

Assessment UPdate

November–December 2014
Volume 26, Number 6



Progress, Trends, and Practices in Higher Education

A Cross-Disciplinary Approach to Evaluating the Senior Thesis

Carol Trosset and Steven Weisler

CAPSTONE THESES—THE CULMINATING PROJECT IN CERTAIN CURRICULA—ARE an important place to perform summative assessment, as these artifacts should provide evidence of the highest skill level that students achieve. The goal of this project is to develop a rubric that might be used to evaluate advanced student work across all analytical, research-based, disciplines. Our aim in developing such an approach is to address a certain type of student work, rather than a single skill or discipline. Senior theses in the analytic mode typically comprise writing, research, and critical thinking components. Because these are common features of capstone work, we argue, it should be possible to develop a cross-disciplinary rubric by which to evaluate them.

We began developing the rubric (see Table 1, page 14) in spring 2006. The first version consisted of eleven components: rationale; dealing with complexity in framing the topic; approach; scholarly context; position; argument; use of data/evidence; insight—seeing patterns and connections; usage, grammar, and spelling; organization; and clarity, style, and readability. Eleven faculty members representing four institutions and eight disciplines used the draft rubric to evaluate eighty-one theses. Initially, all readers evaluated a single thesis and each independently assigned rubric scores. As a norming exercise, the group then discussed the scores they had assigned for each component, which eventually led to their agreeing on how to describe the quality of that particular thesis. Over the next two days, evaluators scored the remaining theses and discussed how to improve the rubric.

Creating a rubric that was both interdisciplinary and discriminated adequately between varying levels of quality entailed (1) providing sufficient detail in the text of each cell to assist readers in placing marginal works at appropriate levels; (2) finding evaluative language that seemed appropriate to the widest variety of disciplines (such as referring to “evidence” as well as “data”); and (3) raising the standard of Level 2 so that a wider range of weak student work would fit Level 1, and that Level 3 would represent quite high-quality work, reserving Level 4 for truly exceptional work that

CONTENTS

ARTICLES

A Cross-Disciplinary Approach to Evaluating the Senior Thesis **1**
Carol Trosset and Steven Weisler

Involving Undergraduates in Assessment: Assisting Peers to Provide Constructive Feedback **3**
Adriana Signorini

Strengthening Foundations for Assessment Initiatives through Professional Development **5**
Frederick Burrack and Christopher Urban

COLUMNS

Assessment Measures **7**

From the States **9**

Recommended Reading **11**

Assessment Update

Progress, Trends, and Practices in Higher Education

November–December 2014

Volume 26, Number 6

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, *Queens University of Charlotte*

T. Dary Erwin, *James Madison University*

Cecilia L. Lopez, *Harold Washington College*

Marcia Mentkowski, *Alverno College*

Jeffrey A. Seybert, *Higher Education Consultant*

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

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. 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 (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 © 2014 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

is probably seen only once in several years. For example, under the “argument” category, Level 1 was originally defined as “No argument, perhaps a simple assertion,” and was revised to read “Weak, invalid, or no argument; perhaps a simple assertion.” The group also developed the policy that in any row, a thesis must fully satisfy all the lower levels to be scored at a higher level. Also, no fractional scores should be given. If a thesis does not fully meet the qualities listed in a particular cell, it should receive the next lower score.

One advantage of using a rubric is that it ensures that all evaluators will consider the same components of student work.



In order to explore the validity of the instrument, we later asked each of three groups of about twenty evaluators to read one of three theses that had been included in the original project. Half of each group evaluated the thesis with the use of the rubric; the other half assigned a grade and listed the criteria they had used. The goal was to attempt to validate the evaluative rubric categories by observing whether they were also considered by faculty readers who were not using the rubric. Although evaluators assigning grades used a variety of criteria, all of these were already included in the rubric. This result is consistent with previous research in which we found that narrative evaluations of capstone theses used the rubric components, but that no individual evaluation discussed more than four of the eleven criteria. Two of the rubric components (insight and dealing with complexity) were virtually unused by the graders. One advantage of using a rubric is that it ensures that all evaluators will consider the same components of student work.

