Assessment UPdate
January–February 2012
Volume 24, Number 1
Progress, Trends, and Practices in Higher Education

Expectations for Assessment Reports: A Descriptive Analysis

Keston H. Fulcher, Matthew Swain, Chris D. Orem

Resources designed to assist institutions with their outcomes assessment endeavors abound. Books, workshops, and conferences provide tips ranging from writing good objectives to mobilizing resources for improving assessment processes. From our perspective, one of the gaps in these resources is information about how to write a good assessment report. Such reports are often a program's best, or at least most official, vehicle for conveying information about student learning and the program experiences that contribute to such learning. Further, the best reports reveal insights into decision-making, particularly the factors that influence faculty to modify curriculum and pedagogy to improve student learning. This information is important within a program and to university administrators and external stakeholders such as regional accreditors. Responding to this need, a small but growing number of universities have developed rubrics to evaluate assessment reports. Ory (1992) referred to such evaluations of assessment as *meta-assessment*. This promising practice represents a shift in thinking by institutions from "Do our programs participate in assessment?" to "How well do our programs conduct assessment?" In this article we provide a descriptive analysis of how institutions operationally define good assessment via these meta-assessment rubrics (MARs).

Our experience suggests that assessment reports typically underrepresent both the strength of assessment within programs and the thoughtfulness of faculty in identifying and resolving problems related to student learning. If this is the case, faculty may ask how to write reports that fairly reflect their programs' assessment processes.

This inquiry is similar to a familiar classroom situation. Students ask faculty for advice about how to write a good paper. Without guidance students complain about their inability to read the teacher's mind. Conversely—and not surprisingly—a greater proportion of students submit papers that are consonant with expectations when the instructor provides clear directions. In the classroom context, many instructors use rubrics to convey expectations. A well-developed writing rubric, for instance, lists the criteria to be evaluated (e.g., content, organization, mechanics) and provides behavioral descriptions associated with levels of performance for each criterion such as *beginning, developing, accomplished,* and *exemplary*.

Institutional leaders have applied an analogous strategy to evaluate assessment reports. They (a) construct a MAR that specifies the characteristics of a good assessment report; (b) share the MAR with programs prior to report submission; and

Contents

ARTICLES

Expectations for Assessment Reports: A Descriptive Analysis 1
Keston H. Fulcher, Matthew Swain, Chris D. Orem

Editor's Notes 3
Trudy W. Banta

Getting SMART with Assessment: ACTION Steps to Institutional Effectiveness 5
Eric Daffron, Sandra J. Jordan

Borrowing Techniques from Computational Linguistics to Process Qualitative Data 7
Forrest Houlette

Department Chair Perspective on Learning Outcomes Assessment at Kansas's Public Universities 9
C. B. Crawford, Larry Gould

COLUMN

Assessment Measures 10
Gary R. Pike

View this newsletter online at wileyonlinelibrary.com
doi:10.1002/au.241
© 2012 Wiley Periodicals, Inc.

Assessment Update

Progress, Trends, and Practices in Higher Education

January–February 2012
Volume 24, Number 1

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
Frances W. Oblander, *director of institutional research & effectiveness, South 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, *La Trobe University, Victoria, Australia*
T. Dary Erwin, *James Madison University*
Cecilia L. Lopez, *Harold Washington College*
Marcia Mentkowski, *Alverno College*
Jeffrey A. Seybert, *Johnson County Community College*
Peter J. Gray, *United States Naval Academy*
Gary R. Pike, *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 1200, San Francisco, CA 94104-4594. Periodicals Postage Paid at San Francisco, CA, and additional mailing offices. Individual subscriptions are $127 per year (institutional $211). Back issues are available in limited supply at $29 per issue. To order, phone toll-free (888) 378-2537 or fax (888) 481-2665. Visit our Web site at www.josseybass.com. *Postmaster:* Send address changes to Assessment Update, Jossey-Bass, One Montgomery St., Suite 1200, San Francisco, CA 94104-4594.

Copyright © 2012 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.

(c) systematically assign ratings to assessment reports based on the MAR. The results from this meta-assessment are typically used for two purposes. Results can be used to facilitate diagnostic feedback to an individual academic program. They can also be aggregated across programs at the college or university level to indicate overall assessment strengths and weaknesses. Examples of these uses are emerging in the literature (e.g., Bloom, 2010; Fulcher and Orem, 2010).

While conducting our own meta-assessment at James Madison University (JMU), we wondered how other institutions operationalized good assessment through rubrics.

> The best reports reveal insights into decision-making, particularly the factors that influence faculty to modify curriculum and pedagogy to improve student learning.

Specifically, what criteria for assessment were most common? In the next section we provide a short description of the method used to investigate this question followed by general results and discussion.

Method

We attempted to procure an exhaustive set of MARs throughout the United States using three strategies: (a) consulting the North Carolina State University's Internet Resources for Higher Education Outcomes Assessment (following Ephraim Schechter's suggestion), (b) requesting MARs from assessment practitioners on the ASSESS Listserv (Fulcher, 2010), and (c) scouring the Internet using a variety of search terms such as "assessing assessment," "evaluating assessment rubric," "meta-assessment rubric," and others. From these sources we located 51 MARs.

(continued on page 14)

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 140 Administration Bldg., 355 N. Lansing St., Indianapolis, IN 46202–2896. ■

Editor's Notes

A Surprising Reaction

Trudy W. Banta

The new degree qualifications Profile (DQP) provided the topic for discussion at the opening session of the 2011 Assessment Institute in Indianapolis on October 30. The DQP was the choice of Cecilia Lopez, associate vice chancellor for arts and science for the City Colleges of Chicago (CCC), who served as provocateur for the panel that included Peter Ewell of the National Center on Higher Education Management Systems, George Kuh of the National Institute on Learning Outcomes Assessment, Tom Angelo of LaTrobe University in Australia, Jeff Seybert of Johnson County Community College, and me.

Cecilia suggested that we focus on the DQP as a result of watching reactions to it among faculty in her consortium of two-year institutions. She told us via e-mail before the Institute, "Just getting their hands on the DQP has energized the CCC assessment chairs in a way I find fascinating. Hearing faculty tumble all over themselves talking and sharing what they are doing keeps me going! I've ordered enough DQP copies for all CCC faculty and have offered our district as a place for pilot testing DQP-inspired measures."

