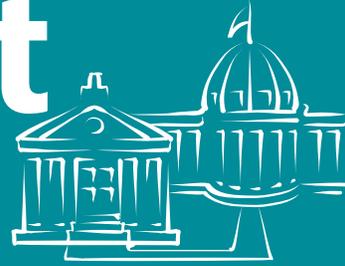


Assessment UPdate

July–August 2015
Volume 27, Number 4



Progress, Trends, and Practices in Higher Education

Assessment: Learning by Doing

Patricia Terry

ACTIVITIES THAT DO NOT COME TO US NATURALLY ARE LEARNED BY DOING THEM, Aristotle (n.d.) observed: “[M]en become builders by building houses, harpers by playing on the harp.” Similarly, faculty members responsible for assessment on their campuses often lack training and instead learn by practicing it in their own institutional contexts. We may seek information about assessment through study and conferences, but we also learn from rolling up our sleeves and digging in. At Gonzaga University, learning assessment by doing it has been an incremental and imperfect but valuable process that has helped us identify strengths in the current university core curriculum and influenced our planning for a new core.

As the coordinator of a set of linked courses in our core, I began looking for ways to evaluate their effectiveness in 2007. However, a background in rhetoric and literature did not equip me with research methods for measuring student learning. Even more problematic, the curriculum was established long before most of our faculty were thinking about assessment or taking it seriously, and program learning outcomes had never been specified. Since 1975, students had been required to enroll concurrently in “Thought and Expression” (T&E)—100-level courses in Composition, Critical Thinking, and Speech—and in 1995, we began enrolling cohorts of twenty students in limited numbers of linked sections. The links did have stated objectives: to reinforce skills taught in the three courses and to foster integration by creating a learning community for first-year students. However, neither Thought and Expression nor the twenty-six-year-old core curriculum had explicit program learning outcomes that would allow for direct assessment. Still, to investigate the hypothesis that linked courses helped students improve, we had to start somewhere.

We experimented with indirect and direct assessment over the next few years. First, a colleague experienced in a variety of research methods helped design and pilot a student questionnaire related to the T&E objectives. In 2008, we began by surveying students in the linked courses to ask how they perceived conceptual integration among the three courses, whether they could apply what they learned in one course in another, and how they experienced community. In 2011 and 2012, we administered a version of the student survey in several sections of T&E that were not linked. During this time,

CONTENTS

ARTICLES

- | | |
|---|---|
| Assessment: Learning by Doing
<i>Patricia Terry</i> | 1 |
| Editor’s Notes
<i>Trudy W. Banta</i> | 3 |
| Supporting and Evaluating Learning Outcomes Assessment through Academic Program Reviews
<i>Ingrid Novodvorsky, Debra Tomanek, Ryan M. Foor, Gail D. Burd</i> | 4 |
| Outcomes Assessment at Small Colleges: Achieving Two Goals with One Assessment
<i>Jennie Brown, Heather Corday Allard, Dré Goode, Rachel Rossetti</i> | 5 |
| Unexplored Variables in Course-Embedded Assignments: Is There an Elephant in the Room?
<i>Mary Kay Jordan-Fleming</i> | 7 |

COLUMNS

- | | |
|--|----|
| Assessment Measures
<i>Gary R. Pike</i> | 8 |
| From the States
<i>Peter T. Ewell</i> | 10 |

Assessment Update

Progress, Trends, and Practices in Higher Education

July–August 2015

Volume 27, Number 4

Editor

Trudy W. Banta, *professor of higher education and senior advisor to the chancellor, Indiana University–Purdue University Indianapolis (IUPUI)*

Managing Editor

Karen Elaine Black, *director of program review, IUPUI*

Assistant Editor

A. Katherine Busby, *director of institutional assessment, Tulane University*

Book Review Editor

Susan Kahn, *director of institutional effectiveness, IUPUI*

Consulting Editors

Peter T. Ewell, *National Center for Higher Education Management Systems*

Thomas Anthony Angelo, *University of North Carolina at Chapel Hill*

T. Dary Erwin, *James Madison University*

Cecilia L. Lopez, *Harold Washington College (ret.)*

Marcia Mentkowsky, *Alverno College*

Jeffrey A. Seybert, *Higher Education Consultant*

Peter J. Gray, *US Naval Academy (ret.)*

Gary R. Pike, *professor of higher education, IUPUI*

Assessment Update: Progress, Trends, and Practices in Higher Education (Print ISSN 1041-6099; online ISSN 1536-0725 at Wiley Online Library, wileyonlinelibrary.com) is published bimonthly by Wiley Subscription Services, Inc., A Wiley Company, at Jossey-Bass, One Montgomery St., Suite 1000, San Francisco, CA 94104-4594. Presorted Standard U.S. Postage Paid at Providence, RI. Individual subscriptions are \$135 per year (institutional \$257). Back issues are available in limited supply at \$29 per issue. To order, phone toll-free (800) 835-6770 or email cs-journals@wiley.com. Visit our Web site at www.josseybass.com. *Postmaster:* Send address changes to Assessment Update, Jossey-Bass, One Montgomery St., Suite 1000, San Francisco, CA 94104-4594.

Copyright © 2015 Wiley Periodicals, Inc., A Wiley Company. All rights reserved. Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030; (201) 748-8789, fax (201) 748-6326, www.wiley.com/go/permissions.

JOSSEY-BASS
A Wiley Brand

we also attempted a direct assessment to complement the survey. A composition instructor of both regular and linked sections designed an assignment that would allow us to compare students' performance in these classes. Using his rubric, faculty from the three departments read and scored the essays. Though they were not able to identify patterns of differences, the experiment taught worthwhile lessons about sample size, identifying a common language across disciplines, and even the timing of the scoring session.

The student survey did not provide direct evidence that the linked courses develop students' skills of critical thinking and thoughtful expression, but it did give us indirect evidence that other program objectives were being achieved. A colleague in Institutional Research compiled the ten semesters' worth of data collected in linked T&E classes and also compared the fall 2011 and 2012 survey data from both linked and nonlinked classes. The results showed that, overall, students in the linked sections perceived themselves to have a better understanding of how the courses were connected and to have benefitted more from relationships with their classmates. While the 1,095 responses collected from the linked sections since 2008 showed consistently high scores for questions about whether linked classes created a community and helped students learn, the added responses from the non-linked classes provided a basis of comparison. Looking at the mean scores from the Fall 2012 linked section surveys ($n = 40$) and those from the non-linked surveys ($n = 64$), the greatest gap between the nonlinked and the higher linked ratings appeared in statements like these:

- The relationships I developed with students in my T&E classes helped me learn.
- Familiarity with my classmates in my T&E classes helped me stay more engaged in these classes.
- I spent some time outside of class studying with students from my T&E classes.

In 2011, on most questions about integrating concepts among the courses and connecting with other students, the mean scores were higher in the linked sections ($n = 116$) than those in the nonlinked sections ($n = 137$). In 2012, the linked sections' mean scores were *(continued on page 16)*

Call for Contributions

The editor welcomes short articles and news items for *Assessment Update*. Guidelines follow for those who would like to contribute articles on outcomes assessment in higher education.

- **Content:** Please send an account of your experience with assessment in higher education. Include concrete examples of practice and results.
- **Audience:** *Assessment Update* readers are academic administrators, campus assessment practitioners, institutional researchers, and faculty from a variety of fields. All types of institutions are represented in the readership.
- **Style:** A report, essay, news story, or letter to the editor is welcome. Limited references can be printed; however, extensive tables cannot be included.
- **Format:** In addition to standard manuscripts, news may be contributed via letter, telephone, or fax (317) 274-4651. The standard manuscript format is a 60-space line with 25 lines per page. Articles may be sent to aupdate@iupui.edu as a Microsoft Word attachment. Please include your complete postal mailing address.
- **Length:** Articles should be four to eight typed, double-spaced pages (1,000–2,000 words). Annotations of recent publications for the Recommended Reading feature should be 200–500 words in length. Short news items and content for the Memos section should be about 50–200 words long.
- **Copyright:** Articles shall not have been registered for copyright or published elsewhere prior to publication in *Assessment Update*.
- **Deadlines:** Each issue is typically planned four months before its publication.