(continued on page 15)

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. ■

Involving Undergraduates in Assessment: Assisting Peers to Provide Constructive Feedback

Adriana Signorini

SOON AFTER IT WAS LAUNCHED IN 2005, the University of California, Merced (UCM) was designated as a Hispanic Serving Institution. The campus currently enrolls almost 6,200 students, with projected growth to 10,000 by 2020. For a research university, UCM has relatively high percentages of Pell grant recipients (60 percent) and first-generation college goers (62 percent). Overall, our undergraduates, who represent 94 percent of total enrollment at UCM, are particularly at risk for not finishing a college degree.

The Center for Research on Teaching Excellence at UCM sponsors the Students Assessing Teaching and Learning (SATAL) program, which trains undergraduates in research design, data gathering, and effective reporting to support faculty with their assessment projects. In reciprocal communication with students, SATAL students, as well as instructors, need to make sure that feedback is constructive. For example, a student's comment, "This class is too early," is not helpful for an instructor who is looking for ways to improve learning in a class. One of the research projects carried out by the SATAL program last spring was the feedback initiative (FI), targeting the goal of assisting students to provide constructive feedback.

Feedback Initiative

Background. Students are often asked to reflect on their learning to provide feed-

back to peers as well as instructors. For instance, in a "flipped" classroom, an instructor frees class time for students to engage in more collaborative learning assignments, such as peer review. Midsemester or final course evaluations also have students reflect on their progress as learners as well as the instructor's teaching effectiveness. According to Bloom's taxonomy of learning, the ability to evaluate is a skill at the higher

students, and the Panadero and Jonsson (2013) rubrics to mediate improved performance and self-regulation. In the FI a rubric offers students criteria for identifying levels of performance by analyzing the components of the feedback process and by asking who, what, when, where, why, and how to enhance the effectiveness of the feedback. Five instructors of a freshman writing course participated in FI; they taught 221 students in twelve

For feedback to be a powerful learning tool, instructors should train students to use effective feedback practices, and a feedback rubric could provide the framework to promote this learning effectively.



end of the taxonomy, which most undergraduates are still developing. However, since very early in their college lives, students evaluate instruction, typically with little or no formal training. For feedback to be a powerful learning tool, instructors should train students to use effective feedback practices, and a feedback rubric could provide the framework to promote this learning effectively.

The SATAL Program developed the Feedback Initiative (FI) to synthesize research on giving and receiving feedback. Ultimately three lines of research guided the design of this empirical study on feedback: Brinko's (1993) feedback-giving practice to improve teaching, the Hattie and Temperley (2007) analysis on what information should be fed back to

sections. To assess the impact of FI on students' feedback, the SATAL program collected direct and indirect evidence.

Training. In a fifty-minute in-class presentation, students were trained to provide constructive feedback. A team of two SATAL students led each classroom presentation, which included PowerPoint slides that presented (1) a working definition of feedback, (2) situations in which feedback is required, (3) a feedback rubric with criteria for providing constructive comments, and (4) a follow-up group activity. SATAL students guided their groups by scaffolding application of the rubric and guiding the wording of constructive feedback. The groups reconvened and presented to the

entire class their revised feedback. By using a slideshow presentation and lesson plan, the SATAL program ensured that the project was presented in similar ways in each of the participating twelve sections in this first-year writing course.

Assessment. To assess FI, the SATAL students collected and analyzed various forms of direct and indirect evidence. During the in-class FI presentation, students completed a pre- and posttest with feedback samples recorded before and after the activity to document the presentation's impact on students' comments and students' level of engagement with feedback training. After the presentation, students completed a minute paper in which they summarized what they learned and noted any questions that remained unanswered on the topic. Also, instructors assisted with the FI data gathering by providing a verbal summary of their impressions of students' engagement during the FI presentation session and their analysis of final course-evaluation results for the statement, "This course has taught me to give and attend to feedback." The SATAL program conducted focus-group sessions with freshmen from the participating writing sections, and finally, SATAL students shared their own perceptions of FI.