So what is the DQP? During my opening remarks, I asked the roomful of 980 academics sufficiently interested in assessment to attend the Institute how many knew enough about the DQP to explain it, and only about 30 people raised their hands. Clearly, the word about the DQP and its potential to influence assessment has not yet reached a broad segment of higher education.

The Lumina Foundation for Education has sponsored the development of the Degree Qualifications Profile, and Cecilia drew on the Lumina report with that name (2011) for her definition: A framework for illustrating clearly what students should know and be able to do once they earn a degree at any level—associate, baccalaureate, or master's—regardless of concentration, major, or field of specialization.

From the Lumina report, the DQP defines competence "in ways that emphasize both the cumulative *integration* of learning from many sources and the *application* of learning in a variety of settings." Thus, the DQP "can offer benchmarks for improving the quality of learning." In addition, since "every stated learning outcome should lead to and support assessment," the DQP "is also designed to encourage colleges and universities to enhance their assessment practices and/or develop new assessments" (p. 2).

The Lumina Foundation's Big Goal is to raise the percentage of baccalaureate degree holders in the US to 60 by 2020. Clearly this number should not be achieved by lowering standards. Ensuring degrees with integrity is a critical component of Lumina's goal. Again quoting the 2011 Lumina report, "Use of the Degree Profile over time should yield . . . reference points for accountability that are far stronger than test scores or tallies of graduates, research dollars, student satisfaction ratings, job placements, or patents" (p. 2).

Just before the Assessment Institute, I read a long-overdue annual report on assessment in IUPUI's School of Liberal Arts. It contained a section I had not seen before outside our Schools of Nursing and Medicine: a "cumulative *integration* of learning," showing what graduates of the school "should know and be able to do once they earn a degree at any level: associate, baccalaureate, master's, or PhD"! I was encouraged because this signaled the possibility of a new era of attending to the cumulative nature of learning instruction and assessment in our School of Liberal Arts. This strengthened my impression that the focus on the DQP in the opening session of the Institute would be relevant, timely, and helpful for our participants.

We have a tradition of asking for questions and comments on 5" x 7" cards at the end of the opening plenary session. A second panel provides responses to a

The DQP defines competence "in ways that emphasize both the cumulative *integration* of learning from many sources and the *application* of learning in a variety of settings."

selection of the submitted questions at the opening session of the Institute on the second morning.

Cecilia, Peter, George, Tom, Jeff, and I left the stage at the conclusion of the opening session pleased that our discussion following Cecilia's introduction had proceeded smoothly. A few hours later, however, we were surprised to read some of the comments participants had written on their 5" x 7" cards:

1. How much is Lumina paying you all to capture this audience and sell us on this tool? Please do not think this question is snide . . . think "transparency."
2. Peter Ewell indicated the DQP was designed to be "uniquely American." In the recent years of this conference I have seen this distinguished panel focus more and more on centralizing the profession . . . central outcomes . . . and yes broad but still, nearly universal outcomes. To me that undermines the greatest strength of American higher education—its diversity. How does this approach honor the diversity of American higher education? How is Lumina's financial impetus and involvement in accreditation a good thing? This sounds like "lining everybody up" and making us all goosestep to the same outcomes. How is such centralizing and philosophical homogeneity a good thing?
3. The initial proposition about DQP was as a common communication tool to stakeholders about the nature and value of the degree, but the conversation suggests it becomes an institutional template for curricular and assessment reform. Which is the cart and which is the horse? If this is not an accurate statement about the degree as it currently exists, is it desirable that institutions in general all aim to remake themselves in its image?

Hmmm . . . Since the DQP is a concept, not an assessment tool, and is available to all free of charge, we had not considered the possibility that we would be perceived as selling a product. And since the Lumina Report says "every institution should expand this Degree Profile by adding outcomes that are specific to its mission and by aligning them with assessments in use or under development" (p. 2), we do not view the DQP as an attempt to homogenize higher education in the US.

The DQP emphasizes the "cumulative integration" of knowledge by using progressively more complex levels of learning derived from Bloom's Taxonomy of Educational Objectives (Bloom, 1956) to describe student learning outcomes at associate, baccalaureate, and master's levels. At the associate level, the learning outcomes listed as examples begin with the verbs 'describe,' 'illustrate,' and 'generate' (Lumina, 2011, p. 10). This occasioned the following comments from three Institute participants who represent two-year institutions:

1. Why does the DQP assume/state that only the lowest level of learning and generating knowledge on Bloom's Taxonomy occurs at the community-college level?
2. How can the Degree Profile represent more effectively the levels of learning taking place at the associates level? In our classrooms, students often move to higher levels of Bloom's Taxonomy as our students construct, evaluate, produce, and differentiate, to name a few of Bloom's verbs that are relevant.
3. We pride ourselves on "high standards," which means we teach *some* things at a baccalaureate or even a master's degree level (while other things are taught at the high school level). How do we convince faculty to at least grade the "C" level by the associates standards of the DQP without being accused of lowering standards?

Obviously most of us on the panel were surprised at the number and intensity of the negative comments about the DQP. The concept is barely a year old (published in January 2011), and clearly it is not yet widely understood.

What is your experience in thinking about/using the DQP? What are its advantages/possibilities or disadvantages/challenges from your perspective? Please send e-mail to tbanta@iupui.edu. I would be pleased to publish some of your comments. ■

References

Bloom, B. S. (1956). (Ed.). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain.* New York, NY: Longmans, Green.

Lumina Foundation for Education. (2011). *The Degree Qualifications Profile*. Indianapolis, IN: Author.

Getting SMART with Assessment: ACTION Steps to Institutional Effectiveness

Eric Daffron, Sandra J. Jordan

Universities hire, promote, and recognize faculty for their disciplinary expertise, their scholarly accomplishments, their teaching effectiveness, and their ability to contribute to the mission of the institution. But the same faculty who shine in the traditional areas of teaching, research, and service may not have the interest, knowledge, or training to engage successfully in institutional effectiveness activities. Other pressing priorities—preparing for classes, writing grants, serving on committees, and conducting research—make it difficult to find more time to learn the finer points of assessment and continuous improvement. Additionally, faculty often finds assessment terminology foreign and assessment processes complex and time-consuming. To complicate matters, faculty and even administrators may misunderstand the purpose of these activities. Considering the evolving and sometimes conflicting information about accountability in the academic press and elsewhere, faculty are left to wonder whether assessment results will inspire meaningful discussions about improving student learning and institutional operations, or whether their results will go quietly into an administrator's files or become the basis for unfair judgments about them or their institution from within or without. In addition, faculty and administrators often confuse the assessment of student learning with other aspects of institutional effectiveness. As a result, they use indirect measures, such as graduation rates or attendance at lectures, as proxies for the direct assessment of student learning. The combined effect of all these issues may be an institutional effectiveness plan that is truly ineffective.