Please address mailed contributions and comments to Trudy W. Banta, Editor, *Assessment Update*, Suite 4049 University Hall, 301 University Blvd., Indianapolis, IN 46202–5146. ■

Linking Outcomes Assessment with Valued Campus Processes

Trudy W. Banta

WE KNOW THAT AN ESSENTIAL way to engage faculty in outcomes assessment is to connect assessment with a process that individual faculty, or groups of faculty, truly value. That process differs among individuals and groups, and even may differ over time for the same faculty. No method or approach works with everyone all the time. This is what keeps outcomes assessment so interesting—and challenging!

Some of the processes that may capture the attention and involvement of faculty colleagues at one time or another in their careers include efforts to enhance student learning, curriculum review and revision, program review, the scholarship of teaching and learning, evaluation of the effectiveness of a new initiative inside or outside class, faculty development, and—of course—rewards and recognition. We are privileged to have in this issue several illustrations of processes that have drawn faculty into outcomes assessment.

In the first article, Patricia Terry capitalizes on the interest of faculty at Gonzaga University in evaluating the effectiveness of linking three courses in the core curriculum, a process initiated in 1995 but previously not evaluated. First, faculty developed a questionnaire to ask how students perceived integration among the three courses. Then a composition instructor designed an assignment and an accompanying scoring rubric that facilitated comparison of students' performance in the linked sections with that of students in nonlinked classes. Data collected over ten semesters suggested benefits for students enrolled in linked sections, and

the finding interested faculty sufficiently that they recommended retaining linked courses in a new core curriculum. In addition, instructors with consistently high ratings on questionnaire items shared their instructional approaches at a faculty meeting.

Academic program review is the valued process with which learning outcomes assessment is linked in the article by Ingrid Novodvorsky, Debra Tomanek, Ryan M. Foor, and Gail D. Burd. In 2011 academic affairs principals at the University of Arizona instituted a requirement that program self-studies prepared in advance of campus visits by respected peers include sections on outcomes assessment practices, findings, and uses of findings in undergraduate and graduate programs. Academics in the Office of Instruction and Assessment (OIA) made this requirement more attractive by offering concurrent consulting services related to assessment. For instance, faculty in programs preparing self-studies in a given year are offered workshops in which the standards to be used by administrators in evaluating assessment plans and practice are explained. When self-studies are submitted, OIA principals rate the assessment sections using a rubric with four criteria: Expected Learning Outcomes, Assessment Activities, Assessment Findings, and Changes in Response to Findings. This *stick* (the assessment requirement) and *carrot* (consultation on assessment) approach has elevated the level of attention given to outcomes assessment at this research university.

As at Gonzaga, interest in improving student learning provided a catalyst for faculty engagement in outcomes assessment

at Franklin Pierce University. But finding a more systematic approach to program assessment also grew out of program review, as it did at the University of Arizona. In an effort to give students in psychology an opportunity to apply their course-based learning in survey design, sampling data collection and analysis, and presentation of research findings, psychology faculty member Jennie Brown involved students Heather Corday Allard, Dré Goode, and Rachel Rossetti in a study designed to elicit evaluative data from alumni for use in program review. Student research assistants were able to raise the response rate on an e-mail survey from 9 percent to 39 percent by following up using Facebook or personal e-mail messages. This approach provides a win-win situation in which students gain valuable experience in applying their classroom learning while faculty obtain valuable information for informing their judgment about the strengths and weaknesses of their academic offerings.

In the fourth article, Mary Kay Jordan-Fleming reveals a sobering truth that has begun to dawn on those like columnist Peter T. Ewell who are engaged in the current effort to validate the Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics for widespread use: Not only do we need to be concerned about interrater reliability for use of the rubrics, but we also need to ascertain that the assignments giving rise to the work to be rated are valid stimuli for calling forth critical thinking and quantitative literacy skills, for example. That is, are the assignments valid for

(continued on page 16)

Supporting and Evaluating Learning Outcomes Assessment through Academic Program Reviews

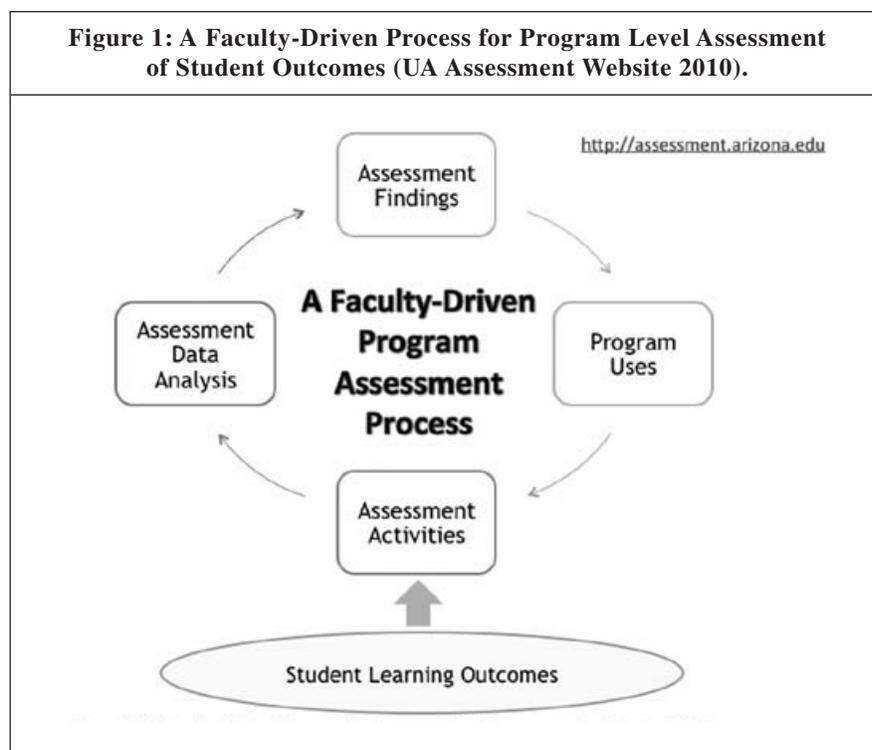
Ingrid Novodvorsky, Debra Tomanek, Ryan M. Foor, Gail D. Burd

THE UNIVERSITY OF ARIZONA (UA) IS a research-intensive, land-grant institution with an enrollment of more than 41,000 students across 205 academic programs. Programs undergo Academic Program Review (APR) every seven years, with an average of 30 programs reviewed each year. In the fall of 2011, learning outcomes assessment became an integral part of APR expectations, with required sections of the APR Self-Study Report on assessment practices and results in undergraduate and graduate programs. In this article, we describe how support and accountability for outcomes assessment are built into our APR process, early signs of how the process is working, and reflections on lessons learned.

A Model for a Faculty-Driven Outcomes Assessment Process

In its latest accreditation review by the Higher Learning Commission, the UA was found to have a weak plan for tracking learning outcomes assessment across academic programs. To address this shortcoming, we developed a model built on best practices that could be used by all programs to improve their outcomes assessment practices (Figure 1).

The model depicts a cycle that starts with identification of student learning outcomes. These outcomes provide the basis for creation of assessment activities, analysis of data collected through the activities, and reporting of findings. This



faculty-driven, within-unit process yields student learning evidence that is used to drive discussions of program changes and improvements (Banta et al. 2009; Maki 2004; Suski 2009; Walvoord 2010).

Buy-In and Support: The Carrot-and-Stick Approach

The next step of our comprehensive plan for improving learning outcomes assessment was to identify a process for operationalizing the new assessment model across campus. A problem with outcomes

assessment at our institution has been the low priority it holds in many programs. We decided to use a *carrot-and-stick* approach to change this. Faculty members take seriously the APR process, a review conducted by both upper administration and an external evaluation team of peers (see <http://www.academicaffairs.arizona.edu/academic-program-review>). Requiring information about the unit's learning outcomes assessment practices in its APR would elevate the importance of

(continued on page 12)

Outcomes Assessment at Small Colleges: Achieving Two Goals with One Assessment

Jennie Brown, Heather Corday Allard, Dré Goode, Rachel Rossetti

TWO MAJOR HURDLES to developing and engaging in effective program assessment are having too little time and not enough money (Haywood 2010). These drawbacks can be even more difficult at small institutions where faculty have higher teaching loads, institutional resources are fewer, and there is no assessment office, assessment committee, or assessment guru. Nevertheless, we will discuss a method for including student researchers in program assessment in a way that can lighten faculty workload while also providing a valuable educational opportunity for undergraduate students.