Results

Students' Samples. Student feedback samples from before and after the presentation and a peer-review session were collected to include as direct evidence of student learning. The *before* FI presentation feedback samples briefly address weaknesses in grammar, style, and content. *After* the FI presentation, student samples demonstrate their awareness of audience and purpose and closer attention to the rubric criteria. In decreasing order of frequency, the rubric criteria most used by students when providing feedback to their peers were the following:

1. Offer specific suggestions that model appropriate behavior (215, or 90 percent)
2. Focus on content rather than on the person (146, or 61 percent)
3. Provide a balance of positive and negative feedback (138, or 58 percent)
4. Include accurate and specific data that are clear about irrefutable evidence (119, or 50 percent)
5. Keep comments nonjudgmental and descriptive rather than evaluative (109, or 46 percent).

Minute Paper. After the FI presentation, students completed a minute paper activity. Among the most-repeated responses, students mentioned that they learned "How to provide constructive feedback" (78, or 33 percent), "Providing positive feedback" and "How to phrase negative feedback" when providing constructive feedback. With regard to the questions students still had on FI after the presentation, 182 students (82 percent) did not have any further questions. A majority of the students offered no suggestion on how to improve the FI presentation (118, or 56 percent).

Instructors' Observations. Reflecting on what went well, instructors reported the following findings:

1. The presentation was interactive, informative, and helpful for planning upcoming peer-review activities (5, or 100 percent).
2. The class was engaged and participated well during the presentations.
3. Writing samples on the whiteboard and making changes to these was very helpful for students "to model how the commentary on the board could be revised."
4. Students found helpful the idea of sandwiching positive and negative feedback (2, or 40 percent).
5. FI would influence the way instructors would provide feedback to students in the future, because now students were providing them with more specific information as a result

of giving and receiving more constructive peer feedback. Instructors could be more effective at helping students, and thus students could become better writers: "The presentation will also shape my commentary to students." This result was also evident during one-on-one conferencing (2, or 40 percent).

Focus Group Summary Report. On a 5-point scale with 5 being the best, most of the students rated their feedback skills as 4 (35, or 67 percent) or 5 (11, or 22 percent) after the presentation. Most of the students agreed or strongly agreed that their ability to assess and provide constructive feedback could be improved through training, and that a rubric could effectively guide them (38, or 74 percent). Although some of the students concluded that having a feedback rubric was helpful, they noted that it was not just the rubric that assisted them, but also the assistance they received from presenters (33, or 65 percent). Students stated that they would utilize the rubric in the future (41, or 80 percent). Also, by providing better feedback to their peers, they believed they became better writers.

Final Course Evaluations. Instructors collected students' comments about FI in the midsemester and final course evaluations. FI was highly valued when students responded to the statement: "This course has taught me to give and attend to feedback." Most of the students (85 percent) rated this statement "frequently" or "always," and some attributed their improved ability to offer useful feedback to the FI session directly. Students also mentioned FI usefulness in other parts of the course evaluations, as in the statement: "Identify and evaluate aspects of this course that have been especially helpful to you."

SATAL Students' Reflections. All five SATAL students who participated affirmed the need to train peers to provide
(continued on page 13)

Strengthening Foundations for Assessment Initiatives through Professional Development

Frederick Burrack and Christopher Urban

The Need for Professional Development

UNIVERSITY ASSESSMENT administrators face many challenges when working to enhance student learning assessment across their campuses. Issues often confronted include (a) faculty and programs that focus on content dissemination as a primary goal of instruction, (b) faculty that are familiar with only a few traditional assessment measures, (c) difficulty sharing effective

but lack of assessment strategies. Many involved in teaching are familiar with traditional content/knowledge assessments, but have not learned more effective and efficient assessment techniques. Some forms of alternative and authentic assessments remain unused or unknown to many. Training on the utilization of a variety of student learning assessment measures is paramount.

Assessment coordinators observe many effective assessment processes that

tunity to communicate the purpose and value of student learning assessment, thereby moving programs toward meaningful practice.

Kansas State's Approach to Assessment

Kansas State University began implementation of a campus-wide student learning assessment process in 2004 as a result of recommendations from its accreditor, the Higher Learning Commission. Currently, each academic program documents the student learning outcomes (SLOs) essential for diploma credentials; identifies appropriate and specific assessments for each outcome; administers assessments in outcome-level courses; annually collects student achievement data to be discussed among faculty; and reports assessment results, interpretations, and programmatic decisions. Our goal is to progress toward a greater number of programs with the assessment process used to spur discussion and decisions on program improvement.

