

CONTENTS

Page 1
*Practical
Assessment*

Page 2
*Taskstream
Announcement*
Annual Meeting

Pages 3-4
*Faculty Learning
Communities*

Pages 5-7
*Reporting
Realities*

Pages 8-11
*Not My Bunnies
and Not My Hat*

Pages 12-13
*The Assessment
Diaries*

Pages 14-15
Member Q&A

Practical Assessment

by David Eubanks

Welcome to the spring *Intersection*, the quarterly publication of the AALHE. This edition marks a transition from a members-only periodical to public distribution and calls for papers. Our invitation to send us your thoughts on what constitutes “practical” assessment filled our inboxes with submissions. We selected a few that we thought represented a diverse set of ideas for inclusion in this issue. Other great pieces will appear in subsequent editions.

One thing is clear from the experiences you will read about here: not everything works. My favorite quote about practicality, for which I have not found a definitive attribution, is: *In theory, theory and practice are the same. In practice, they aren't.*

It can be disheartening to attend a conference and see only successful examples of practice or to see a pristine theory in colorful schematics, knowing that one's own efforts fall short. The articles within show that assessment can also be a series of unfortunate events, punctuated by victories often enough to keep some of us out of therapy some of the time.

My own contribution to practicality is an idea I learned the hard way. I flunked a job interview one time by being too eager to impress and consequently told a college dean that his assessment methods weren't very good. As in, “Oh, no. That won't work. Let me show you how to do it right.” Yes, it's a painful memory.

Subsequently, I learned the “yes, and...” trick that is now a standard tool for working with staff and faculty on assessment. The idea is simple: don't tell someone who has invested time and energy into a project that they aren't doing it right. Instead, help them understand how to modify and augment their current strategy. This leaves them in charge and you as the guide rather than overseer. Saying NO invokes the “you break it, you bought it” idea and has the opposite effect of what we usually want, because it absolves the person of all responsibility except for what we now tell them to do. It may help to think of a NO as a detonation followed by a mushroom cloud.

The other possibility exists too: that they have actually solved the problem in their own unique way, and it just looks odd at first glance. It may even be a new way to do things that improves on the methods we usually recommend.

Starting with yes is also the first rule of improv, so you can think of it as cross-training.

Taskstream Announcement

by Tara Rose, AALHE President

I am eager to share with you exciting news. The AALHE board has recently approved a partnership with Taskstream. Through technology, Taskstream is a company that similarly shares a vision in advancing effective assessment practices. Over the next three years we will work collaboratively in conducting research on current practices and trends in the art of assessment and publish articles and white papers to advance the assessment field. Further, Taskstream will provide steadfast support with our annual conference. As many of you know, our annual conference is known as a great networking opportunity for assessment professionals (limited to only 300 people).

The positive result of this partnership cannot be accomplished without your support – the support from our 600+ members. It is you that helps this organization flourish and we thank each and every one of you. I thank you for your dedication to our organization, to scholarship, to your institutions, and most importantly your dedication to students.

I hope you will join me in our excitement and support our decision to partner with Taskstream.

AALHE 2016 Annual Conference

by Leah Simpson

The theme of the 6th annual conference, which will be held in Milwaukee, Wisconsin, June 6th- 8th, is *Assessing What We Value: A Focus on Student Learning*. The 2016 AALHE Annual Conference is a premier forum for assessment practitioners in higher education. No matter if your institution is small or large, private, public, or for-profit, you will take home something valuable from this year's sessions. Our conference is designed to: provide professional development, connect assessment professionals with others in the field, and lead and advocate for best practice in assessment.

Featured sessions:

Monday, June 6th

- Pre-conference kick-off: Conversation with the AALHE board and conference committee
- Plenary: Thomas Chapel, Centers for Disease Control and Prevention (CDC), "Potholes on the Road to Good Monitoring and Evaluation."

Tuesday, June 7th

- Accreditors Special Session: Join a panel of representatives from regional accreditors and campus assessment professionals
- Keynote: Kathleen Wise and Charles Blaich, Wabash College, "Leading with Assessment: Ensuring that what Counts for Us Matters for our Students."

Wednesday, June 8th

- Plenary: Jeremy Penn, AALHE Board Member, "If a Student Learns and There's No One Around to Assess it, Does it Make a Sound?"

To register, review the conference schedule, or learn more about the 2016 AALHE Annual Conference, navigate to www.aalhe.org.

Faculty Learning Communities



Jeanne Mullaney

The Community College of Rhode Island (CCRI) is the state's only public associate's level postsecondary institution and the largest in New England with almost 17,000 students taking classes on four campuses and two satellites. Like our sister institutions across the country, we have sought to engage faculty in the assessment of student learning in the usual ways: we have invited experts to speak and consult on campus, sent faculty to conferences, offered workshops on campus, attended department meetings, and worked with faculty one-on-one. Over the past decade, a number of faculty members took part in these activities and began carrying out assessment-related tasks at the department and program level, but we still needed to find a way engage them in assessing CCRI's general education program.

Since sustained faculty development activities are known to be more preferable for promoting long-term change in practice than stand-alone workshops and activities, we tried a faculty learning community (FLC) model that would last for six months and entail monthly meetings. The catalyst for creating the FLC was CCRI's participation in the Multi-State Collaborative to Advance Learning Outcomes Assessment pilot project. Along with public post-secondary institutions in eight other states, CCRI agreed to collect 75-100 samples of student work that demonstrated skills in each of two areas — quantitative reasoning and written communication — in the fall 2014 semester. These were to be assessed against the AAC&U's Quantitative Literacy and Written Communication VALUE rubrics.