Our collective experience at several colleges and universities has convinced us that institutions can minimize, if not eliminate, many of these problems if they at least separate the assessment of student learning from the assessment of other aspects of the institution. Once separated, these two parts of institutional effectiveness assume their own processes and time lines as well as separate networks of training, feedback, and support. In the end, the two parts should recombine in a holistic system of institutional effectiveness linked to periodic program review and resource allocation.

Faculty often finds assessment terminology foreign and assessment processes complex and time-consuming.

Establishing and supporting effective institutional assessment and planning begin with communicating the challenges, generating interest, suggesting solutions, and linking those solutions to best practices, articulating the value of assessment and planning, and finding the right campus groups and individuals to move the projects forward. Key to identifying the right travel companions is explaining how institutional effectiveness activities will contribute to improvements that are meaningful to stakeholders. Individuals leading institutional effectiveness activities should understand the unique traditions of their institution, as well as best practices of the academy, and should know when to be flexible and where to compromise. Additionally, as assessment teams are built, it is wise to learn what the various members of the team value. Administrators want to meet accreditation standards, to maximize effectiveness, and to create processes that lead to improvement and reportable results. Faculty members on the team may be preoccupied with other priorities, including meeting promotion, tenure, and merit criteria; preserving time for teaching and research; maintaining independence; and designing processes that are not cumbersome and time-consuming—yet lead to improvements.

Assessment teams will more likely work effectively if they understand the difference between the assessment of student learning and other forms of institutional assessment. In our experience, it can be helpful to uncouple the processes and differentiate learning assessment from program evaluation. The process of learning assessment begins with establishing the institution's overarching characteristics of graduates. The broad characteristics become essential learning goals agreed upon by the academic stakeholders and spanning across all baccalaureate programs at the university. Then department faculty set their own goals and outcomes by engaging in department-wide discussions. After departments identify those goals and outcomes most significant to their programs,

they complete assessment plans and submit those plans to the assessment committee.

An assessment reporting format called SMART uses a memorable acronym to help communicate the process:

S=Sharpen learning goals into outcomes and set achievement targets (e.g., the percentage of majors who meet an outcome at a satisfactory level or higher) to which the department aspires.

M=Map the outcome across the program curriculum. A simple matrix showing the relationship between courses and outcomes may reveal that an outcome central to a department's mission is nowhere addressed in its curriculum.

A=Assess the outcome. How precisely will the department measure the outcomes that it has selected?

R=Review the results. This section lists the actual findings of the assessment.

T=Transfer the information into action. Often called "closing the loop," this section details how the department plans to improve its assessment methods or the curriculum in the future.

The SMART form, which includes instructions, is provided electronically with hyperlinks to definitions and examples. For the planning process, only the S-M-A sections of the form are completed. After completing their assessments, departments report their results by completing the R and T sections of the form.

An Assessment Committee is central to the process outlined above. The committee reviews the plans and provides each department with suggestions for improvement, recommendations for alignment and efficiency, and connections between units assessing similar outcomes. When the year's activities are reported, the Assessment Committee offers encouragement, compiles results, and reports institutional progress in the aggregate. Aggregating findings has the virtue of protecting individual departments from undue public exposure while keeping the entire institution abreast of assessment activities and their results. After all, no department likes to spend time and energy on assessment without the assurance that other departments are also participating and that results have been put into meaningful action. The Assessment Committee, though crucial to the success of any assessment program, is just one part of the larger assessment agenda. A culture of assessment includes development opportunities for faculty, alignment with faculty evaluations, a communication plan, and celebrations of success.

Like the assessment of student learning, the assessment of other parts of the institution should bear a brief name or a snappy acronym that reminds individuals of the process and helps them to distinguish it from SMART. Take, for instance, ACTION, an acronym that stands for the various steps in the following process (which is not identical to, but rather a convenient amalgam of, processes that we have used in the past):

A=Align goals with strategic directions. In this system, institutional stakeholders establish six-year strategic directions, the President's Cabinet sets annual priorities within those strategic directions, and divisions such as academic affairs align two-year goals with those strategic directions. By setting two-year goals, divisions make short-term commitments to more focused parts of larger strategic goals and, based on biannual achievements and evolving internal and external conditions, adjust those goals over the six-year period.

C=Create objectives and action steps. Colleges (to use academic affairs as the leading example) set annual objectives in response to divisional two-year goals, and departments within each college set annual action steps to fulfill those objectives. As a result, each unit customizes the direction, goal, or objective set above it in a way that makes sense for that unit. Even as the original strategic direction becomes more individualized, every action step should clearly correspond to an objective, which, in turn, should correspond to a goal, and so forth, so that no matter how customized a strategic direction becomes, every unit in the institution is committed to fulfilling it.

T=Tie action steps to targets. As each department establishes a specific action step that it will take in support of the objective set at the college level, it also sets a target to which it aspires and against which it can measure the action step's effectiveness.

I=Identify measures. Often the most challenging part of institutional effectiveness, setting a good measure, is also one of the most crucial steps. Sometimes departments begin with simple measures, such as project completion, but over time, they should mature in their planning by adding attendance counts, satisfaction surveys, and increasingly more sophisticated measures. Depending on the action step, departments may use data from SMART to support ACTION. For example, if a French department plans to refurbish a language lab, it could set project completion as well as increased achievement of language outcomes as measures. Using SMART to support ACTION not only demonstrates sophisticated planning practices but also illustrates the integration of various planning processes that we advocate.

An assessment reporting format called SMART uses a memorable acronym to help communicate the process.

While A, C, T, and I belong to the planning phase of the process, O and N belong to the reporting phase.