Each program has an assessment coordinator who collects the assessment data designated in the assessment plan and creates a report for the program's faculty. Faculty discussions are added as the narrative for the report, which is submitted to the College Assessment Review Committee (CARC). Each CARC includes a representative from each program. Annual Progress Reports (APR) are peer-reviewed by each CARC. Simultaneously, the Office of Assessment reviews

Many involved in teaching are familiar with traditional content/knowledge assessments, but have not learned more effective and efficient assessment techniques.



assessment practice because of departmental or institutional silos, (d) programs addressing assessment issues with little foundation from best practice research, and (e) resistance from those who see student learning assessment as one more thing to do and unrelated to teaching and research.

One way to employ professional development to address these issues is to alter the fundamental instructional paradigm that challenges assessment goals. Broadly, this involves shifting from content coverage to a focus on identifying how students make sense of and apply what has been taught (Miller 2012). Other times the issue that hinders implementing assessment initiatives is not a difference in educational understanding,

are nearly invisible to others on campus who might benefit from these examples. A challenge is finding ways to break down disciplinary boundaries so all educational units can benefit. Through sharing good practice in professional development offerings, many challenges can be eliminated. There are also many programs that implement student learning assessments for compliance but do not use the process to identify student learning needs or to implement improvements to their programs. This is a problem because closing the assessment loop—making programmatic alterations that result in enhanced student learning—is the ultimate purpose of student learning assessment. The response to minimal compliance serves as an oppor-

the electronically submitted reports and provides feedback. Each program has an opportunity to reply and make revisions to their report. The feedback and the programs' responses are sent from the Office of Assessment to each dean with a summary, and a university summary is submitted to the provost.

Professional Development as a Foundation

The assessment structure of each university will determine the goals for professional development. In our situation, assessment that is student-centered as well as authentic to the discipline serves students, faculty, and staff best. Inculcating this paradigm across the institution will ensure that student learning assessment is of value to our constituents and leads to continual, meaningful improvement. Professional development in assessment contributes to the gradual shift toward this paradigm and helps programs make their assessment processes more efficient and meaningful.

We have identified several professional development outcomes for participants that, taken together, will help our assessment process succeed to its fullest. Participants will possess a student-centered philosophy and knowledge of techniques to improve assessment processes. Participants will demonstrate skills in specific topics of need, such as learning to write SLOs, developing and using rubrics, and assessment planning and mapping. Participants will engage in collaborations with both internal and external partners to implement ideas beyond one department or institution. Knowledge is key to moving faculty from resistance through compliance all the way to active involvement in assessment processes (Miller 2012). Tying professional development to specific outcomes provides the foundation for strengthening assessment practice university-wide.

Professional Development Examples: Support To Attend Conferences. Conference attendance can provide an indi-

vidual or team with training in assessment, opportunities for collaboration, and knowledge that can help develop an assessment paradigm and overcome resistance. It is not uncommon for resistant individuals to return with enthusiasm to strengthen their program's processes.

For many years, our university's Office of Assessment has supported faculty to attend conferences with the mission of bringing back updated information and strategies to their areas. In 2013, these opportunities were formalized and more effectively marketed through the creation of a mini-grant program. Preference is given to those presenting a session on assessment or completing other related professional development. Each

Tying professional development to specific outcomes provides the foundation for strengthening assessment practice university-wide.



recipient is required to share what was learned with their program and report to the Office of Assessment how they integrated the knowledge to improve assessment in their program.

Support to attend assessment workshops focused on developing an assessment project has led to several campus-wide improvements over the last few years. In 2012, a team was sent to a week-long institute to develop a project focused on introducing High-Impact Practices across our campus. What developed was a set of sessions presented at our Institute for Student Learning Assessment. In 2013, a second team was sent to create an introductory professional development workshop series, The Academy for Student-Centered Learning.