We planned to collect the samples from students who had completed more than 75% of their coursework, so for our Associate's level students this meant those who had completed more than 45 credits. To ensure a wide sample, we would collect only seven to ten artifacts from any instructor or course. We knew that we would need to approach dozens of faculty in order to gather the requisite number of samples. The FLC model we developed involved working directly with a small group of faculty members who would in turn each engage several colleagues. In this way, peer-to-peer faculty development was intended to ripple outward from the faculty learning community to a broader group of faculty.

We recruited the nine FLC members strategically to include faculty from disciplines that were more likely to give certain types of assignments in their classes, so that our FLC included one member each from biology, business, English, law enforcement, mathematics, physics, psychology, radiography, and sociology. We also intentionally targeted new faculty, so our FLC consisted of six new, one mid-career and two seasoned faculty members. These colleagues met six times over the summer and fall of 2014 to become familiar with the VALUE rubrics for written communication and for quantitative literacy and to consider the kinds of assignments that would align well to the rubrics. In a survey in which eight of nine members responded, 80% had never seen or heard of the VALUE rubrics prior to taking part in this project, although 75% reported that they had routinely used other rubrics. The FLC members also discussed CCRI's Definition of An Educated Person and the alignment of the MSC project to measuring student progress toward the goals that CCRI faculty have outlined for our graduates. The members studied the rubrics and considered their alignment to assignments that they regularly gave their students.

Although the ostensible goal of the group was to collect samples of student work that aligned to the VALUE rubrics so that we could fulfill our commitment to the MSC project, as much of the conversation during the FLC sessions was about improving teaching in general as it was about enhancing specific assignments. Indeed, in one of the first meetings, when members stated their reasons for joining the FLC, several mentioned that they wanted to take advantage of the opportunity to "network with others who are enthusiastic about teaching" or simply to "become a better teacher." One expressed the aim to "apply quantitative reasoning in class in purposeful activities." As a result, in addition to allowing us to gather the necessary samples of direct evidence of student learning, this model of sustained collegial interaction in a community of practice promoted faculty involvement in activities that allowed them to fulfill personal goals and improve their teaching.

(Continued on page 4)

Faculty Learning Communities

Continued from page 3

In the process of working to verify the alignment of assignments to the VALUE rubrics, the FLC members became familiar with some of the diverse assignments meant to elicit written communication and quantitative literacy across the disciplines at our institution. They were inspired by the originality of their colleagues, and they began to act as consultants for each other as the group adapted assignments to better align to the rubrics. In particular, the FLC members focused on how essential the assignment instructions were to guide students and prompt them to demonstrate the type of skills faculty wanted them to in the work. They considered how to improve their directions, as well as other aspects of their teaching, and in the process, they strengthened not only their course materials and assignments, but also their pedagogy.

Once the fall 2014 semester began, the FLC members worked with a total of thirty-six colleagues, educating them about the VALUE rubrics and working with them to modify their assignments if they were amenable to doing so. Five of the members made presentations at department meetings, so even faculty who opted not to participate heard about the project and learned about the VALUE rubrics from their peers. By January, we had assembled eighty-six samples of student work that demonstrated written communication skills and 122 samples of work that demonstrated quantitative reasoning from courses in eighteen disciplines across the three divisions of our institution. We met our collection goal for the MSC project in addition to meeting several professional development goals related to the scholarship of teaching and learning. One of the FLC members presented at a conference about her experiences working with the faculty in her department to redesign assignments to strengthen students' quantitative literacy. That faculty member and another created YouTube videos describing the process they used for redesigning an oceanography lab to incorporate more quantitative literacy skills and developing a quantitative literacy activity in psychology. These videos and others are available on the [MSC website](#) for faculty to reference.

In May 2015, five of the nine FLC members participated in a training session on the written communication VALUE rubric. They experienced the norming and calibration process and practiced scoring two samples, thus building our institutional capacity to use the rubrics to assess student work in the future. Of those five members, at least one will participate in the scoring as part of the MSC project as well.

The FLC and faculty members who had contributed samples to the project reconvened during the fall 2015 semester to consider the data from the pilot project. We were pleased to note that the demographics of the student sample closely matched the CCRI student population in gender, ethnicity, and age. Taking into consideration that the results may not be generalizable, the group reviewed student achievement and evaluated patterns of strengths and weaknesses. Overall, our students' quantitative literacy scores were a bit stronger than the aggregate in several dimensions, and their written communication scores were comparable to those of the aggregate for the thirty-one community colleges that participated in the MSC. However, in the area of sources and evidence in written communication, our students didn't demonstrate the ability to "use credible and relevant sources to support an develop their ideas" (from the Written Communication VALUE rubric) as well as we thought they should have. One of the solutions we explored was to ask our librarians to create workshops for faculty on how to develop students' skill in this area. We held one workshop this spring and have a series planned for the next academic year.

The faculty learning community paradigm has been successful at CCRI because it fostered collegiality as the members collaborated and came to view assessment differently. The faculty growth associated with participation in the FLC created a favorable reputation for it, and the model continues to grow as this year's cohort connected with and engaged a new group of colleagues in assessing the general education skills of CCRI's graduates.

Jeanne Mullaney is Professor of Foreign Languages and Cultures, and Faculty Assessment Coordinator at The Community College of Rhode Island. She can be reached at jmullaney@ccri.edu.