(continued on page 13)

Borrowing Techniques from Computational Linguistics to Process Qualitative Data

Forrest Houlette

Analyzing qualitative data collected for assessment requires coding of the information to allow classification of the data into categories (Rogers & Goodrick, 2010). For example, students frequently report on our student survey that they experience the "runaround" at Sullivan University. To assign a definition to this term and to assess the programs related to this issue, we asked a group of students to write about the last time they had received the "runaround." Or, if they had never had the experience, we asked them to write about an example of good customer service they had received. On completion of this exercise, we were left with a stack of stories. We faced the question of how to interpret the stories in a way that would lead to actionable assessment.

To solve the problem of interpretation, we turned to techniques that have arisen in computational linguistics. In this domain, the problem is how to make sense of large volumes of text without human intervention. Three techniques—entity recognition, theme extraction, and event segmentation—allow a computer to find sense in the text and then make use of that meaning to achieve particular goals. For example, a computer could process thousands of police reports and plot a map of where specific types of events like injury traffic accidents occur most frequently. Fortunately, the human brain is extremely capable of performing the same type of processing. For our limited data set, engaging a computational linguist to solve the problem would have been overkill. Teaching coders to apply the techniques is easy and very straightforward.

First, the coder reads a document and makes a two-column table for each type of analysis. In the left column, the coder records the item found, giving it a name. In the right column, the coder tracks the frequency of occurrences of the item. Coding involves reading the document, recording the items, and placing a tick mark in the right column to track the frequency of the item. After coding, analysts make use of the tables to create descriptions of the data. In our case, we wanted to move from a set of stories to specific descriptions about who was involved, what they were doing, and how often they were doing it. We felt that this approach would give some meaning to the term "runaround." More importantly, we wanted some insight into how to control the process in order to improve student life.

Coding involves reading the document, recording the items, and placing a tick mark in the right column to track the frequency of the item.

We began with named entity recognition. Named entities are things that are atomic in nature (Molla, 2008). That is, like an atom, they cannot be subdivided without changing what they are fundamentally. Named entities are the important persons, places, and things in the text—entities that could be identified with a proper noun. Our coder found mainly office names, some people's names, and instruments. We identified these offices: Admissions, Advising, Career Placement, Enrollment Services, Financial Planning, Housing, and Scheduling. Financial Planning was mentioned the most often. We found one instrument: paperwork. Whenever the students experienced either good customer service or the "runaround," paperwork was involved. Students mentioned very few names of people, and we found the people they did mention were not relevant to understanding the problem.

Next, we worked with event segmentation. Events are blocks of text that describe a coherent sequence of activities (Zacks & Swallow, 2007). An event has a beginning, a middle in which activity takes some direction, and an end. We found several events and gave them appropriate names: an employee did not possess the right information; promises are not fulfilled; housing rules are too restrictive; the employee solved my problem. We recorded the frequencies with which the events occurred. On the "runaround" side, we found that missing information and promises not fulfilled were the most frequent events.

After events, we turned to themes. Themes are topics that are repeated often enough to achieve significance (Paradis & Berrut, 1996). A theme that repeated in the papers reporting problems was staff turnover. Typical statements like the following appear: The person who initially helped me left the university, and someone who knew little about my situation or how to resolve it took over. A second theme that appeared in the data was lost or mishandled paperwork.

Items mentioned most were financial aid forms and transcripts. A significant theme that appeared was the financial aid hold. During this event, the university prevents a student from attending class until the hold is cleared. Having to miss class because of a hold resulting from lost or mishandled paperwork clearly evoked frustration. The final theme that emerged from the data was miscommunication. Students attempt to get answers and cannot get the answers they need. They cannot find staff members who have the answers. They report having to involve entities outside the university in order to get information into the hands of the staff who need it.

By this point, we had some better understanding of the elements that contribute to our students' perceptions of the "runaround." However, we needed to connect entities, events, and themes in meaningful ways. To make these connections, we used the deep case hypothesis to build links among items. (Fillmore, 1967, provides the most accessible explanation.) This hypothesis assumes that verbs cast roles for the nouns associated with them. The verb *break*, for example, casts roles for an object, an agent, and an instrument. Some agent used an instrument of some kind to affect the object, thus breaking it. We chose verbs that were associated with the experience, and we assembled sentences to describe the experience captured in our stories.

We learned through this process that the "runaround" is a very complex experience. The Financial Planning office was the office most mentioned as giving the "runaround." But it was also the office most mentioned as giving good customer service. As we compared the sentences that described the two experiences, we discovered that the sentences were in fact describing the same experience. The sentence for both experiences begins as follows:

> Somehow Financial Planning lost my paperwork . . .

How the story proceeds from that point is different depending on the student's personality and point of view at the time of the experience. The sentence can end in two ways:

> . . . and I got a financial aid hold and that was the worst experience I have had at school.
>
> . . . and they called me in and took care of it right away and I was only on hold a day or so.

We learned that the "runaround" cannot be solved by investing in customer service training alone. We also need to teach students ways to cope with the university bureaucracy. Our assessment led to action plans to address the issue on multiple fronts.

Stakeholders focused on building customer service and customer management skills. Staff took the Dale Carnegie Course on Communications and Human Relations and a series of SkillSoft modules. Subsequent data collection from focus groups and from interviews with 23 pharmacy technician students indicated that after a year, our efforts paid off. The pharmacy technician students, when specifically asked about their experience with Financial Planning, reported no "runaround" issues. In fact, they reported far fewer negative experiences of any kind than our initial study would have predicted.

We chose verbs that were associated with the experience, and we assembled sentences to describe the experience captured in our stories.

These four approaches borrowed from linguistics enabled us to code and analyze the qualitative data we collected. What is useful about the tools is that the codings arise from the data. You do not need an external theoretical framework to define how to classify the data. Instead, you have a procedural framework that generates the codings. Once they are generated, any appropriate analytical framework can be applied to interpret the data. This approach allows you to manage qualitative data gathered for assessment in an efficient and effective manner. ■

References

Fillmore, C. (1967). The grammar of hitting and breaking. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in transformational grammar* (pp. 120–133). Waltham, MA: Ginn.