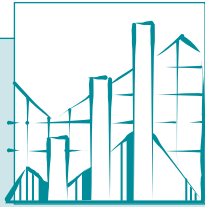
Professional Development Examples: The Academy for Student-Centered Learning. To formalize what in the past had been scattered attempts at professional development, a set of workshops and trainings was organized within a formal structure titled the "Academy for

Student-Centered Learning." Partnering with curricular and co-curricular units to plan the Academy provided guidance on content, speakers, scheduling, and marketing. Participants gain membership in the Academy by completing four introductory workshops: Student-Centered Learning and the Whole University; Student-Centered Learning Outcomes; Measuring Student Learning; and Continual Improvement of Student Learning. Incentives for attending Academy workshops are provided to drive attendance as well as promote additional assessment initiatives. Participants are given a book, free lunch at the annual Institute for Student Learning Assessment, and preference for mini-grant funds. The

audience for the workshop includes new faculty, graduate students, and faculty interested in learning about assessment in a student-centered paradigm.

Professional Development Examples: Targeted Workshops. One-time targeted workshops are useful to meet the needs of a large population of faculty and staff with a diverse range of assessment knowledge, as well as to promote specific assessment initiatives. As part of a larger pilot project to promote use of the AAC&U VALUE Rubrics, a day of workshops was planned. An AAC&U representative presented workshops for those participating in the pilot, for those with a general interest in the VALUE Rubrics, and for a specific department that expressed interest.

In 2013, a college brought in a speaker to prepare for the assessment part of an accreditation visit. The two-day training allowed the entire college to meet, learn, and discuss assessment issues—something previously limited to assessment committees. Because the
(continued on page 12)



Assessment Measures

Course Embedded Assessment Designs: Lessons from the MSC

Gary R. Pike

During the past year, I have been fortunate to participate in the Multi-State Collaborative to Advance Learning Outcomes Assessment (MSC). The MSC is a project involving the State Higher Education Executive Officers Association (SHEEO), the Association of American Colleges and Universities (AAC&U), and nine states (Connecticut, Indiana, Kentucky, Massachusetts, Minnesota, Missouri, Oregon, Rhode Island, and Utah). The goal of this pilot project is to collect authentic evidence of student learning with the use of the AAC&U VALUE Rubrics in order to assess student learning and development at the state and institution levels (State Higher Education Executive Officers Association [SHEEO] 2014b). For the pilot study, sixty-seven two-year and four-year colleges and universities will be collecting student artifacts. Assessments will focus on written communication and quantitative literacy, with critical thinking also being assessed in some states. Ultimately, the data will be used to gauge the validity and reliability of the assessment model (SHEEO 2014b).

My role in the MSC has been as a member of the MSC sampling subgroup, assisting institutions to develop sampling plans for collecting student artifacts. I believe the sampling framework developed by the MSC may be useful to other institutions interested in using authentic, course-embedded assessment procedures to evaluate student learning. In addition, having spent much of the summer reviewing and commenting on sampling plans, I noticed some similarities in institu-

tions' approaches that may be useful for other colleges and universities as they sample student products.

Because the goal of the MSC is to provide representative samples of student work across multiple institutions and multiple states, the sampling parameters for the project were tightly prescribed. While not essential for a single institution interested in assessing student learning, the specificity of the sampling plan may be useful to institutions in thinking about whom to assess using course-embedded assessments. The ultimate goal of the MSC sampling process is to obtain seventy-five to one hundred independent artifacts per outcome per institution. (Some of the smaller institutions participating in the MSC project have formed consortia of three institutions, allowing them to provide twenty-five to thirty-five independent artifacts per outcome per institution.) Institutions wishing to collect larger samples of student artifacts have been encouraged to do so, although the institutions themselves will be responsible for scoring those artifacts (MSC Sampling Subgroup 2014).

In order to ensure that the samples of student artifacts will accurately represent graduates from each institution, the eligible student population to be sampled from is clearly defined. The eligible student population for the study is those students who have completed at least 75 percent of the credits needed for graduation. For baccalaureate institutions requiring 120 credits for graduation, all students should have completed at least 90 credits. For institutions requiring 60 credits for an

associate degree, eligible students must have completed 45 credits. For this study, students may have earned the credit hours at their current institution, or they may have transferred the credits from another institution. Students may be either full-time or part-time, may be enrolled in day or evening courses, and may be enrolled in face-to-face, online, or hybrid courses (MSC Sampling Subgroup 2014).