Photo credit: David Fischbach

Reporting Realities



Gray Scott

Although some faculty scoff at the validity of assessment, I find it harder to sell its reliability—particularly when the assessed artifacts are written or otherwise complex and drawn from across our general education courses outside of a controlled research environment. When reliabilities are reported honestly for such assessments, the numbers will almost always raise eyebrows among those who know statistics.

Some recent reports on reliability, such as Ashley Finley's 2011 article on VALUE rubrics in *Peer Review*, seem to suggest that broad assessments across-the-curriculum can meet peer-review-journal standards for reliability. Certainly the way Finley's study is often described (as having shown the VALUE rubrics are "highly reliable") fosters that impression. The Finley study attempted to gauge reliability of forty-four trained raters on three of the fifteen VALUE rubrics using a chance-corrected kappa statistic. Finley reports that the trained rater pool came up with an average kappa across three rubrics of .18. Her target was .70. Noting correctly that some researchers treat near-agreement as agreement, Finley collapsed the five levels of performance from the rubric down to just three, which raised the average kappa to .69. It should be added, however, that the VALUE rubrics only have four of those rating levels (1 to 4) on the page; level 0 is implied and an option for scorers who think to use it, but out-of-sight can be out-of-mind, leading one to speculate that perhaps Finley's final kappa reflects a scale with only two levels to it. In either case, the adjusted figures have contributed to an impression that assessment, which rarely if ever is conducted under controlled research conditions, can meet research-level benchmarks, yet the unaltered numbers tell what might be a more typical story.

Indeed, her unaltered numbers look a lot like the ones we're seeing in our own work. We shouldn't be at all surprised. Real cognitive work, and particularly written work, is messy, complex, and multifaceted, not a single skill, but an orchestration of interrelated skills. David Dunning and Justin Kruger have shown that our ability to judge performance depends to great extent on our own skill levels which are likely to vary considerably across raters and their own suites of skills. Put those two facts together, and we have a recipe for widely varying judgments of any complex student performance.

The degree to which raters can differ was demonstrated in a once-better-known 1961 study. Sydell Carlton, Paul Diederich, and John French recruited fifty-three professionals representing six fields and asked each participant to sort 300 college-student essays into nine "grades" based on their quality using whatever criteria they deemed appropriate. When every essay had been graded fifty-three times, it turned out that every essay in the stack had received at least five different grades. A full third of the essays (101) had received every grade possible.

Diederich reported and built on that work in his landmark *Measuring Growth in English*, a slim 1974 guide that I think more assessment professionals ought to read. What Diederich found was that, when it comes to complex artifacts like essays, raters tend to see many of the same issues, but because it's difficult to juggle many criteria, we rely on single-criterion heuristics, which can vary by rater. The grammar purists slam papers for comma splices while cheerfully endorsing a paper that is light on content. A reader who bases judgments on voice finds little to celebrate in a paper that is stale in tone but otherwise exceptionally researched.

Diederich took the most common factors and built one of the earliest and most-used writing rubrics out of them — the Diederich scale, which remained a staple in writing research until the late 1980s and which played a role in his consulting work with school districts. But even when he worked with faculty on the scale for some time, Diederich found that he could rarely obtain tetrachoric correlations between pairs of raters higher than .50, corresponding to a reliability of about .67 (p. 33).

(Continued on page 6)

Reporting Realities

Continued from page 5

In the January 2016 issue of the *Intersection*, David Eubanks makes a persuasive case for what I think is a healthier attitude toward reliability in assessment, showing both that too much agreement can be a sign of problems and that we routinely accept lower levels of reliability when making decisions based on online customer reviews. By way of illustration, Eubanks compares what might look like low Gini coefficients and Fleiss kappas for a sample assessment to parallel values for public rating systems and finds that the assessment can fare considerably better.

A customer-review benchmark might be a tough sell for faculty on the crunchy side of the curriculum, however. When closing the loop and describing results for stakeholders, we need an approach that lands somewhere between blithely but honestly reporting a free-marginal multi-rater kappa of .2 and engaging in shell games to produce expected reliabilities. Finley's side-by-side presentation of three different interpretations of reliability for VALUE rubrics offers one model, enabling multiple perspectives on the data.

In closing the loop with our university community, we include multiple views on reliability. Reports feature not only rater reliability for each criterion but also two kinds of mean scores. One mean score is based on an assumption that when multiple raters assess an artifact, the average will be close to the truth; the other is based on majority-rule judgments among raters for each artifact. In addition, we include Likert-scale survey feedback from raters and faculty alike on each criterion, which can help us identify criteria in need of revision.

In the spirit of assessing our new assessment system, we have been playing with two other strategies. The first is what we are calling a *criteria map*. For a given objective (say Critical Thinking), we plot each criterion in a chart with two axes, one for the scores and one for reliability. But because, in our approach, we let faculty indicate which criteria they think apply to the task they had students perform, we have a third variable to plot: how often a given criterion was used. Excel's bubble-chart option enables the building of charts with three variables, with the values in the last column determining the sizes of the bubbles, as seen in the figure on page 7. We multiplied raw 1-3 scores by 100 to put these on a more intuitive 100-300-point scale.