As part of a recent program review process, the psychology department at Franklin Pierce University developed a more systematic method for program assessment. The faculty had been doing assessment informally for many years, observing student performance and listening to students' comments, analyzing this information individually and in department meetings, and then making changes to curriculum based on that analysis. However, department faculty acknowledged that more systematic research, including an alumni survey as well as more direct measurement of student learning outcomes, would lead to better feedback and a better program.

Psychology faculty decided to involve students in research, as psychology undergraduates learn about proper survey design techniques, sampling, data collection and analysis, and the presentation of research findings—all of which can

be part of the outcomes assessment process. Additionally, program evaluation is a major area of applied psychological research. Psychology undergraduates usually are briefly exposed to program evaluation in their research methods courses, but generally they do not put it into practice or have much experience with it. The ability to evaluate a program is a skill that students can use in many work settings and in graduate school to determine whether programs are meeting their goals successfully. This is true for intervention

experience, which is critical for graduate school admission in psychology. RAs were selected based on their knowledge of research methods, attention to detail, writing ability, statistical skills, and professionalism. The RAs worked with the lead researcher to develop measures for assessing key concepts in which department faculty were interested. These concepts included what faculty expect graduates to achieve in their time in the program, assessment of program learning outcomes, and how alumni perceive that

The ability to evaluate a program is a skill that students can use in many work settings and in graduate school to determine whether programs are meeting their goals successfully.



programs for drug and alcohol abuse as well as academic programs—although with the latter it may be more difficult to operationalize and measure outcomes.

Psychology faculty developed a proposal that included the use of student research assistants (RAs) to assist a faculty member (lead researcher) in an outcomes assessment project using data from the program's alumni. Department faculty saw this as one valid way to assess program outcomes (additional methods would include more direct measures). One psychology faculty member acted as the lead researcher on the project and liaison between the RAs and the department. Funding to pay RAs was granted by the university, although many students would likely have volunteered for this research

their education affects them in their current career or in graduate education.

Methodology

Assessment design and implementation took place over two years and involved three RAs. The lead faculty member gave the RAs a list of items of interest within the department, including the program's learning outcomes. RAs then worked together to create a survey instrument based on these items. This questionnaire included measures of the participants' demographics (date of graduation, other education received, current job/education, income, and so on—information that would be useful for recruitment of new students and for potential internship contacts), measures

of the program's learning outcomes (participants rated their agreement/disagreement with these items and were allowed to make comments if they liked), and questions about the participants' perceptions of the curriculum. RAs worked with the lead researcher to remove any biased wording or double-barreled questions. Once the questionnaire was in its final stages, the lead researcher worked with the RAs to obtain Institutional Review Board (IRB) approval and approval from the Alumni Relations Office.

Once IRB and Alumni Relations Office approvals were granted and the questionnaire was completed, the questions were put online using Survey Monkey to obtain data electronically from alumni. Since the current learning outcomes had been in place for the past five years, only alumni who had graduated in this period were considered for participation. Alumni Relations aided assessment efforts by e-mailing alumni and asking them to participate in the survey. Unfortunately, this resulted initially in a 9 percent response rate—well below average (39 percent, according to Cook et al. 2000). However, response rates increased within one week to 39.6 percent by having one RA contact alumni through the department's Facebook page or via e-mail.

Once data collection was completed, an RA calculated means and standard deviations for all of the quantitative items (some items were scored on a 7-point scale from "strongly agree" to "strongly disagree"). Participants also had an opportunity to provide comments on some items. Participants' comments had to be coded or categorized based on similarity to determine if there were recurring themes. With the help of the lead researcher, the RA put together a presentation and report to share the findings with the department and administration.

Lessons Learned

Psychology department faculty benefited greatly from this process, as they

were able to obtain important information on their alumni. They saw this as a great opportunity for experiential learning and research experience for the RAs involved. RAs were able to provide excellent assistance in this project, which made the project more feasible given the small amount of time and money available for assessment. The RAs were able to gain new insights, in addition to their research experience. One insight demonstrated that a program (academic or otherwise) is never completed, but is constantly in development. RAs also got an inside look at what teaching faculty do and the care and thought that go into assessment and into program change and development. RAs were able to apply and practice the skills they had learned in their courses (including research design, data collection, data analysis, and communication of findings) for real-world customers (psychology department and university administration). They were also able to use what they had learned in their course work to make a difference for their university, which was empowering. RAs reported gaining an understanding of how to be professional and friendly to participants and faculty members and became more comfortable in these interactions. RAs gained an understanding of the time, attention to detail, and effort that goes into program evaluation. RAs also learned that the point of program assessment is to develop a better program and that it will never lead to a "perfect" program.

Psychology department faculty were able to enhance their understanding of the program's strengths and weaknesses from the perspective of their alumni. The specified strengths included the professors individually, the small size of the department, and the diversity of the curriculum. Additionally, on the items with rating scales (there were twenty-five, rated from 1 = "strongly disagree" to 7 = "strongly agree"), nineteen were rated 6 or higher, indicating that the majority of students agreed

that they had achieved these program outcomes. In terms of areas for development, six items were rated between 5 and 6. These scores demonstrate general agreement that the skills were learned, but the means were lower, suggesting that the program should focus more on developing students' skills in statistics and research methods, the influence of the nervous system on behavior, and the effects of genes and environment on behavior. An additional area to consider for improvement is students' preparation for the program's comprehensive exam (students would have liked more class time to review for this).

Overall, this was perceived as a beneficial and effective process. Now that it is in place (survey online, method for involving RAs, etc.), it can be continued and adapted over time with the help of future RAs and the input of department faculty and administration. Although faculty members outside psychology and related disciplines may not be trained in research involving the operationalization and measurement of human perception and performance, this process could be replicated by other academic areas with the help of a few psychology faculty members and students. ■

References

- Cook, C., Heath, F., and Thompson, R. 2000. "A Meta-analysis of Response Rates in Web- or Internet-Based Surveys." *Educational & Psychological Measurement* 60 (6): 821.
- Haywood, A. 2010. "Academy for Assessment of Student Learning Preliminary Results." Presented to Higher Learning Commission OASIS Advisory Team, Lisle, IL.

Jennie Brown is an associate professor of psychology, and Heather Corday Al-lard, Dré Goode, and Rachel Rossetti are research assistants at Franklin Pierce University in New Hampshire.

Unexplored Variables in Course-Embedded Assignments: Is There an Elephant in the Room?

Mary Kay Jordan-Fleming

HAPPENED UPON A COMIC STRIP THE other day that showed several people seated in a comfortable room. One poor man was being suffocated by an enormous pachyderm wedged onto his lap. The caption read, “Only Alan was prepared to acknowledge the elephant in the room.” At times, I feel certain that Alan makes his living in the assessment field.

Assessment scholars call increasingly for authentic course-embedded measures; indeed, the advantages are significant. But equally compelling are the validity concerns raised when faculty-generated prompts are not aligned to learning outcomes and rubrics. As the use of course-embedded assignments gains popularity, assessment professionals would be well advised to consider potential sources of unwanted variability and how to control them.

Course embedding captured national attention when the Association of American Colleges and Universities (n.d.) initiated its Valid Assessment of Learning in Undergraduate Education (VALUE) project, producing a common set of extensively tested rubrics to help institutions benchmark student learning in sixteen key areas. The rubrics move “students’ own complex college work—projects, writing, research, collaborations, service learning, internships, creative performances, and the like—to the center of the assessment equation” (Sullivan 2014, 1). The National Institute for Learning Outcomes Assessment advocates authentic work as the means by which students demonstrate the competences in the Degree Qualifica-

tions Profile (Lumina Foundation 2011) and is building an online library of exemplars. Significant advantages of course embedding include: (a) students are motivated to perform because assignments are also graded by course instructors; (b) there is a seamless fit with course content; (c) there is opportunity for meaningful developmental feedback; and (d) the assignments incur no added cost or requirements for students. Used in combination with university-wide rubrics and trained faculty raters, course-embedded artifacts are an appealing alternative to standardized tests. They promote closing the loop and can deepen investment in the process, promoting faculty leadership of efforts to improve teaching and learning.