In addition to defining the eligible student population for the study, the MSC sampling subgroup established several parameters for selecting a sample. The purpose of the sampling parameters is to provide as diverse and representative a sample of student artifacts as possible. The protocols require that students be selected from a variety of disciplinary programs or areas. Participating institutions are told that samples drawn from only one or two majors or programs are unacceptable. Similarly, samples should be drawn from multiple courses and multiple faculty members within disciplinary areas. A limit of seven to ten artifacts can be collected in total from any one faculty member or any one course; only one artifact can be collected per student; and only one outcome (i.e., writing, quantitative literacy, or critical thinking) can be assessed with the use of an artifact (MSC Sampling Subgroup 2014).

Recognizing that sampling can be accomplished in a variety of ways, the sampling subgroup identified three different sampling approaches. All three approaches are described in detail in the "MSC Sampling Parameters and

Suggested Sampling Methods” available on the SHEEO website (www.sheeo.org/node/774). The first approach is to begin with students as the primary sampling unit. Institutions will first identify eligible students enrolled in fall 2014 and the courses in which there are enrolled. Next, instructors of those courses will be contacted to determine if it would be appropriate to assess one of the learning outcomes in that course. The instructors also will be asked if they agree to include an appropriate assignment in the course and submit the student artifacts for assessment in the pilot study. From the list of eligible students in courses taught by willing faculty, random samples of one hundred students will be selected for each learning outcome. Institutions are also encouraged to generate backup samples by removing students in the first sample from the original list and then drawing additional random samples for each outcome. Artifacts from students in the backup samples could be used should students in the original samples leave the institution or not complete the assessment.

The second approach for drawing a representative sample is to start with courses in which eligible students are likely to be enrolled. In some instances, the courses may be identified easily as capstone courses. The sampling subgroup recommends that institutions use data from the fall 2013 semester to identify possible courses to be included in the study. Several institutions have used their institutional research offices to help them identify courses in which eligible students are likely to be enrolled. Once courses have been identified, institutions will contact instructors of those courses to solicit their participation in the project. A list of eligible students enrolled in the courses taught by participating faculty members will be generated. Random samples, including backup

samples, of eligible students will then be selected and submitted for scoring. Because of the requirements in the sampling parameters, no more than ten student artifacts can be selected from any one course.

The third approach begins by identifying faculty members who would likely be willing to participate in the study, who teach courses that enroll eligible students, and who teach courses that are conducive to assessing either writing, quantitative reasoning, or critical thinking. Faculty members will be asked if they would be willing to participate in the study by including in their course an assignment representing one of the three learning outcomes and submitting the student work for scoring in the pilot study. A list of eligible students enrolled in participating faculty members’ courses will be generated and random samples of one hundred students, plus backup samples, will be selected. Again, no more than ten students can be selected for any one faculty member.

Members of the sampling subgroup quickly realized that although there are three distinct approaches for selecting samples of student artifacts, all three are dependent on the willingness of faculty members to participate in the project. Moreover, the success of the project depends on the active participation of faculty members because they must design assignments in their classes that are appropriate for eliciting student work that can be assessed using the VALUE Rubrics. In order to encourage faculty participation, SHEEO and the MSC project staff have provided a variety of resources for faculty members. In order to increase faculty members’ understanding of the MSC, a series of frequently asked questions (FAQs) has been posted on the MSC website (SHEEO 2014c). These FAQs provide background information about the Multi-State Collaborative project

and the VALUE Rubrics, the purposes of the study and how data will be used, and estimates of the time commitment required and resources available to faculty members. The FAQs also address potential concerns faculty members may have about the confidentiality of results and the protection of human subjects.

Additional details about the assessment project are provided in the paper “At A Glance: Potential Benefits and Challenges for Faculty Participants in the Multi-State Collaborative Pilot Study” available on the MSC website (SHEEO 2014a). The MSC project staff also have developed a series of webinars on the benefits and goals of the MSC project, the VALUE Rubrics, and how to design class projects so that they yield student artifacts that can be assessed with the use of the VALUE Rubrics. When providing feedback to institutions concerning their sampling plans, members of the sampling subgroup also stressed the importance of involving faculty members in the sampling process.