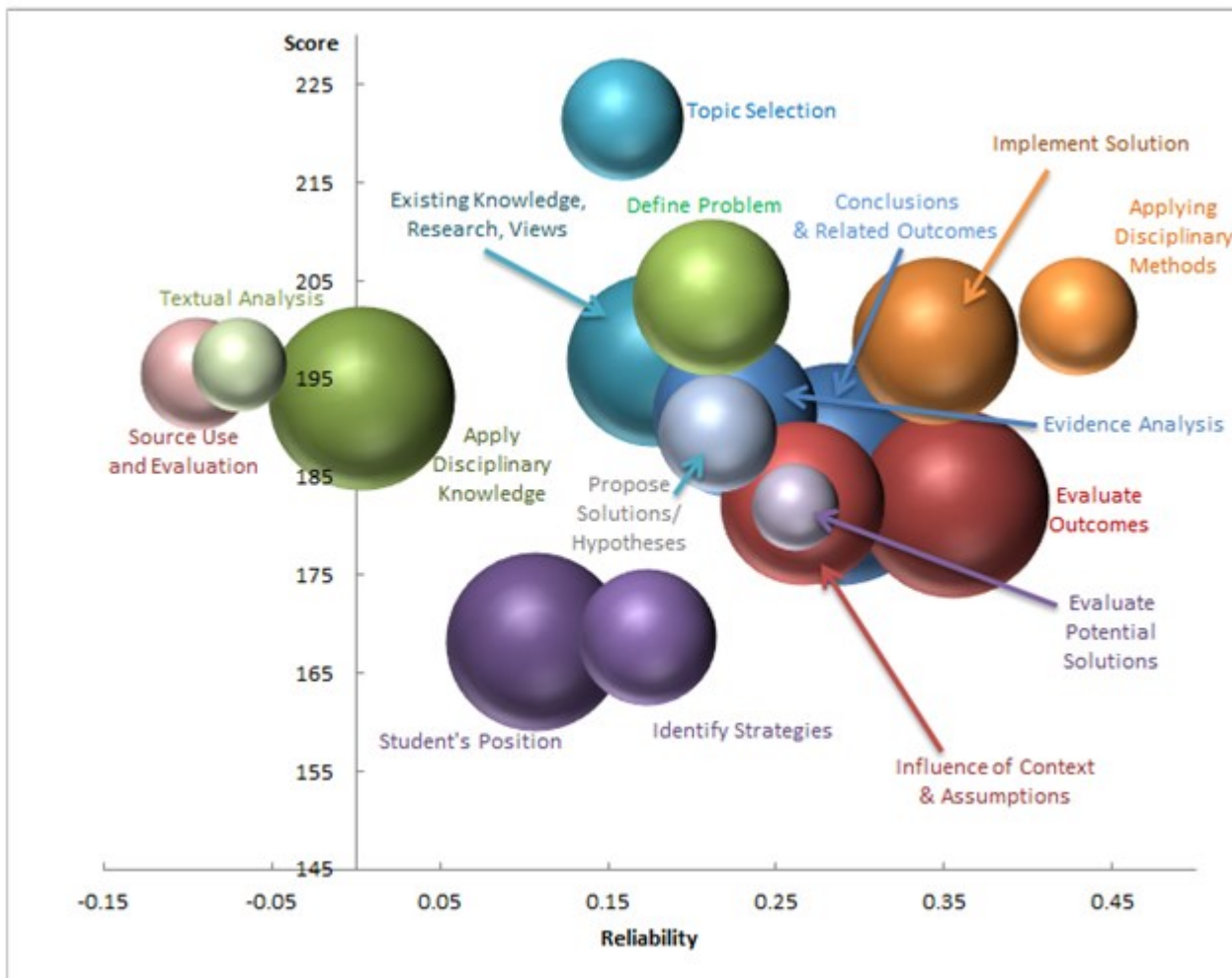
A criteria map enables community-wide assessment of the assessment itself, and we think the resulting discussion and evolutionary approach improves buy-in as we use the map to home in on and address problems. For instance, we can debate recommendations about which criteria (e.g. Textual Analysis) ought to be cut entirely due to both low-reliability and low-employment. We can identify popular criteria that seem relatively reliable (e.g. Implement Solution) and focus more rater training on them, seeing them as high-yield. In the case of lower-reliability, low-scoring criteria, like Student's Position, we might see whether the scores improve if we find a way to improve their reliability or whether, perhaps, faculty were devising assignments based on very different interpretations of the same criterion, leading to confusion among raters about exactly what those criteria should mean.

Finally, we can use the criteria map to set up the second strategy I alluded to, which is to frame hypotheses about what is going on in the above data—instead of assuming that we know—and then deliberately and explicitly test them in later assessments. Reliability and validity are ultimately about confidence in interpretations. Treating interpretations as hypotheses to be tested and then challenging them can lower anxieties about what will be done with such data, encouraging a greater degree of comfort with the fuzzier numbers common to assessment. Since piloting our new assessment approach in Spring 2015, we have been discussing an array of hypotheses, for instance that sharp gains between sophomore and junior year may be more related to attrition than to learning or that rewriting Student's Position might improve both reliability and scores, as well as possible ways to test them.

(Continued on page 7)

Reporting Realities

Continued from page 6



By reporting several indicators of reliability, by inviting assessment of our criteria through criteria maps, and by starting our process of interpretation with hypotheses, we appear to have defused a lot of the natural (and understandable) resistance to the new general-education assessment process, even while we are forthcoming about the data's shortcomings. The low reliabilities inherent to this messy practice now serve as entry points, rather than obstacles, to dialogue. Though we adopted these practices in part as change-management strategies, they seem to be worth keeping in the long run.