We have only begun to scratch the surface of the design and measurement issues involved in course embedding. Maki (2014) observed that “poorly designed assignments lead to poorly executed student work.” Both Seifert (2012) and Greenhoot and Bernstein (2012) urged aligning assignments with rubrics. Pierakos et al. (2013) found that both the structure and complexity of tasks influenced problem-based learning. To date, the most thorough consideration of relationships among course design, assignment specifications, and assessment was offered by Ewell (2013), who called for robust course-embedded assessment utilizing curriculum mapping and assignment templates that mirror the rubrics.

Research at our institution (Jordan-Fleming 2013) examined essay prompts from six courses in which faculty stated

that they required evidence of interdisciplinary synthesis. On average, only 33 percent of students in these courses met the criteria for mastery, but further analysis discovered a marked performance range, 0 to 74 percent passing rates, across the assignments. Scrutiny of the assignment specifications revealed troubling differences in the clarity of instructions, with the poorest-performing classes having been told to “include interdisciplinary approaches, if appropriate.” Clearly, a prompt that allows students to opt out of using an interdisciplinary approach will not elicit their best work in doing so. The disconnect between what instructors thought they had asked students to do and what they had actually asked students to do is one of several potential sources of unwanted noise in measures of student learning, and one that few institutions measure or control. Assessment processes that incorporate third-party raters and common rubrics typically begin with an all-important training session designed to develop consensus about rubric standards among assignment readers, a step that Seifert (2012) identifies as key to establishing the reliability of rubrics. I would suggest that, in addition to norming the assignment readers, we norm the assignment *writers* if we want to establish validity.

A new project at our institution is attempting to do just that. Initiated in conjunction with a backward-designed curriculum revision, we require all courses fulfilling general education to submit

(continued on page 12)



Assessment Measures

The CLA+

Gary R. Pike

Since its introduction in 2002, more than 600 institutions have used the Collegiate Learning Assessment (CLA) to evaluate student learning at the institution level. Recently, the Council for Aid to Education (Council for Aid to Education [CAE] n.d.) introduced a new version—the CLA+. A few years after it was introduced, I provided in this column (*AU* 18 (2): 13–14) an overview of the CLA. In a subsequent column, Stephen Wall-Smith (2006) described the experience of one university in administering the CLA (*AU* 18 (3): 14–15). With the introduction of the new assessment, I thought it appropriate to review the CLA+ to see how it differs from its predecessor, the suggested uses of the instrument, and evidence of reliability and validity.

The most basic difference between the CLA and the CLA+ is that the CLA was designed for institution-level assessment, whereas the CLA+ is designed for student-level assessment.



The most basic difference between the CLA and the CLA+ is that the CLA was designed for institution-level assessment, whereas the CLA+ is designed for student-level assessment (CAE n.d.). CAE staff note that the CLA+ can be used with entering freshmen as an advising tool for identifying students' academic strengths and weaknesses. They also suggest that the CLA+ could be used as an admissions measure for transfer students. At the end of a student's

academic career, CAE staff observe that the CLA+ can be used as a competence certification measure. CAE staff also suggest that institutions may wish to place CLA+ scores on students' transcripts to document their critical thinking and problem-solving abilities. Scores for graduating students can also be aggregated at the institution level as a summative assessment of institutional effectiveness. Finally, CAE staff recommend using the CLA+ as a measure of value added, much as was done with the CLA.

Like the CLA, the core of the CLA+ is its performance task, which presents students with a realistic problem or issue. Students must address the problem or issue by suggesting a solution or recommending

a specific course of action. Students have one hour to complete this part of the assessment. As background, students are provided with a documents library with four to nine documents relevant to the problem or issue. These documents may be technical reports, data tables, newspaper articles, and/or memoranda. In completing the performance task, students must present a decision or recommendation and support it with information from the documents provided. In

the sample performance measure provided by CAE, students are told they are the chief marketing officer of an athletic equipment company that sells a brand of inline skates. As a result of a recent skating accident in which a teenager was seriously injured, the company has received negative press, and critics claim the company's advertisements are misleading. Using the information provided, a student is instructed to make a written recommendation concerning whether the company should continue its present ad campaign.

Students' written recommendations are evaluated using a scoring rubric. Separate evaluations are provided for (1) analysis and problem solving, (2) writing effectiveness, and (3) writing mechanics. The analysis and problem-solving dimension of the rubric focuses on the ability of students to arrive at a logical conclusion and support that conclusion with appropriate information. The writing effectiveness dimension evaluates a student's ability to construct organized and logical arguments, including convincing evidence, and the writing mechanics dimension examines a student's command of written English (language, syntax, and grammar). Each dimension is evaluated on a 6-point scale from 1 (low) to 6 (high). A rating of 1 on the analysis and problem-solving dimension indicates that a student stated or implied a conclusion, but provided minimal analysis and support for the position. A rating of 6 would indicate that a student explicitly stated a conclu-

sion or decision, provided comprehensive support for the position, and thoroughly refuted contradictory evidence and/or alternative positions.

Unlike its predecessor, the CLA+ contains twenty-five selected-response questions. Ten of the questions measure scientific and quantitative rea-

Unlike its predecessor, the CLA+ contains twenty-five selected-response questions. Ten of the questions measure scientific and quantitative reasoning, ten assess critical reading and evaluation, and five evaluate a student's ability to critique an argument.



soning, ten assess critical reading and evaluation, and five evaluate a student's ability to critique an argument. In answering the questions, students must read and draw information from accompanying documents. In the sample test, questions concerning scientific and quantitative reasoning require a student to read an article, "Fueling the Future," about how scientists are exploring alternative methods of fuel production to find sustainable and renewable forms of energy. One possible source of fuel discussed in the article is biodiesel fuel from algae. Students are then presented a selected-response question about factors that negatively affect biodiesel algae's ability to be an alternative fuel source. The article provides charts and graphs, and several questions are designed to measure a student's ability to extract important information from them.

CAE staff report that interrater agreement (i.e., correlations) for the performance tasks ranges between 0.67 and 0.75 across four performance tasks (CAE 2014). Across two forms of the CLA+, alpha reliabilities for the selected-response items were 0.80 and 0.78. Using a stratified coefficient

alpha, CAE staff report that reliability coefficients are 0.85 and 0.87 for combined performance and selected-response parts of the assessment.

In one study, students who completed the CLA+ were asked to report how much time they had remaining when they completed the performance

tasks and selected-response questions. In most cases, 90 percent or more of the students reported having at least some time remaining. Students were also asked how much effort they put into the assessment. Almost 45 percent reported they put forth a moderate effort, and 34 percent indicated they put forth a lot of effort. Slightly more than 15 percent of the students reported they put forth their best effort, whereas slightly less than 6 percent reported putting forth little or no effort. When asked how engaging the tasks were, slightly more than 20 percent indicated the tasks were not at all engaging or slightly engaging, almost 52 percent reported the tasks were moderately engaging, nearly 25 percent stated the tasks were very engaging, and less than 4 percent stated the tasks were extremely engaging.

As evidence of construct validity for the CLA+, the CAE staff report the results of a 2008 study that examined the correlations among CLA scores, scores on the Collegiate Assessment of Academic Proficiency (CAAP), and scores on the Measure of Academic Proficiency and Progress (MAPP) (Klein et al. 2009). The high positive correlations among the

three tests are reported as showing that the tests measure the same underlying constructs. CAE staff have also examined the face validity of the CLA+. Samples of students completing the CLA+ were asked how well the different parts of the assessment measured the skills of writing, reading comprehension, mathematics, and critical thinking and problem solving. For writing, reading comprehension, and critical thinking and problem solving, approximately 70 to 75 percent reported that the assessments measured the skills from moderately to very well. More than 82 percent of the students indicated that the CLA+ measured mathematics skills not well at all or slightly well. ■

References

- Council for Aid to Education (CAE). n.d. *CLA+: Comparing CLA to CLA+*. Retrieved from http://cae.org/images/uploads/pdf/cla_vs_cla_plus.pdf
- Council for Aid to Education (CAE). 2014. *CLA+: Reliability and Validity of CLA+*. Retrieved from http://cae.org/images/uploads/pdf/Reliability_and_Validity_of_CLA_Plus.pdf
- Klein, S., Liu, O. L., Sconing, J., et al. 2009. *Test Validity Study (TVS) Report*. Supported by the Fund for the Improvement of Postsecondary Education. Retrieved from http://cae.org/images/uploads/pdf/13_Test_Validity_Study_Report.pdf
- Wall-Smith, S. B. 2006. "Assessment Measures: The Collegiate Learning Assessment at Harris-Stowe State University." *Assessment Update: Progress, Trends, and Practices in Higher Education* 18 (3), 14–15.