Molla, D. (2008). *AFNER—Named entity recognition.* Retrieved from http://afner.sourceforge.net/index.html

Paradis, F., & Berrut, C. (1996). Experiments with theme extraction in explanatory texts. In *Second international conference on conceptions of library and information science, CoLIS*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.52.7064

Rogers, P. J., & Goodrick, D. (2010). Qualitative data analysis. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (3rd ed., pp. 429–453). San Francisco, CA: Jossey-Bass.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*, 80–84. Retrieved from http://dcl.wustl.edu/pubs/ZacksCurrDir07.pdf

Forrest Houlette is director of Institutional Research at Sullivan University in Louisville, Kentucky.

Department Chair Perspective on Learning Outcomes Assessment at Kansas's Public Universities

C. B. Crawford, Larry Gould

In early 2010, a survey method was designed and implemented to study the perceptions of assessment among department chairs of public higher education institutions in Kansas. Academic departments were identified (N = 194), and each department was contacted through the respective chair. Sixty-five (34%) department chairs responded. A 20-item questionnaire assessed perceptions of assessment. Eight additional questions were related to department characteristics.

Results and Data Analysis

Most department chairs responding have more than ten FTE faculty. Most departments have both on-campus and distance education operations (63%). Most programs offer undergraduate and graduate programs (85%), and a slight majority (57%) of departments graduate fewer than 50 students annually from their undergraduate and graduate programs. Some diversity of discipline is reflected in the departments responding, with social sciences, engineering, sciences, education, and medical fields comprising 75% of the sample. Over 90% of department chairs responding report that the primary generation of student credit hours either comes from major classes or is split between major classes and general education.

Relative to the issue of assessment, 94% of respondents reported that they have at least implemented an assessment plan. Over 80% have collected assessment data, and more than 60% have made program changes based on their assessment findings. Regarding generalized faculty attitudes about assessment, more than 80% of responding department chairs indicated that faculty members are motivated to carry out assessment activities, with the most common response indicating that faculty members have already implemented changes based on their assessment results.

The most favorable response among all the survey items relates to the need to conduct assessment to meet specialized accreditation requirements: 88% indicated strong agreement with the statement. More than 85% of these department chairs base assessment on program-level learning outcomes. Only 3% of department chairs reported disagreement with assessment being an important part of the culture of their institution, and only 6% disagreed with the statement that assessment is an important part of their departmental culture. A mere 3% of respondents said that improvement of student learning is not a high priority, and only 5% indicated they are not comfortable discussing assessment results with other faculty in their department. Over 60% of respondents believe that they will need to increase departmental effort related to assessment over the next two years.

> More than 80% of responding department chairs indicated that faculty members are motivated to carry out assessment activities.

While considerable commitment to assessment is expressed by responding department chairs, there is a less favorable impression of accountability and institutional support for assessment. Some 50% of the department chairs feel pressure from campus administration to improve student learning. While most respondents agree that their respective institutions have provided support for assessment, over 25% disagree. About 25% of respondents indicate that their faculty members do not access training materials related to assessment. Almost 50% of the department chairs do not share learning outcomes assessment results with prospective or current students. A majority of respondents disagreed with the need for greater statewide efforts related to assessment. Nonetheless, over 67% of responding chairs agreed that departments should be reviewed regularly for progress on improving student learning.

Only 10% of the chairs disagree with the use of traditional assessment methods (pretest/posttest, exit exam, capstone performance) and fewer than 15% disagree with the use of rubric-based assessment processes. About 50% of the departments use some external examination to validate student learning and almost 60% use indirect measures as a proxy for direct assessment. Less than 10% of departments

(continued on page 13)

Assessment Measures

Criteria for Evaluating Campus Surveys

Gary R. Pike

One of the first columns I wrote for *Assessment Update* dealt with criteria used to evaluate assessment instruments, particularly standardized tests (Pike, 1989). This year, several articles have called into question the appropriateness of using surveys and students' self-reports of their college experiences in assessment and educational research (see Herzog & Bowman, 2011; *Review of Higher Education*, 2011). Given the widespread discussion about surveys, I thought it would be appropriate to revisit the topic of evaluating assessment instruments and focus on surveys.

Standards for evaluating the validity of educational measures have evolved over time. Early validity frameworks focused on criterion-related validity and adopted the view that validity was a characteristic of the instrument (Cronbach, 1971; Cureton, 1951). More recently, educational measurement specialists have argued that validity refers to the ways in which data are interpreted and used (Kane, 2006; Messick, 1989). According to Messick (1989) and Kane (2006), an assessment instrument can be valid for one interpretation or use—but not for another. It is significant to note that Messick's validity framework served as the foundation for the definition of validity used in the *Standards for Educational and Psychological Testing*. The *Standards* defines validity as ". . . the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 9).

In my original column, I noted that Messick's validity framework contains three criteria for evaluating educational measures. These criteria are the substantive, structural, and external components of validity (Loevinger, 1957). The substantive component of validity asks whether the content of the test or survey adequately represents constructs it is designed to measure. The structural component of validity refers to the extent to which the structure of the scores produced by a test or survey accurately represents the theoretical structure of the measure. For example, if a survey is designed to measure three aspects of the student experience, the data from the survey should represent three dependable scores. The third element in Messick's framework is the external component, and it focuses on the extent to which the relationships between assessment scores and external measures are consistent with theory. This component is similar to Cureton's (1951) concept of criterion-related validity. If scores on a scale are presumed to be related to academic success, then observed scores should be related to measures of student success.

From my perspective, the most important step in evaluating an assessment instrument is defining the uses that are to be made of the assessment data. As Michael Kane observed in his chapter on validity: "The evidence needed for validation necessarily depends on the claims being made. Therefore, validation requires a clear statement of the proposed interpretations and uses" (Kane, 2006, p. 23). Before selecting any assessment instrument, institutions and assessment professionals should clearly identify how the data will be used. Obviously, the recommendations of an instrument's developer should guide decisions about uses of assessment data derived from it. Descriptions of intended use should identify the types of inferences to be made (e.g., are the data to be used to evaluate learning, predict retention, or measure satisfaction) and the appropriate unit(s) of analysis (e.g., students, programs/groups, or the institution). Only after the intended uses of the instrument have been clearly defined can validity be accurately assessed.

A concrete example may help to clarify the importance of intended interpretations and uses of data in evaluating assessment instruments. For purposes of illustration, I compare two popular surveys of students' college experiences: the Cooperative Institutional Research Program (CIRP) surveys and the National Survey of Student Engagement (NSSE). To further simplify the example, I limit my discussion to CIRP's *Your First College Year (YFCY)* survey and NSSE's results for first-year students.