Further Reading

Diederich, P.B., French, J. W., & Carlton, S. T. (1961). Factors in judgments of writing ability. *ETS Research Bulletin Series*, 1961(2), i-93.

Diederich, P.B. (1974). *Measuring growth in English*. Urbana, Illinois: National Council of Teachers of English.

Finley, A. (Fall/Winter, 2011). How reliable are the VALUE rubrics? Peer review. Retrieved from AACU.org on March 31, 2016.

Gray Scott is Assistant Professor of English at Texas Woman's University. He can be reached at grayscott.twu.edu.

Not My Bunnies and Not My Hat



First row (left to right): Gretchen Hazard and Jessica Jacovidis; second row (left to right): Kathy Clarke and Jeanne Horst.

Institutions in Virginia are required to report to the State Council of Higher Education for Virginia (SCHEV) on key areas of student learning. These areas focus on student learning in general education, because all schools offer some type of general education programming. Virginia institutions report to SCHEV on six areas of student learning: critical thinking, writing, communication, information literacy, quantitative reasoning, and scientific reasoning. However, it is not just accountability that drives the assessment process. The primary goal for conducting assessment is to evaluate and improve students' learning.

James Madison University's (JMU) general education program is organized around five domains, each focusing on a different aspect of general education. Each of the five areas is developed around a core set of student learning outcomes, which gives each participating discipline flexibility in the way in which courses can be offered. For example, JMU's critical thinking courses are offered from a variety of perspectives: media arts, business, philosophy, and history. The departments and instructors have freedom in delivering the shared learning outcomes via any modality, using their disciplinary expertise. Our assessments provide accountability information, but more importantly, they provide evidence about whether our students are achieving the desired learning outcomes. In our domain, we use six or more different assessments, some of which are embedded in the courses and others that are administered on our university-wide assessment days. Most of the assessments were developed by faculty teaching in the domain of interest and assessment specialists in the Center for Assessment and Research Studies. For critical thinking, we use a commercial test called the Critical-thinking Assessment Test (CAT).

Randomly-selected sophomores take the Critical-thinking Assessment Test (CAT), which is designed to assess and promote the improvement of critical thinking and real-world problem-solving skills. This instrument was developed by faculty members from Tennessee Technological University with support from the National Science Foundation and is the product of extensive development, testing, and refinement with a broad range of institutions, faculty, and students across the country. For additional information on the CAT please visit www.tntech.edu/cat.

We assumed that faculty would be interested in hearing about the assessment results, but this is an ongoing challenge. We describe below some of our strategies to capture faculty interest in assessment results in one of JMU's five general education domains.

Level of analysis: Cohort versus course. One of the complaints of faculty is that assessment data do not apply to their own courses. Faculty members care strongly about their own students' learning. However, assessment data provided in aggregate about a cohort may be most useful for program-level conversations while telling faculty little about their specific courses. Traditionally, when we present faculty with information several levels

Quotes from General Education Faculty

I don't need to look at assessment results or change the way I teach. What I want is for my students to understand, and I can see THAT in their eyes.

I don't know what data to look at to make meaning of the results. And, when you make meaning for me, the results don't connect to what I teach.

I think the results of assessment are interesting, but I haven't a clue what to do with the results other than looking at them.

I have a lot of complaints about the instrument. I don't think the test has anything to do with my class.

(Continued on page 9)

Not My Bunnies and Not My Hat

Continued from page 8

removed from their courses, they are uninterested (Fulcher, Good, Coleman, & Smith, 2014). We have been pulling rabbits out of our hats, only to have faculty respond “those are not my bunnies and that’s not my hat.” Try as we might, faculty have a difficult time translating program-level assessment results into something applicable to their own courses. Institutionally, we may value program-level assessment for accountability and program assessment, but if we want faculty to use results, we need to make it meaningful to them. While necessary, providing program-level data will not do that. We continue to struggle with how to make program-level results meaningful to faculty.

Embedding Assessments. One solution is to embed the assessments within the course. For example, both our communication and information literacy assessments are embedded within the fundamental human communication course. Both tests are competency examinations that are strongly aligned with the student learning outcomes. Because the course is required for all entering first year students, we have data on over 4,000 students each year. Findings are reported out at the level of student learning outcomes, aggregating results across sections of the course or of the entire first year class. Moreover, we compare the current year’s cohort to those from previous years. Using this information, faculty members who teach or lead a section of the fundamental human communication course are easily able to identify where students are doing well and where they are not. For example, assessment findings indicated that students perform least well on the learning objective related to listening. The communication faculty members are afforded the opportunity to discuss ways in which they can strengthen students’ skills in that area. Because the test is so tightly aligned with the outcomes, we can identify improvement (or lack thereof) in subsequent cohorts of students.

Similarly, the information literacy assessment findings are reported by outcome. Faculty librarians are able to modify tutorial learning content to address areas of student need based on test results. Additionally, examination of assessment results for various student groups has allowed for targeted interventions for low performing student groups. By embedding these tests within required courses, and tightly aligning the tests with the student learning outcomes, we are able to provide program-level information that is also useful at the individual course level.