Gary R. Pike is a professor of higher education and student affairs in the Indiana University School of Education—Indianapolis. He can be contacted at pikeg@iupui.edu.



From the States

Multistate Redux: Two Initiatives Expand in Scope?

Peter T. Ewell

In 2013, I reported in this column (*AU* 25 (1): 10–11) on two emerging multi-state assessment initiatives that were in their initial stages and that were expected to expand significantly in scope and state coverage. The first was Massachusetts's innovative statewide student learning assessment initiative based on senior-year student assignments and papers rated by faculty using the Association of American Colleges and Universities' (AAC&U) Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics. The second was the Western Interstate Commission for Higher Education's (WICHE) pilot of a multi-state, multisector, student-tracking initiative designed to determine the feasibility of linking unit-record data across postsecondary education, elementary and secondary education, and the workforce (*AU* 25 (5): 9–10, 12). Both efforts have developed considerably and promise to break new ground beneficially in the area of multistate assessment.

As I mentioned in that first 2013 column, the Massachusetts effort was watched closely by a number of interested states and culminated in an exploratory meeting, hosted by the State Higher Education Executive Officers (SHEEO) and AAC&U, to gauge interest in forming a multistate pilot project using the same methodology. This meeting was held in early 2012 with delegations from sixteen states attending. Outcomes were sufficiently promising that funding was obtained from the Bill and Melinda Gates Foundation. The resulting project, dubbed

the Multi-State Collaborative to Advance Learning Outcomes Assessment (MSC), was launched in December 2013 with sixty-nine institutions in nine states—Connecticut, Indiana, Kentucky, Massachusetts, Minnesota, Missouri, Oregon, Rhode Island, and Utah. The objective was to conduct a full-scale, multistate demonstration of the methodology pioneered by Massachusetts. This project involves drawing representative samples of student work, archiving and storing these “artifacts” to be evaluated as evidence of student learning in quantitative literacy and written communication (critical thinking was also assessed by some campuses but was not required), scoring these artifacts using trained faculty raters, and developing a reporting function to communicate results to various stakeholders. Because the effort was a pilot composed of institutions participating voluntarily, there is no attempt to generate representative performance information for each state, although interstate comparison is an eventual goal of the MSC.

The sampling plan called for collecting seventy-five to a hundred artifacts for each competence produced by students in the fall of 2014. As of this writing, the MSC has collected 8,811 artifacts generated by 1,182 assignments (2,930 in quantitative literacy, 3,712 in written communication, and 2,169 in critical thinking). The artifacts have been archived for scoring and analysis in a specially constructed Web-based data platform commissioned from TaskStream. Creation of the TaskStream platform was a major

achievement in itself as it provides a single location for storing, archiving, manipulating, and scoring artifacts seamlessly and efficiently. A cadre of faculty scorers drawn from each state was recruited, and scorers were trained to use the VALUE rubrics in February of this year. Scoring is now under way by 170 scorers from sixty-five institutions and state agencies. Twenty percent of the artifacts are rated by two trained scorers; double scoring will not only bolster reliability but will also enable interrater reliability statistics to be computed. The demonstration will be completed in September 2015, and Gates has supplied funding to undertake an operational phase of the initiative, which will launch in 2015–2016. In this phase, the objective will be to use the now proven methodology to generate valid state-level summary scores for each higher education sector that can be publicly reported and compared. Maine and Texas have joined the original nine participating states in the MSC for the operational phase, and thus far there have been serious conversations about participation with North Carolina, Hawaii, Colorado, and South Dakota.

Meanwhile, WICHE's multistate, multisector student-tracking initiative, entitled the Multi-State Longitudinal Data Exchange (MLDE) and also supported by Gates, has continued to develop. The original pilot involved linking unit-record postsecondary education, elementary/secondary education, and workforce data in four states—Hawaii, Idaho, Oregon, and Washington—and proved that it was

feasible to administer such a complex data exchange. Students from two starting cohorts, one consisting of high school students graduating in 2005, and one consisting of new entrants into postsecondary education in 2005–2006 were tracked over six years and yielded nearly 200,000 individual records to be matched. Moreover, the pilot generated enough actual data on student and workforce mobility within this four-state region to assess the utility of such an effort, although proven, technical feasibility remained a challenge.

Probably the most important technical issue was identity resolution—ensuring that records matched from quite different data environments really belonged to the same person. For example, WICHE analysts excluded 4 percent of the students examined in employment/earnings matching because available evidence indicated that the Social Security numbers (SSNs) associated with these students were shared with at least one other individual with distinctively different personally identifiable characteristics, or more than one SSN was supplied for the same individual.

In terms of what was found, the data exchange improved on information available to state policymakers. For example, sharing data made it possible to determine where college students ended up after graduation for at least 7 percent more college graduates than states could determine using their own data resources. Some states experienced greater levels of information enhancement through the exchange because they were small and isolated; greater levels of information enhancement are also expected as the number of states participating in the initiative increases.

These promising results from MLDE's four-state pilot effort prompted the Gates Foundation to allocate four

more years of support in July 2014. Their expectation is that at least ten more states will begin similar multi-state data-sharing efforts using experience gained in the pilot. Like the original four states, these additional states are expected to join as geographically contiguous “clusters,” and the original four states will continue participating. Criteria established for states joining the initiative include support for the MLDE among key stakeholders including their ability to identify policy issues that might be illuminated by longitudinal data of this kind, a substantial number of students likely to cross the state's borders in search of employment or further education, and demonstrated technical capacity to exchange data with other entities, together with experience with the legal and regulatory aspects of doing so.

Throughout the spring, WICHE project staff conducted a series of regional meetings intended to share results of the pilot and gauge interest among states in joining. Assuming that the target of ten additional state participants is met, project working groups will be drafting updated Memoranda of Agreement (MOAs) for all states. The plan going forward, moreover, is for WICHE to cease playing the central governance role in the initiative in July 2016, so another priority is for participants to evolve a free-standing governance structure and a related business model to assure long-term sustainability.

In short, both multistate initiatives have grown in sophistication and in the number of states participating since I wrote about them two years ago. As they develop, however, there are still challenges to face. The first is “initiative fatigue.” State agency capacities are stretched these days with budget challenges and escalating needs to respond to external demands for information. At the same time, personnel

capacities are limited. It is telling to see many of the same senior staff from common participating states at both MSC and MLDE meetings. Their participation is gratifying and always immensely fruitful, but suggests that the “bench” available for states to staff such multistate initiatives is thin. A second challenge is governance and, in particular, managing the transition from active leadership by SHEEO and WICHE supported by Foundation funding, to self-organized, free-standing organizations with sustainable business plans. Both MSC and MLDE will be working actively on these matters in the coming year. The final question is about state policy use of the resulting data. Both initiatives will need to build a compelling case to stakeholders that the additional information about student progress and performance they will generate are worth the substantial continuing investments in people and infrastructure these efforts will demand.

Definitive answers to these challenges will have to await the end of the coming year. But two conclusions seem clear at this point. First, multi-state data initiatives are coming of age with respect to both state coverage and data sophistication. Second, these efforts are capable of generating more and better information for both policy decision making and accountability than states can achieve using their own data resources alone. They are also proving more straightforward and effective in generating useful information than the many federal proposals to produce similar information that are emerging in Washington as reauthorization looms. ■

Peter T. Ewell is vice president of the National Center for Higher Education Management Systems (NCHEMS).

Unexplored Variables in Course-Embedded Assignments: Is There an Elephant in the Room?

(continued from page 7)

student artifacts for assessment on a designated subset of learning outcomes. Published rubrics and assignment guidelines (similar to Ewell's assignment template) assist faculty in constructing assessment-appropriate prompts. All such assignments require prior approval by the academic assessment coordinator, who examines them to make sure that the designated learning outcomes are clearly embedded. While this approach is proving more labor-intensive than anticipated, its value lies in minimizing one known source of variability in the data.