The most important step in evaluating an assessment instrument is defining the uses that are to be made of the assessment data.

The *YFCY* survey is designed to provide information about the adjustment, involvement, and academic and personal development of first-year college students (Higher Education Research Institute, 2011). Results are reported at the institution and student levels. Institution-level data can be used for assessment and accountability reporting, whereas student level data can be aggregated to the program level and used to evaluate the effectiveness of educational interventions such as learning communities or freshman success programs (Higher Education Research Institute, 2011). Student-level data have also been used in a wide variety of research reports focusing on the success of individual students. To facilitate using institutional- and student-level data, the *YFCY* provides a variety of construct scores (e.g., academic adjustment, faculty interaction, and overall satisfaction) developed using item response theory (IRT).

NSSE's survey, *The College Student Report,* was originally developed to provide institutions with measures of the extent to which students on their campuses were engaged in educational experiences that have been shown to promote student success (National Survey of Student Engagement, 2000). The survey's developers emphasized that while *students* are engaged in the effective educational practices, student engagement is a function of the extent to which *institutions* provide students with opportunities to be engaged in educationally purposeful activities. To facilitate institutional discussions of good practices in education, NSSE introduced five benchmarks: (1) Level of Academic Challenge, (2) Active and Collaborative Learning, (3) Student-Faculty Interaction, (4) Enriching Educational Experiences, and (5) Supportive Campus Environment (National Survey of Student Engagement, 2000). NSSE results are intended for institution- and group-level decision making, although some national research has been undertaken using student-level data (Ewell, McClenney, & McCormick, 2011; McCormick & McClenney, in press). Unlike CIRP's *YFCY* survey, the NSSE benchmarks are not intended to represent underlying constructs; instead, the benchmarks represent clusters of questions about related sets of good educational practices (McCormick & McClenney, in press).

A hypothetical institution trying to decide whether to use either the *YFCY* or NSSE's *The College Student Report* will be faced with three sets of questions about the two surveys. First, how well does the content covered by the two surveys represent the programs, goals, and expected outcomes of the institution? Both surveys provide several useful summary measures (constructs or benchmarks) that can assist in interpreting results; but both developers stress that there are many other items on the survey that can be useful in assessing educational programs. Institutions should not focus solely on summary measures, but should examine all of the items on the survey.

Evaluating the structural component of validity will be very different for the two instruments. The *YFCY* constructs are based on item response theory, not classical test theory, and traditional measures of the structure of a measure (e.g., alpha reliability/internal consistency) can yield inappropriate and misleading results (Higher Education Research Institute, 2011). That is not to say that issues of structural validity do not apply to the *YFCY* survey. Item response theory is based on two assumptions that are critical to formulating appropriate construct measures: (1) unidimensionality (i.e., only one construct is measured by the items in a scale) and (2) local independence (i.e., controlling for the influences of the underlying construct, a student's responses to survey items should be independent of one another) (Hambleton, Swaminathan, & Rogers, 1991). Assessing unidimensionality and local independence are likely to be beyond the capability of assessment practitioners at an institution. However, practitioners should expect the survey developers to provide evidence that these assumptions are met. This information is provided for the *YFCY* survey (Higher Education Research Institute, 2011).

Selecting the appropriate unit of analysis is critically important when evaluating the external component of validity.

The NSSE survey is based on classical test theory and it is tempting to think that traditional alpha reliability would be appropriate for the NSSE benchmarks. However, the NSSE benchmarks were developed for institutional- and group-level decision making, not decisions about individual students (Ewell, McClenney, & McCormick, 2011). As a consequence, group-mean generalizability, not alpha reliability, is the appropriate measure of the internal consistency of the NSSE benchmarks (Pike, 2006b). Furthermore, the NSSE benchmarks are clusters of items, not scales that represent samples from a larger universe of observations, and variability across items should not be considered a source of error when calculating the dependability of the benchmarks (Pike, 2006b). Again, assessing group-mean generalizability is beyond the ability of a single institution, but research has shown that the NSSE benchmarks provide dependable measures when based on 50 or more students.

Selecting the appropriate unit of analysis is critically important when evaluating the external component of validity. Both the *YFCY* and NSSE surveys are primarily designed to provide data for decision making at the institutional and group levels. Studies of the validity of these surveys should focus on institutions or groups—not individual students. Findings that the *YFCY* constructs or the NSSE benchmarks do not predict students' grades or persistence is not useful because the surveys were not intended to be used for those purposes. Because the appropriate studies of external validity involve institutions or groups of students, it may be very difficult for institutions to carry out their own validity studies. However, institutions and assessment professionals should demand that survey developers provide evidence of the adequacy and appropriateness of the results provided by their instruments. Both the Higher Education Research Institute (2011) and the National Survey of Student Engagement (Kuh, 2001; Pike, 2006a) have provided evidence of the appropriateness of their surveys for group-level decisions.

In their response to articles appearing in the *Review of Higher Education*, McCormick and McClenney (in press) identified a fourth standard for evaluating assessment instruments. That standard is the *consequential validity* of a measure. Consequential validity refers to the outcomes associated with using an assessment instrument. In other words, are institutions that utilize an assessment instrument able to point to ways in which they have used the information to make improvements in students' educational experiences and learning outcomes? Both NSSE and the CIRP surveys have been used by many different types of institutions for a variety of purposes, and information about institutional experiences with the surveys is widely available. In selecting assessment instruments, institutions and assessment professionals would be wise to consider carefully whether the intended uses of an instrument have yielded positive results at other institutions. ■

References

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.

Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621–694). Washington, DC: American Council on Education.

Ewell, P. T., McClenney, K., & McCormick, A. C. (2011, September 20). Measuring engagement. *Inside Higher Education.* Retrieved from http://www.insidehighered.com/views/2011/09/20/essay_defending_the_value_of_surveys_of_student_engagement

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991) *Fundamentals of item response theory.* Newbery Park, CA: Sage.

Herzog, S., & Bowman, N. A. (Eds.) (2011). Validity and limitations of college student self-report data (*New Directions for Institutional Research Series, no. 150*). San Francisco, CA: Jossey-Bass.

Higher Education Research Institute. (2011). *HERI-CIRP Portfolio.* Los Angeles, CA: Higher Education Research Institute at the University of California Los Angeles. Retrieved from http://www.heri.ucla.edu/PDFs/HERI-CIRP_Portfolio.pdf

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.