So what lessons from the MSC assessment project can be learned by assessment practitioners on college campuses? I believe that colleges and universities interested in using student work generated as part of class activities (i.e., authentic assessment) can learn two important lessons from the MSC. First, obtaining a representative sample of student work requires careful attention to sampling protocols. This means that the eligible student population must be carefully defined. For the MSC, defining the eligible student population required setting the number of credits required for eligibility. It also required determining whether full- or part-time students should be included, whether transfer students would be included, and whether evening students or students

(continued on page 10)



From the States

The Feds are Back: Reauthorization and College Ratings

Peter T. Ewell

Things are heating up in Washington as the first ideas for reauthorizing the Higher Education Act (HEA) are put on the table by House and Senate leaders. As usual, no one expects real action on these proposals for another year, or at least until the new Congress is organized. But a look at some of these proposals may provide clues about what major policy players are thinking about higher education accountability. At the same time, the administration is planning a college ratings system, which, if implemented, may be a game-changer.

Given a divided Congress, there are two competing versions of the HEA out there at the moment. The Democratic version is offered by Senator Tom Harkin of Iowa and is a comprehensive bill. The Republican alternative from House leaders and Senator Lamar Alexander of Tennessee consists of several proposed bills that address many of the same issues piecemeal. Although the bulk of both approaches concern financial details associated with Pell grants and student loans, both also contain accountability provisions in two areas—consumer information and accreditation.

With regard to consumer information, the Harkin bill adds several measures to the existing “College Scorecard” maintained by the Department of Education, including more complete graduation rates (not limited to first-time, full-time) and net price by income. Calculating these additional graduation rates could impose a substantial burden on institutions, which would need to do these calculations themselves in the absence of a federal

student unit record system. Such measures also reflect a continuing federal conviction that consumer information will serve accountability by leveraging consumer choice—a belief that dates as far back as graduation rate reporting under Student Right to Know in 1990. In fact, there is overwhelming evidence that student choice is driven by proximity and price, with scant attention to quality. But there are also some more consequential provisions associated with poor institutional performance. For example, institutions with high student loan default rates and low graduation rates would automatically receive program reviews by the Department of Education—a policy feature reminiscent of performance funding in many states.

Though indirect, accreditation is the other vehicle through which the federal government can hold colleges accountable, so the HEA already has many prescriptions about what accrediting organizations should do. Both bills leave in place the requirement for accreditors to put a priority on evidence of student academic achievement, but they establish an additional requirement that all accreditation self-studies and team reports be made public. This appears somewhat redundant, however, because virtually all recognized accreditors are already planning to publish such materials themselves or are requiring institutions to do so.

Because it will probably take eighteen months to two years for Congress to take action on reauthorization, though, these proposals may easily change. Much more immediate, poten-

tially, are the administration’s plans to create a college ratings system based on a number of standard measures of institutional condition or performance. The most consequential proposal here is that the Department will link performance on these measures with the receipt of student aid funds. As might be expected, this initiative is strongly opposed by the Washington lobbying community, led by the American Council on Education (ACE). But it will probably happen despite this and its technical quality may be higher than many critics fear. A major reason why it will probably be implemented, according to Secretary of Education Arne Duncan, is that President Obama has a strong personal interest in putting the rating system in place and has therefore made it a priority for the Department of Education to accomplish. As a result, the Secretary’s goal is to have a draft of the system available this fall, although other reports put the target date as the end of the calendar year. Either way, we are likely to see it very soon.

The proposed ratings system is also likely to be pretty sophisticated. Its architects at the Department have looked extensively at state indicator systems and performance funding schemes to learn what works best, and are also consulting widely with higher education policy and measurement researchers to surface best practices. For example, they are well aware of the limitations of the current six-year graduation rate calculation based on first-time, full-time, students and have been considering a range of alternatives including ratio measures comparing credentials

produced to students enrolled. Similarly, they know that current measures of graduate earnings based on income a year after completion can be misleading because some job trajectories—particularly for liberal arts graduates—are slower to take off with respect to earnings and do not take graduate school attendance into account. Though hampered by the lack of a national student unit record system, they also are searching beyond the Integrated Postsecondary Educational Data System (IPEDS) for data sources, including data collected by the Census Bureau and the Internal Revenue Service. Finally, they appear to be listening carefully to feedback from the field. For example, an early proposal was to create fixed standards that would govern decisions about institutional access to financial aid. After much consultation,

they now appear to be moving toward sustained or improved performance as a benchmark.