Embedding assessment within courses makes the data more meaningful to faculty. However, course-embedded program assessment is still a step removed from individual faculty. This is an important safeguard, as one fear consistently expressed by faculty is that assessment data will be used as evaluation measures. In order to prevent misuse, our institution has been careful not to present assessment findings at the course or faculty level. However, because multiple faculty may teach the same course, they respond to the assessment data by saying, “that’s my hat, but those aren’t my bunnies.” That is, although the faculty are teaching the same course, the results include data from multiple sections, so faculty can dismiss assessment results as the problem of other course sections, not their own. We are beginning to discuss how this practice limits the usefulness of assessment findings at the level of the individual faculty member. We are continuing conversations with faculty about how we could provide meaningful assessment results, while still protecting against the use of assessment results for faculty evaluations. Although we are a step closer, we are still removed. We have conceived of a new magic trick but are unclear how to perform it.

Infographic reports. Another faculty complaint is that reports are too long and too technical. For years, we presented 100-page reports on general education assessment findings. Although this is “required reading” for the beginning of the year meetings and for program administrators, it became obvious that no one was reading the reports. Thinking that faculty might engage more with shorter, easier-to-navigate reports, we created a short report for each area of assessment. The short reports include 1) student learning outcomes, 2) lists of constituents, 3) short descriptions of the assessment method and results with graphics, 4) progress made over the year, and 5) recommendations for the coming year. We

(Continued on page 10)

Not My Bunnies and Not My Hat

Continued from page 9

were pleased with our new conjuring trick, and we anticipated that lively discussions would ensue. Our audience remained underwhelmed. Timing is one issue. Faculty receive results just after the year has begun, after they have already written their syllabi and planned their classes. Or, they receive the results at the end of the year, when classes are over and they are anticipating a break. Moreover, the data are aggregated and at a level that feels irrelevant to faculty.

We needed to add some pizzazz to our magic, so we began creating infographic reports. These reports are color-coded by assessment area and include graphical depictions of the student learning outcomes. The text is minimal. Since introducing the new infographic reports, we have seen increased discussion about the findings. These new reports have encouraged us to think about data in different ways. Learning outcomes are no longer simply a bulleted list; graphics permitted them to become more process- rather than product-oriented. Additionally, instead of letting the instrument guide the reporting format, we can better organize the information so that we can easily see students' strengths and weaknesses. Moreover, graphic reports allow us to highlight critical data points and make it easier to navigate important results. At the same time, faculty response has not been completely positive, and we have found that beautiful reports do not compensate for assessment data that are not relevant to individual faculty. However, while the infographic approach is one that has been successful, it is not the complete answer to the question about how to capture faculty imagination. As colorful as the routine has become, until we can provide assessment results at a level that is meaningful to the individual faculty members, we will continue to struggle with getting faculty to engage with assessment results. Regardless of how pretty the hat or how cute the bunnies, faculty don't make the connection to their own classes. The magic that eludes us is motivating them to care about all the bunnies instead of just their own.

Assessment as faculty development. Faculty may not trust assessment results if they do not know how they are generated. One way to get them to value the performance is to let them see behind the curtain. We attempt this with both writing and critical thinking assessments. These performance assessments are administered during low-stakes assessment testing sessions outside of class. Faculty members are then recruited, paid (and fed) to spend a day rating student work. Those who participate in the sessions gain first-hand knowledge of student work as well as experience with the rating tools. The university writing rubric is made available to all faculty members. Those who participate in the rating sessions gain expertise on the nuances of using the rubric, which they report as helpful in evaluating their own classroom assignments. We end each rating session with a group discussion of student learning and use of assessment results. At least for a day, we have captured faculty attention and engaged in rich discussions about student learning. If we are lucky, we help them accessorize their hat, ensuring a better link between learning outcomes and course content. However, we don't know if they actually use it with their bunnies.

Because the rating sessions are such a rich source of faculty development, we are hesitant to abandon them. However, they are costly, and because we can only include a limited number of faculty and student work products, this strategy does not necessarily lead to widespread engagement in assessment. Individual faculty members may come away excited, but by the time the assessment reports reach them, they are less excited with the overall program-level report. Faculty appreciate the work of the day, but we have no evidence that they carry it forward either in their own departmental discussions or within their own classes. The lasting benefit is that once they peek behind the curtain, they think differently about assessment than they did before.

Assessment as scholarship. Although the scholarship of teaching and learning is not equally embraced by all disciplines, framing assessment as a research question has captured the imagination of some faculty members. We have found that it leads to collaborations and faculty assessment champions

(Continued on page 11)

Not My Bunnies and Not My Hat

Continued from page 10

across campus. This has taken the form of developing teaching interventions, instrument development and use, conference presentations, and publications. Collaboration among assessment professionals, administrators, disciplinary experts, while engaging with assessment data sets the stage for faculty to become their own magicians.

We have pulled a lot of rabbits out of a lot of hats, added pizzazz to our performance, allowed our audience to peek behind curtains, and invited them to become magicians themselves. Our big takeaway from all of our tricks is that direct relevance to faculty activity is key. Unless we can make assessment results meaningful to faculty, all the magic in the world will not convince them of its utility. Our efforts will continue as we try to address these challenges and engage faculty in assessment for the larger good of improving student learning. The magic is in the learning outcomes. Although it is tempting to think of the magic hat as the discipline or the assessment test or method, what we really want to demonstrate is that the bunnies are affected (knowledge, skills, and abilities) after they come out of the hat.

Further Reading

Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014). *A simple model for learning improvement: Weigh pig, feed pig, weigh pig*. (NILOA Occasional Paper). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Jessica N. Jacovidis is Graduate Assistant, Center for Assessment and Research Studies at James Madison University and can be reached at jacovijn@jmu.edu.