Kuh, G. D. (2001). *The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties.* Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved from http://nsse.iub.edu/pdf/psychometric_framework_2002.pdf

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635–694.

McCormick, A. C., & McClenney, K. (in press). Will these trees *ever* bear fruit? A response to the special issue on student engagement. *Review of Higher Education.*

Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational measurement* (3rd ed. pp. 13–104). New York, NY: American Council on Education and Macmillan.

National Survey of Student Engagement. (2000). *The NSSE 2000 report: National benchmarks of effective educational practice.* Bloomington, IN: Indiana University Center for Postsecondary Research.

Pike, G. R. (1989). Criteria for evaluating assessment instruments. *Assessment Update: Progress, Trends and Practices in Higher Education, 1*(4), 8–9.

Pike, G. R. (2006a). The convergent and discriminant validity of NSSE scalelet scores. *Journal of College Student Development, 47,* 550–563.

Pike, G. R. (2006b). "The dependability of NSSE scalelets for college- and department-level assessment." *Research in Higher Education, 47,* 177–195.

Special issue on student engagement. (2011). *Review of Higher Education, 35*(1).

Gary R. Pike is executive director of Information Management and Institutional Research at Indiana University-Purdue University Indianapolis, 1100 Waterway Blvd., WW53, Indianapolis, IN 46202. His e-mail address is pikeg@iupui.edu.

Getting SMART with Assessment: ACTION Steps to Institutional Effectiveness

(continued from page 6)

O=Observe the actual results of the assessment activities.

N=Notify the planning committee of the results.

Individuals engaged in the six-step process described here receive ample support in the system that we propose. Like SMART, ACTION includes a form with a time line, a template, questions, and other cues to guide individuals through each step of the process so they know which information to provide when and to whom. In our collective experience, we have learned that a well-designed form, pretested by volunteers throughout the institution before full implementation, can increase compliance, reduce confusion, and enhance satisfaction with the process. In addition, a planning committee that provides written feedback—suggested revisions, collegial prodding, and even words of encouragement—at both the planning (A, C, T, and I) and the reporting (O and N) stages can enhance not only effective planning but also institutional effectiveness itself. Just as students benefit from frequent, prompt feedback, individuals engaged in planning processes benefit from the same feedback. In fact, without it, planning may fail. Finally, websites, faculty and staff development workshops, and other resources can provide users with the tools they need to succeed in moving the institution forward in planning.

So far, we have advocated separating SMART assessment from ACTION planning in an effort to allow units to focus deliberately on one part of institutional effectiveness at a time without confusing the assessment of student learning with other parts of institutional assessment. However, ultimately this separation is artificial, as true institutional effectiveness stems from a careful, holistic examination of all processes in which a unit is engaged. Thus, we encourage a moment in the annual planning and reporting cycle in which SMART and ACTION recombine in a program review. While some institutions conduct program reviews every five years or so, an annual program review provides an opportunity for each unit to reflect on its programs and services more frequently and to adjust as necessary. Depending on the unit's mission, a program review could include, in addition to SMART assessment and ACTION planning, other standard indicators of success and efficiency, such as enrollment in majors, course scheduling statistics, and faculty scholarship (to use the example of an academic unit). In any case, institutional planners, such as the President's Cabinet, should examine program reviews and allocate financial, human, and other resources.

Colleges and universities employ planning, improvement, and assessment as tools to help advance the institutional mission. The SMART assessment and ACTION planning processes can be initiated across the institution—in academic departments, for co-curricular activities, and in student services. By focusing upon what matters most and engaging people across the institution, this process moves planning off the shelf and provides a vehicle for change that increases effectiveness and the value of education provided by the institution. ■

Eric Daffron is vice provost for curriculum and assessment at Ramapo College of New Jersey and Sandra J. Jordan is provost and vice president for Academic Affairs at Georgia College in Milledgeville, Georgia.

Department Chair Perspective on Learning Outcomes Assessment at Kansas's Public Universities

(continued from page 9)

confirmed the use of an e-portfolio, and about 30% use a course management system to collect and archive student learning assessment results.

Discussion and Recommendations

Several important conclusions regarding faculty interest emerge from the survey data.

1. Nearly all department chairs have implemented an assessment plan.
2. Faculty members are motivated to carry out assessment activities, and improving student learning is a high priority for them.
3. Departments base assessment practice on program-level student learning outcomes, supporting the critical need for establishment and articulation of such outcomes.
4. Faculty view assessment as an important part of the culture of their department, and few report being uncomfortable discussing assessment results with colleagues.
5. A majority of department chairs acknowledge that more assessment

work will need to occur in the near future.

Institutional commitment to assessment was not viewed so favorably as department faculty interest in assessment. These findings could be summarized as follows:

1. Department chairs believe that assessment is an important part of the culture of the institution.
2. Many chairs report feeling pressure from their institution to improve student learning, but insufficient institutional resources and training for assessment are provided.
3. About half of department chairs report being comfortable sharing assessment results with students.
4. Department chairs do not want greater statewide effort related to assessment, but most feel there should be a regular review of assessment to improve student learning.

Variation in the use of assessment models emerged from the Kansas survey data:

1. Departments use traditional assessment methods (pretest/posttest, exit exam, capstone performance), indirect assessment methods (satisfaction surveys), and external examinations to validate student learning.
2. Most departments support the use of rubrics, but adoption of e-portfolio and course management systems is not yet widespread.
3. Programs that are delivered both on-campus and at a distance make use of traditional and rubric-based assessment. ■

C. B. Crawford is assistant provost and Larry Gould is provost/chief academic officer at Fort Hays State University in Hays, Kansas.

Expectations for Assessment Reports: A Descriptive Analysis

(continued from page 2)

The next step was to identify common criteria among these rubrics. Based upon the assessment model at JMU, and after reviewing about a dozen MARs, two of the authors (Fulcher and Orem) developed a checklist for evaluating the MARs. The checklist contained over two dozen criteria—some broad, such as "objectives," and others much more specific, such as whether quantitative or qualitative methods were used.