Still, other sources of unwanted variability may lurk. What about assignment complexity or content familiarity? For example, both statistics and calculus courses may fulfill a quantitative literacy requirement, but the types of problems given in those classes will vary along dimensions that may affect measurement of quantitative literacy. Similarly, one assignment assessing global citizenship may have a narrow scope while another builds on a complex set of readings and thorough discussion, thus "seeding" the problem with more possibilities for satisfying rubric expectations. Variations of this type affect the validity of course-embedded measures and will surely obfuscate our conclusions about student learning.

Backward curricular design, curriculum mapping, sound pedagogy, assignments carefully aligned to valid and reliable rubrics, carefully trained raters, and closing the loop are all critical. Each of these steps will build the integrity and usefulness of course-embedded data on student learning. Realizing that goal will require intentional and well-coordinated faculty development efforts that are appropriately staffed, funded, and seamlessly integrated with curriculum and assessment. Assignment writing is an important meta-professional skill that cannot be assumed even among the most experienced faculty. ■

References

- Association of American Colleges and Universities. n.d. *VALUE: Valid Assessment of Learning in Undergraduate Education*. Retrieved from <http://www.aacu.org/value/>
- Ewell, P. T. 2013. *The Lumina Degree Qualifications Profile (DQP): Implications for Assessment* (NILOA Occasional Paper #16). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Greenhoot, A., and Bernstein, D. 2012. "Using VALUE Rubrics to Evaluate Collaborative Course Design." *Peer Review*, Fall 2011/Winter 2012: 22–26.
- Jordan-Fleming, M. K. 2013. "Characteristics of Courses and Assignments that

Develop Students' Interdisciplinary Problem-Solving Skills." Presentation at the annual conference of the Association for the Assessment of Learning in Higher Education, June 3–5, Lexington, KY. Lumina Foundation. 2011. *Degree Qualifications Profile*. Indianapolis, IN: Lumina Foundation. Retrieved from http://www.luminafoundation.org/publications/The_Degree_Qualifications_Profile.pdf

Maki, P. L. 2014. "Using AAC&U's VALUE Rubrics to Assess Authentic Student Work Across Massachusetts' Public Higher Education Institutions: An Alternative to Standardized Tests." Presentation at the annual meeting of the Higher Learning Commission, June 13, Chicago, IL.

Pierrakos, O., Nagel, R., Watson, H., et al. 2013. "Not All Problems Are Created Equal: Intentional Design of Learning Experiences Grounded on an Innovative Problem Based Learning Model." Presentation at the Conference on Higher Education Pedagogy, Virginia Polytechnic Institute and State University, February 6–8, Blacksburg, VA.

Seifert, L. 2012. "Assessing General Education Learning Outcomes." *Peer Review*, Fall 2011/Winter 2012: 9–11.

Sullivan, D. F. 2014. *It's Time to Get Serious about the Right Kind of Assessment: A Message for Presidents*. Retrieved from <http://www.aacu.org/value/right-kind-of-assessment>

Mary Kay Jordan-Fleming is academic assessment coordinator and professor of psychology at Mount St. Joseph University in Cincinnati, Ohio.

Supporting and Evaluating Learning Outcomes Assessment through Academic Program Reviews

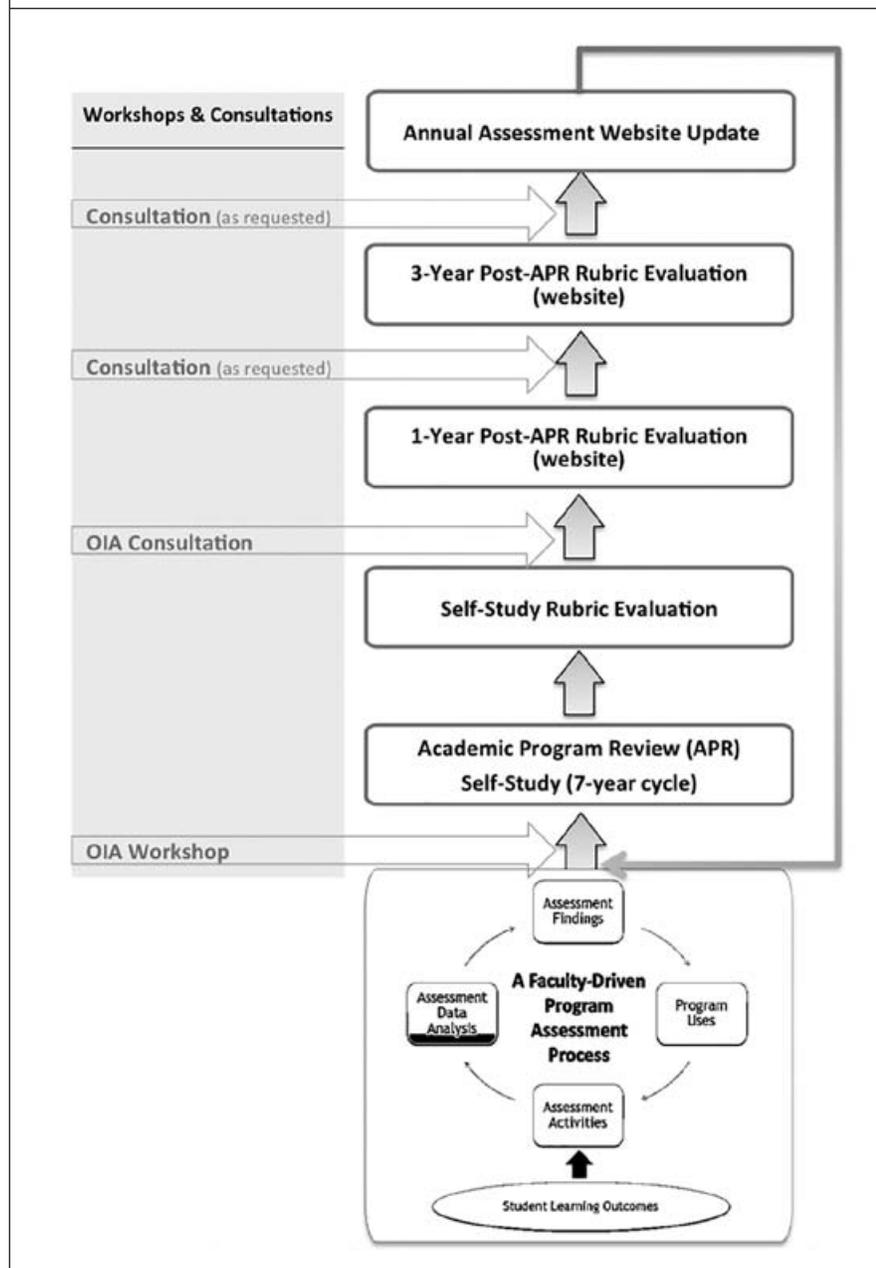
(continued from page 4)

those practices. The requirement thus became the *stick* in our approach. However, we knew that programs would need help implementing and sustaining learning outcomes assessment. Thus, the APR could be an entry point for the *carrot* part

of the approach (e.g., consulting activities from the Office of Instruction and Assessment [OIA] staff members). The resulting carrot-and-stick approach for operationalizing the UA outcomes assessment model is illustrated in Figure 2.

Each of the programs preparing for an APR in a given year is invited to attend an OIA workshop on learning outcomes assessment. Workshop participants examine the rubric that will be used to score their learning outcomes assessment plans and practice scoring two sample plans. After the workshop, OIA assessment staff members consult with programs about best practices in outcomes assessment and the assess-

Figure 2: The APR-Enabled 7-Year Assessment Improvement Cycle (UA Assessment Website 2010).



ment sections of their APR Self-Study Reports.

Once a program submits its APR Self-Study Report, OIA assessment staff members review the sections that describe learning outcomes assessment in the undergraduate and graduate programs and score these assessment plans using a rubric with four criteria: Expected Learning Outcomes, Assessment Activities, Assessment Findings, and Changes in Response

to Findings. Whether a program has updated assessment information on the UA assessment Web site is also noted on the rubric. Examples of outcomes assessment reporting are available at http://assessment.arizona.edu/learning_outcomes under Academic Programs in the left sidebar. To see the rubric used by OIA staff members, go to <http://assessment.arizona.edu/node/1376>. The scored rubrics are sent to program heads in advance of their

APR site visits. The scored rubrics are also sent to the senior vice provost for Academic Affairs, the administrator in charge of APRs and campus-wide outcomes assessment. In this way, learning outcomes assessment becomes a key item for discussion during the final APR meeting, which includes the program head, appropriate dean, senior vice provost, and the provost.

