The Community of Practice is helping us demonstrate how assessing student learning and achievement is critical to internal constituents (faculty, administrators, peer reviewers) focused on improving educational quality and external constituents such as policy makers interested in understating the value of higher education. All of these efforts are designed to ensure that

WASC's peer reviewer training centers on institutional quality and improvement that account for the myriad unique institutional types and cultures in our region.

***What are some common misconceptions institutions have regarding reporting on assessment of student learning and accreditation expectations?***

I think one of the misconceptions is failing to account for a variety of assessment reporting needs - it's not always a one size fits all model. Different disciplines may require different epistemological lenses because they value different types of knowledge and knowing; Art History faculty might want to use a methodology and reporting structure that looks very different than an Economics faculty because they're trying to understand student learning from different perspectives.

A similar misconception is approaching assessment by trying to address only what the accreditor wants rather than what is meaningful to faculty and the institution. It's similar to a student asking, what do I need to know for the test? It should be more about what an institution needs to know about its students and their learning in order to celebrate successes and improve curriculum and instruction as necessary. Accreditation is a peer based quality assurance system, and the misconception can sometimes be "us" and "them" approach to assessment. There is no "us" and "them" because it's all "we." Accreditation standards emerge, are applied, and are interpreted by peers in the higher education community.

***Do you foresee any regulatory changes that will impact institutional assessment efforts?***

That's a good question, but difficult to answer because the policy landscape is volatile and it's difficult to predict the trajectory things will take as the process of reauthorizing the Higher Education Act develops. The best way to insulate your college from regulatory change is to make assessment an institutional priority. If assessing learning is part of the fabric of an institution's culture, it doesn't matter if regulations shift between process or outcomes – either way, the institution is systemically collecting, analyzing, interpreting, reporting and most importantly using data on student learning.

***What other thoughts or suggestions would you like to share with our readers?***

Well, I'm starting to think about assessment in terms of a developmental trajectory for institutional growth. Assessment can foreground broader work for institutional improvement because it privileges learning at the center of the effort. When reflection about student learning becomes an intentional institutional practice it can integrated functional areas. Assessment can get people in Academic Affairs and Student Affairs collaborating and contributing to common goals so there is a true sense of shared responsibility for students' and their learning. This is something I hope to see more of in the future.

**Reference**

Bolman, L.G., & Deal, T. (2008). *Reframing organizations. Artistry, choice, and Leadership*. San Francisco: Jossey-Bass.

---

Jonathan Keiser is Deputy Provost of Academic and Student Affairs at City Colleges of Chicago. He can be reached at [jkeiser@ccc.edu](mailto:jkeiser@ccc.edu).



## Registration opens in January 2018

- [Reserve your hotel room](#)
- Apply to become an [Institutional Conference Host](#)
- Submit your [session proposal](#) for AALHE's Eighth Annual Conference  
by January 20, 2018

Association for the Assessment  
of Learning in Higher Education

**AALHE**

PROMOTING ASSESSMENT FOR LEARNING

# INTERSECTION

## **Intersection Editorial Board**

Jane Marie Souza, Editor-in-Chief, University of Rochester

David Eubanks, Furman University

Jacob Amidon, Finger Lakes Community College

George Klemic, Lewis University

Gray Scott, Texas Woman's University

Josephine Welsh, Missouri Southern State University

Jana M. Hanson, South Dakota State University

Alison Witherspoon, American College of Education

Michelle Rogers, Des Moines University

Elizabeth Smith, University of Tulsa

Jeff Barbee, Indiana University School of Medicine

Steven J. Michels, Sacred Heart University

*[Intersection](#) is the quarterly publication of the*

*[Association for Assessment of Learning in Higher Education](#)*

2901 Richmond Road, Suite 130-318, Lexington, KY, 40509 | 859-388-0855 | [info@aalhe.org](mailto:info@aalhe.org)

*This work is licensed under a*

*[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#).*