Kathy Clarke is Interim Director of Assessment & Planning, Libraries and Educational Technologies at James Madison University and can be reached at clarkeke@jmu.edu.

Gretchen Hazard is Cluster 1 Coordinator, General Education, University Studies at James Madison University and can be reached at hazardga@jmu.edu.

S. Jeanne Horst is Assistant Professor and Assistant Assessment Specialist, Center for Assessment and Research Studies at James Madison University and can be reached at horstsj@jmu.edu.

AALHE Membership

If you are not already a member, please consider joining us! As a member of AALHE, you have access to a wealth of resources that will help you as an assessment professional: the annual conference and conference proceedings, webinars (including access to the webinar archive), and networking opportunities in a quickly growing assessment organization.

Joining is simple: navigate to the application form at <http://www.aalhe.org/join-aalhe/>. Current annual dues are \$120 for standard membership and \$90 for graduate students.

The Assessment Diaries



Lisa Wiehle

The accreditors are coming... We've all been there. More and more, assessment of student learning has been in the spotlight, ready or not. In 2006, Elgin Community College in Northern Illinois was visited by the North Central Association (NCA) for our 10 year review. One of the concerns highlighted by the panel was the institutionalization of assessment. A faculty committee had been formed, processes had been developed, lots of data was being collected, but our feedback report said they'd be coming back in 2009 to check in on our progress.

The committee continued to meet, processes and paperwork were brought online, still more data was generated. There were forms and due dates, reports were dutifully organized into files, and most faculty knew that if it was February 1st or October 15th, something was owed to the Assessment Office. We felt good about our progress. The focus visit was successfully completed in the spring of 2009. In the final report, the team made the following observations:

- The institutional assessment plan appears to diminish the value of narrative stories about how classroom teachers are applying changes in response to student learning experiences.
- The college is strongly encouraged to find additional ways to document and bring forward assessment success stories. Without the power of narrative stories, quantitative assessment data may not be as easily understood or contextualized.
- Finding ways to celebrate and recognize assessment success and encourage more faculty sharing of the impact of intentional practice in assessment would help demonstrate ongoing institutionalization of a culture of evidence.

Okay, got it. Stories, not data. Wait... what?! As a former institutional researcher, I know that data and evidence are must-haves. But the feedback made sense. We still kept busy with measurement and reporting, but we also began organizing a campus-wide event for the following spring. We would invite a speaker, encourage our faculty to present workshop sessions, and conclude with a celebratory lunch. (And cake. Always have cake.) We titled the event The Assessment Diaries.

That first year, I had to do a lot of personal outreach and call in some favors to encourage proposals. I was initially surprised by some responses to these requests. Faculty were surprised I thought others would be interested in their assessment projects, or they questioned whether their results were significant enough. A few even told me they didn't really like talking in front of people. It turns out that some instructors don't consider teaching to be public speaking! That year attendance was what we called "intimate." But the feedback overwhelmingly confirmed we were on the right track. The setting was casual and low key, and dedicated time away from the pace of a busy semester allowed for reflection on student learning. People met new colleagues and were surprised they could still learn a trick or two from a program very different than their own.

The college committed to running the event annually, and we decided to stick with the Assessment Diaries brand. Attendance has grown steadily. Each year we create a theme and give away some sort of "diary" – a notebook or notepad. I love seeing them used in meetings throughout the year. For the past two years, students from the college's design club have created the program, and we hope to create more opportunities for student participation in the future.

The event has become institutionally recognized as a platform for faculty and staff to share their broader experiences in teaching, learning and student success as opposed to "just" assessment. Session topics have been quite varied and have included classroom techniques, program-level assessment, placement research, general education, technology integration, faculty sabbaticals, and co-curricular and student services projects. A selection of session titles illustrates their diversity.

(Continued on page 13)

The Assessment Diaries

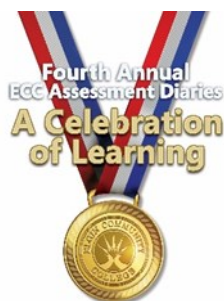
Continued from page 12

Selected session titles:

- The Power and Potential of Alumni in the Student Success Equation
- Graduate Follow-up Studies
- Integrate Writing, Critical Thinking, and Active Learning into Your Classes
- Reading Strategies in Content Courses
- Prompting Student Participation with Google Classroom
- Assessing Student Presentations
- Faculty/Library Partnership for Student Research Projects
- Using Data to Revise Math Placement Policies
- Creating Rubrics in D2L to Assess Student Work (note: D2L is a learning management system)
- Global Competence and Multi-Cultural Education

A formal awards ceremony was added in 2016 to recognize good assessment work in the categories of course assessment, program review and curriculum design/re-design. Nominees were announced prior to the event (which we believe drove some registrations), and panels of faculty, staff and administrators anonymously scored them with a rubric to determine the winners.

A fourth category, President's Choice, was also created to recognize broader efforts towards student success. We provided a list of potential nominees to our president who then chose the winner and kept the secret until the ceremony. In his presentation remarks, he indicated that the winners tackled large barriers for student success which were identified and understood using data, and subsequently have used assessment data to demonstrate improvement. In our inaugural year for this award, there was a tie, recognizing a cross-departmental team which implemented the Student Planning Module within Ellucian/Colleague, and a college-readiness effort with university and secondary-school partners to develop curriculum for a fourth-year math/stats course to be offered to high school seniors within the district. Distinctive trophies were given to the winners and already a healthy competition towards next year has begun.