After the APR final meeting, OIA assessment staff members consult with program faculty to develop post-APR improvement plans. The plans are shared with the appropriate associate dean and the senior vice provost. Again, OIA assessment staff members are available to consult with programs on an as-needed basis throughout the post-APR year. At the end of that year, a program's updated information on the assessment Web site is scored using the same rubric, and the heads and their associate dean receive copies of the scored rubrics. At this point, the scored rubrics can be used by a unit to gauge improvement in outcomes assessment practices.

We expect programs to update their pages on the UA assessment Web site annually; we check for updates each summer. The next OIA evaluation occurs three years after the APR, at which point the relevant pages on the assessment Web site are scored again, with rubric scores sent to program heads and their associate deans. This midcycle check should support programs in sustaining best practices in learning outcomes assessment. We also expect the associate deans to monitor their programs' progress on assessment.

A Comparison of UA's Assessment Reporting with Peer Institutions

We believe our open approach to reporting assessment plans and updates is unique, at least among our peer institutions. Nearly all of our peer institutions have assessment sites accessible by searching for "academic program assessment" from the main institutional Web pages. Similar to our assessment process model (see Figure 1), the University of California, Davis conceptualizes their assessment process

through stages including plan, inquire, analyze, and apply (Regents of the University of California 2014). The University of Iowa Office of Assessment Web site includes the Assessment Exchange, an online database for departments to review other departments' assessment plans and updates (University of Iowa n.d.). However, these materials are available only through network log-in. Academic program learning outcomes were readily available for review at other institutions' sites (see University of California at Los Angeles and University of Florida).

An important aspect of our assessment approach is the open access that everyone has to each of our academic programs' student learning outcomes, assessment activities, assessment findings, and changes in response to findings. This open access to assessment information should encourage accountability for units to sustain their outcomes assessment practices. On our assessment site (<http://assessment.arizona.edu>), a unit's outcomes assessment information is available by clicking Academic Assessment, and then Academic Programs; a list of colleges is displayed. From there, users can choose a college and then a unit.

Each academic unit is further organized into pages for its undergraduate and graduate programs.

Tracking Program Improvement Across Time

At this point, four cohorts of programs have gone through the APR process requiring the reporting and scoring of learning outcomes assessment plans. Figure 3 shows the average rubric scores on the four criteria for each of the cohorts.

We are pleased to note a gradual improvement over four years in the quality of programs' learning outcomes, while acknowledging more work to be done on other criteria. (The drop in Cohort 4's average scores for the last three criteria is primarily due to modification of the rubric criteria.) However, we have observed that once a program has identified clear and measurable outcomes, the other components of the model develop over time. To illustrate this point, Figure 4 shows the average APR and one-year post-APR rubric scores for the first two cohorts, both of which have completed their post-APR years. Cohort 3's post-APR evaluation will be completed during the summer of 2015.

All members of Cohorts 1 and 2 demonstrated improvement in Outcomes and Assessment Activities from their APR to one year post-APR scoring. We anticipate that these improvements will be reflected in Assessment Findings and Changes in Response to Findings when these cohorts' Web pages are again reviewed three years after their APR year (summer of 2015 for Cohort 1). We acknowledge the importance of providing time for sufficient data collection before findings can lead to productive faculty discussions and decisions about changes.

In addition to the changes in rubric scores across time, we are encouraged by the increased number of programs that now seek our assistance with assessment the year before their APR year. During the 2014–2015 academic year, we offered a workshop series to Cohorts 5, 6, and 7; 64 percent of programs in those cohorts attended at least one workshop. Also, for 13 of the 43 programs in Cohort 3 (2013–2014), the outcomes assessment plans presented in the APR Self-Study Reports were sufficiently strong to warrant elimination of the post-APR consultations with them.

Figure 3: Average Rubric Criteria Scores for Four Cohorts.

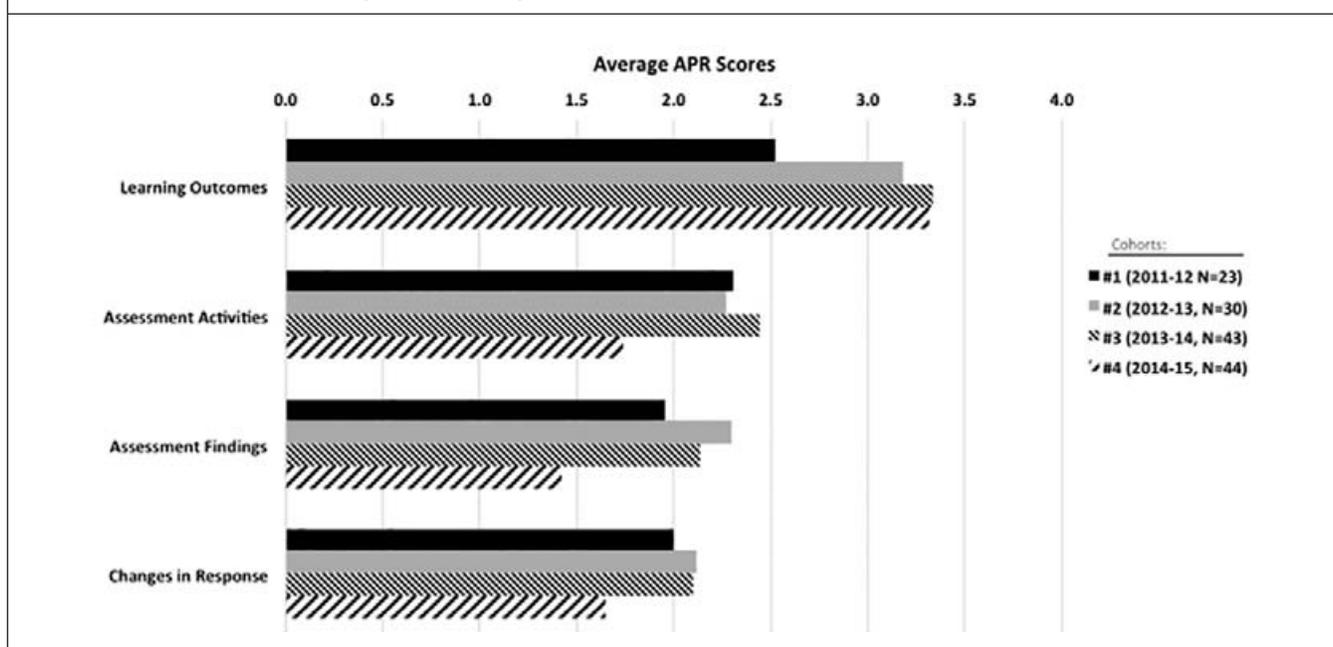
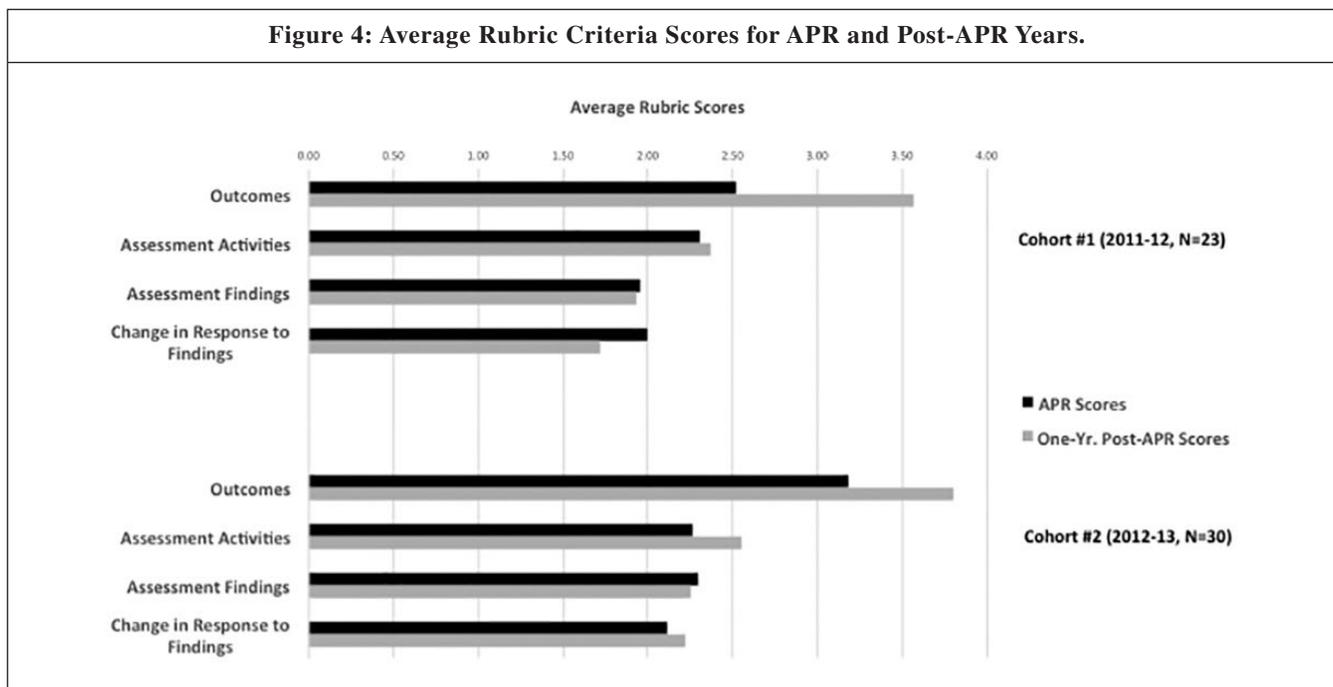