Using the checklist, raters examined each MAR to determine the presence or absence of particular criteria. Criteria that were prominent in the headers or explicitly stated as a behavioral anchor in the MAR were rated as present. If a criterion was not explicitly stated, we rated it as absent. For example, if a rubric emphasized communication with stakeholders but did not mention faculty, the "communication with stakeholders" was noted as present but not "communication with faculty in the program."

To determine consistency of using the checklist, a subset of 10 MARs was evaluated by the same three raters. The Cohen's Kappa coefficients for this analysis indicated a moderate level of consistency among them.

Table 1 reveals the findings. (For clarity of presentation we organized the results hierarchically.) The vast majority of MARs referenced objective/student learning outcomes (94%), methods (90%), results (73%), and use of results (75%). These four major criteria are listed in the left-hand column of the table. The other more specific criteria are categorized by the major criteria to which they best correspond and by prevalence: common (>50%), moderately common (25%–49%), and less common (10–24%). It should be noted that three additional criteria were particularly rare, appearing in less than 10% of the MARs. They are listed in the footnotes of the table.

Discussion

Most meta-assessment rubrics explicitly referenced objectives/student learning outcomes, methods, results, and use of results. The level of detail within those major criteria, however, differed dramatically. For example, some MARs only required raters to determine if results were present or not. Other MARs probed this area more deeply, prompting raters to examine the clarity of the results, the degree to which the results corresponded to objectives, and the analyses by which the results were derived.

Some criteria were quite common across MARs (e.g., specificity of objec-

Table 1. Prevalence of Criteria in 51 Program-level Meta-Assessment Rubrics

Major Criterion	Common	Moderately Common	Less Common
Objectives/Student Learning Outcomes	• Specific	• Student-oriented • Links to guiding document (e.g., Mission Statement) • Objective/Curriculum relationship (e.g., curriculum map)	
Methods	• Multiple measures • Type of measures (direct vs. indirect)	• Instrument/Objective match • Specification of criteria for success/a priori desired results • Data collection process	• Quantitative methods • Qualitative methods
Results		• Clear • Related to objectives • Related back to criteria for success/ a priori results	• Appropriate analysis • Reasonable interpretation
Use of Results	• Used for program improvement	• Communication with stakeholders	• Dissemination of results to faculty within program • Dissemination of results to people external to the program faculty • Used to petition for resources/Tied to budget planning • Connected to outcomes/objectives • Used to improve assessment process
Rare (<10%): • Participation of students in assessment process • Good psychometrics (reliability) • Good psychometrics (additional validity information)			

tives, direct measures, multiple measures, and use of results for program improvement). This finding seems consistent with topics that are typically emphasized in introductory assessment workshops. We were also pleased that many criteria associated with methods were moderately common, such as instrument-to-objective match, criteria for success, and data collection processes. From our anecdotal experience, these topics are not so commonly discussed in assessment forums. Not all important methodological considerations were common, however. Despite their emphasis in the *Standards for Education and Psychological Testing* (APA, 1999), validity and reliability were rarely mentioned.

From a macroperspective, we were surprised that over 50 institutions—and likely many more that we did not consider—used rubrics to evaluate the assessment processes of their academic degree programs. (You can find most of these MARs at http://www2.acs.ncsu.edu/UPA/assmt/resource.htm; search using "evaluating assessment.") Moreover, many of these institutions provided supplemental documentation to help programs conduct and report good assessment. For example, some institutions provided real or hypothetical exemplary assessment reports. Others provided guides for conducting assessment. It is our thinking that the combination of such resources with the MARs is ideal for the betterment of program assessment.

Before offering recommendations, we suggest the following caveats. First, different institutions may have different aspirations for assessment. Thus, while one MAR may be appropriate for a particular school, it may not be appropriate for another. Second, this study was not designed to rank the MARs from best to worst. The study was intended only to ascertain the prevalence of particular assessment content within the MARs. Third, our institution has spent considerable time developing a MAR. As such, we are biased toward it. Fourth, we were most attracted to MARs that had excellent supporting documentation. Consequently, we recommend three sites or files that include a MAR and additional supporting documentation:

- James Madison University: http://www.jmu.edu/assessment/JMUAssess/APT_Help_Package2011.pdf
- Oakland University: http://www.oakland.edu/?id=12886&sid=250
- Washington State University: http://communitylearning.files.wordpress.com/2010/12/guide-to-effective

JOSSEY-BASS™
An Imprint of
WILEY

One Montgomery St., Suite 1200
San Francisco, CA 94104-4594

www.josseybass.com

Periodicals
Postage Paid at
San Francisco, CA,
and additional
mailing offices

978-1-118-36336-2

-program-learning-outcomes-assessment-rubric-august-2010.pdf

In summary, a growing number of institutions engage in meta-assessment. This promising practice enables programs to receive feedback about the quality of their assessments. Additionally, this feedback can be aggregated across programs indicating the overall quality of assessment at an institution—useful information for upper administration or an accrediting body. One should note that meta-assessment is often a time-consuming endeavor. For example, at our institution teams of two raters evaluate over 100 assessment reports. Considering the time it takes to read the assessment reports, rate them numerically, provide comments, adjudicate rater discrepancies, and disseminate the results to programs and other stakeholders, well over six hours are spent on each report (or over 600 hours total at JMU). Given such resource requirements, it is important that institutions begin documenting that this meta-assessment process contributes to improved assessment and, ultimately, to improved student learning. ■

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bloom, M. F. (2010). Peer review of program assessment efforts: One strategy, multiple gains. *Assessment Update 22*(5), 5–7, 16.

Fulcher, K. H. (2010, October 25) Subject: Rubrics to evaluate quality of assessment. In ASSESS. ASSESS@LSV.UKY.EDU. Retrieved from http://lsv.uky.edu/scripts/wa.exe?A2=ind1010&L=assess&T=0&F=&S=&P=8992

Fulcher, K. H., & Orem, C. D. (2010). Evolving from quantity to quality: A new yardstick for assessment. *Research and Practice in Assessment, 4*(1), 1–10. Retrieved from http://www.virginiaassessment.org/rpa/5/FulcherandOrem.pdf

Ory J. C. (1992). Meta-assessment: Evaluating assessment activities. *Research in Higher Education, 33*(4), 467–481.

Keston H. Fulcher is associate director at the Center for Assessment and Research Studies, Matthew Swain is a graduate student, and Chris D. Orem is a doctoral candidate at James Madison University in Harrisonburg, Virginia.