One last comment from our 2009 visit report served to remind us “assessment is not an end in itself, but a means to an end. Improvement of student learning should be at the heart of the institution’s mission.” A new team of peer reviewers came to campus this past October. While we still have forms and due dates, we were confident they would find that the college now views assessment as another means to ensure quality, and to uphold our mission, *to improve people’s lives through learning*.

Lisa Wiehle is Manager of Outcomes Assessment at Elgin Community College, and can be reached at lwiehle@elgin.edu
Photo credit: Mark Carriveau.

Member Q&A: Practical Advice

What do you know now that you wished you had known when you were starting out in assessment?

Assess What You Value

- *When introducing a person or program to assessment, start with something they care deeply about doing well, rather than what might be easiest to assess.*
- *Assessment must align itself to practical knowledge, intuition, critical thinking and decision making opportunity rather than measuring only learned or memorization skills.*
- *Assessment is about effective teaching not accreditation.*
- *Assessment has many levels, and it is important to assess what you value in your organization.*
- *It is important to move faculty away from thinking about “my course” to thinking about “our program.”*

Assessment Is Complex

- *Rule number one is to start small, keep it simple, and focus on doing a few things very well.*
- *It is very difficult! Some assessment is better than no assessment.*
- *Assessment can be messy with few clear-cut decisions and approaches.*
- *Assessment is an ongoing process that never ends.*
- *Have patience; quality takes time and yields the best results.*
- *Assessment has depth, from course assessment to program-specific assessment, as well as general education.*
- *Assessment has so many layers across an institution and it can take a lot of time to learn about how it all fits together.*
- *Many times your assessment work cannot progress because you are reliant on the input of others....The result is a long running list of half-finished projects that will ultimately get completed, but you have to be willing to work with lots of balls up in the air at any one time.*
- *Cultures differ between and among academic units.*
- *Assessment can be flexible, fun, and free of formulaic practices that stifle intellectual curiosity!*
- *Faculty often do not understand assessment even though they think they do. It is not the same as grading. It takes time to explain the vocabulary of assessment and the process.*
- *Assessment touches many college-wide processes, departments, policies, faculty, administrators, and student services staff, which can make it quite challenging to keep all stakeholders apprised of updates or changes to assessment of student learning and how it impacts all facets of the college.*

Working as an assessment professional is every bit as much about your ability to work with people than it is about content expertise.

Methodology Matters

- *Building a curriculum map is important; it can drive everything.*
- *Use valid course signature assignments to assess program outcomes.*
- *Create benchmarks and mastery courses to monitor student learning and growth.*
- *Start with “How to Write Program Outcomes” and then check each department’s work before going on to the next phase of writing assessment plans. Don’t assume departments will follow the guiding principles and write strong outcomes.*
- *Use various methodologies in assessing student learning outcomes.*
- *Assessment is not a mystery; rather, it is an opportunity to tell the story—who, what, how and why.*
- *The “all-in-one” rubric has the potential to provide a lot of detailed feedback about student learning and performance to both faculty and students, as well as assist with documentation of that assessment at the course, program, and institutional level.*

(Continued on page 15)

Member Q&A *Continued from page 14*

What do you know now that you wished you had known when you were starting out in assessment? *(cont.)*

Expect Resistance

- *It all seemed so logical in the beginning, but people fight this tooth and nail.*
- *Assessment evokes strong negative emotions. People hold a variety of understandings and misunderstandings of what assessment is. It is incredible, but detractors in higher education actually do not see that assessment is important.*
- *It is hard to get some faculty and staff to report their assessment data.*

Learn how to fend off attacks from those not wanting to engage in the process and how to argue for the value of assessment beyond accreditation.

What is the most important unanswered question or issue you still have about assessment?

Can It Be?

- *When leading folks down the path to meaningful assessment, can “good enough” be okay as a transitional phase?*
- *For programs that effectively engage with program assessment, can changes in student learning be effectively documented, and can that documentation be credible and trustworthy?*
- *Is it reasonable to expect that the faculty members who teach at an institution could become a community of practice dedicated to improving their craft as teachers?*
- *Will it ever be possible to develop meaningful standards for student learning outcomes within disciplines and fields and across institutions?*

What Are...

- *The best assessment/improvement practices for non-academic units?*
- *Regional accreditors looking for in terms of reporting learning outcomes assessment?*
- *The best strategies for creating and sustaining a culture of assessment?*
- *Meaningful ways of engaging adjunct faculty in the assessment process?*
- *Best practices for disseminating assessment results to multiple stakeholders to keep them engaged in how they are used?*

How?

- *How do we integrate assessment into teaching rather than have it seen as an add-on?*
- *How do we convince faculty that grades are not equivalent to assessment of student learning outcomes?*
- *How do we assess transferable cognitive skills such as critical thinking?*
- *How do we communicate assessment results and planned improvements to a variety of stakeholders? How do we tell the story of assessment and why it matters?*
- *How do we promote the development of action plans (based on data) that are substantial and not just a minor tweak to make the assessment director go away?*
- *How do we ensure faculty have selected appropriate assessments that reflect the rigor of the course and align with learning outcomes?*
- *How do we overcome the challenge of faculty not often seeing assessment as an essential part of their work?*
- *How can we promote meaningful executive sponsorship for assessment?*
- *How do we engender academic responsibility along side academic freedom?*
- *How do we help students understand the relevance of institutional assessment across their college career?*