Figure 4: Average Rubric Criteria Scores for APR and Post-APR Years.



Lessons Learned

The role of the stick. The APR *stick* is a necessary incentive for programs to practice the UA outcomes assessment model. Without attachment to the highly regarded APR, many of our academic programs would revert to inadequate assessment practices. However, many faculty and department heads have told us that without the guidance provided by the OIA assessment team members before and after receiving their rubric-scored evaluations, they would have been challenged to make the necessary improvements.

The importance of the carrot. The coaching that the OIA assessment team members provide to faculty and department heads is very important. We hold consulting meetings to provide feedback on faculty members' ideas or newly created tools (e.g., rubrics) and negotiate steps on programs' required improvement plans. These coaching actions, our *carrot* activities, create goodwill that translates into willingness to work toward a common goal.

Graduate programs' discomfort with outcomes assessment. Faculty members are particularly uncomfortable with applying learning outcomes assessment to

their graduate programs. They feel that graduate student learning cannot be summarized in a single set of outcomes and one-size-fits-all assessment activities. Yet the one-size-fits-all model is exactly how assessment activities are conducted in most units (e.g., comprehensive exams, dissertation defense). Faculty members usually have clear ideas about what constitutes satisfactory learning during these valued assessment activities, although these ideas are not often stated in writing. Providing examples of how outcomes assessment can link to the assessments already practiced in the unit often results in a willingness to try our suggestions. Examples include sample rubrics from other units to assess outcomes at assessment events such as comprehensive exams. Furthermore, we have learned that it is important for us to talk through these examples with the faculty. For an example, see http://assessment.arizona.edu/information_sheets and select *Example: A Graduate Program Outcomes Assessment Plan and a Sample Rubric*.

The value of OIA staffers' faculty experiences. The OIA assessment team members are experienced faculty members with current or previous UA faculty positions. Additionally, we have had ex-

periences working with department-level outcomes assessment while in these faculty positions. Therefore, we share an understanding of the challenges related to faculty workload, balancing professional responsibilities, and uneasiness with an assessment process that feels different from more familiar classroom assessment practices. We believe that the shared understanding helps to build trust in addition to the goodwill created by our *carrot* activities.

Efforts to build awareness of outcomes assessment beyond UA. Our process allows us to showcase the rubric-scored evaluations of a program's assessment practices to the APR reviewers during the site visit. The senior vice provost provides team members with copies of the rubric scores and a description of what the scores mean. Review team members are generally unaware of the priority given to outcomes assessment at our institution and have never indicated that rubric evaluations like ours are used at their institutions. We believe that this action, taken by our high-level administrators, has the potential to build awareness of the importance of outcomes assessment among colleagues from peer institutions. ■

References

- Banta, T. W., Jones, E. A., and Black, K. E. 2009. *Designing Effective Assessment: Principles and Profiles of Good Practice*. San Francisco, CA: Jossey-Bass.
- Maki, P. L. 2004. *Assessing for Learning: Building a Sustainable Commitment across the Institution*. Sterling, VA: Stylus.
- Suskie, L. 2009. *Assessing Student Learning: A Common Sense Guide* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Regents of the University of California. 2014. *Assessment Stages*. Retrieved from <http://cetl.ucdavis.edu/services/consultations/assess/stages/index.html>
- University of Arizona. n.d. Assessment Web site. Retrieved from <http://assessment.arizona.edu>
- University of Iowa. n.d. *Assessment Exchange*. Retrieved from <http://www.uiowa.edu/assessment/learning-outcomes-reports#AssessmentExchange>
- Walvoord, B. E. 2010. *Assessment Clear and Simple: A Practical Guide for Institutions, Departments, and General Education* (2nd ed.). San Francisco, CA: Jossey-Bass.

Ingrid Novodvorsky is director of Teaching, Learning, & Assessment in the OIA; Debra Tomanek is the associate vice provost for Instruction & Assessment; Ryan Foor is assessment coordinator in the OIA; and Gail D. Burd is the senior vice provost for Academic Affairs at the University of Arizona.

Editor's Notes

(continued from page 3)

the purpose of eliciting the intended skills? Jordan-Fleming reports that faculty interest in using course-embedded methods to assess learning outcomes in the general education curriculum at Mount St. Joseph University has alerted her colleagues to threats to the reliability and validity of rubrics. This

finding has increased recognition of the importance of faculty development related to assignment writing.

Embedding assessment in courses (as opposed to taking assessment out of the classroom, as could be the case with a standardized test), involving students as partners

in assessment, evaluating the effectiveness of linking courses in a core curriculum, and making assessment of learning a required component of program review are all processes that can capture faculty interest and engagement in outcomes assessment. As the articles in this issue illustrate, each of these processes can produce opportunities for sharing good practice in teaching and learning, both informally and in more extended faculty development initiatives. ■

Assessment: Learning by Doing

(continued from page 2)

higher on all questions related to integration and community, with the gap between the nonlinked and the linked ratings greatest for items such as these:

- I often thought about the connections among my T&E classes.
- I could use what I learned in my English 101 class in my Philosophy 101 class (and vice versa).
- The assignments in one of the T&E classes supported the content of the other two courses.

Reporting the survey data to different audiences has yielded benefits for instructional practice and curricular planning. In 2011, after collecting the data for five semesters, statistically oriented colleagues and I shared it with the linking faculty. At a meeting devoted to “best practices,” teachers whose scores had been consistently high on particular survey items were invited to talk about the instructional approaches that might have contributed to

those results. The discussion was important not only for the particular teaching strategies that were exchanged but also because it demonstrated to faculty that assessment data can be meaningful. Writing a program review two years later, I could point to the survey to show strengths that could be developed. Further, for the committee tasked with formalizing a proposal for revising the core, the survey data corroborated consistent positive reports from students and faculty and influenced the group’s recommendation to retain linked courses in a new core.

These initial assessment efforts provided information and experience for us to build upon. Having agreed upon student learning outcomes for the current core, we are now in our third year of direct assessment, even as we have continued planning a revised core. Beyond providing evidence of student learning, our assessment work over the past several years has taught us

ways to share results with different audiences, to build a comprehensive assessment plan into a curriculum, and to keep the focus on student learning. We are in a better position for thorough and meaningful assessment in the new core, with program and course learning outcomes, the opportunity for faculty to create embedded assessments, and a program assessment plan that aligns with course-level assessment. As we plan implementation of the new core, we continue to learn by listening to our colleagues on other campuses, but there is no substitute for the valuable lessons that we learned by doing. ■

Reference

- Aristotle. *Nicomachean Ethics*. n.d. Perseus Digital Library. Retrieved from <http://data.perseus.org/citations/urn:cts:greekLit:tlg0086.tlg010.perseus-engl:1103a>

Patricia Terry is professor of English and associate dean in the College of Arts and Sciences at Gonzaga University in Spokane, Washington.