One particularly vexing issue in creating measures of this kind is how to handle situations in which performance is driven by factors over which institutions have little or no control. For example, graduation rates are largely a product of the academic preparation and ability of the students enrolled, so less selective institutions will almost always have lower graduation rates than their more selective counterparts. Among other things, this situation creates perverse incentives for open-access institutions to abandon their missions (if they can) to achieve better performance. Architects of the ratings system are well aware of this situation and are exploring a range of alternatives to addressing it, including

limiting comparisons to selected peers and regression-based models that take a range of institutional characteristics into account to construct an “expected” graduation rate against which actual performance can be compared.

In sum, higher education is likely to be in for a rough ride with respect to accountability over the next couple of years. The fate of the college ratings system, which may well be known as this issue goes to press, will probably tell us a lot about what we will face in reauthorization. However it comes out, college and university leaders and analysts will have a lot to think about. ■

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

Assessment Measures

(continued from page 8)

in distance-learning programs would be included in the research. For the MSC pilot study, all of these groups were included. Whether the same student groups should be included in an assessment design for a given college campus depends on the outcomes being assessed and the educational programs being evaluated.

The second important lesson to be learned from the MSC pilot study is that although a variety of methods for selecting a representative sample of student artifacts exist, all of the approaches depend on the willingness of faculty members to participate in the assessment project. To encourage faculty participation, MSC staff have made information about the goals of the project, the measures to be used, and sampling procedures available on the web. They have also included links to webinars and emphasized to participating institutions the impor-

tance of involving faculty members in every step of the assessment process. Institutions must involve faculty members in planning and conducting assessment efforts and in sharing assessment results in faculty meetings across campus. All of these efforts should be designed to emphasize the benefits of assessment and to address faculty concerns about the use of assessment data. ■

References

- MSC Sampling Subgroup. 2014. *Multi-State Collaborative 2014/2015 Pilot Study: Sampling Parameters and Suggested Sampling Methods*. Boulder, CO: State Higher Education Executive Officers Association. www.sheeo.org/node/774
- State Higher Education Executive Officers Association (SHEEO). 2014a. *At a Glance: Potential Benefits and Challenges for Faculty Participants in the Multi-State Collaborative Pilot*

Study. Boulder, CO: SHEEO. www.sheeo.org/node/774

State Higher Education Executive Officers Association (SHEEO). 2014b. *Multi-State Collaborative to Advance Learning Outcomes Assessment (MSC) Description*. Boulder, CO: SHEEO. www.sheeo.org/sites/default/files/MSC_POR_Description_jc_3-19-2014.pdf

State Higher Education Executive Officers Association (SHEEO). 2014c. *Multi-State Collaborative to Advance Learning Outcomes Assessment (MSC): Frequently Asked Questions for Faculty Participants*. Boulder, CO: SHEEO. www.sheeo.org/sites/default/files/project_file/MSC_FAQ_021014.pdf

Gary R. Pike is a professor of higher education and student affairs in the Indiana University School of Education—Indianapolis, 902 West New York Street, ES3116, Indianapolis, IN 46202 (pikeg@iupui.edu).

RECOMMENDED READING



Susan Kahn

Finley, A., and McNair, T. 2013. *Assessing Underserved Students' Engagement in High-Impact Practices*. Washington, DC: Association of American Colleges and Universities.

We hear a lot about so-called “high-impact practices” these days, and, indeed, the evidence that these practices boost students’ chances of academic success is persuasive (Kuh 2008). In *Assessing Underserved Students' Engagement in High-Impact Practices*, authors Ashley Finley and Tia McNair report on a study that examined the impact of these practices (including learning communities, service learning, study abroad, internships, student/faculty research, and senior capstone experiences) on students from traditionally underserved populations, as well as these students’ access to such practices. Written for an audience of campus practitioners, the monograph aims to address the persistent educational disparities between underserved (i.e., underrepresented minority, first-generation, transfer, and low-income) and “traditionally advantaged” college students.

As Association of American Colleges and Universities (AAC&U) President Carol Geary Schneider and Vice President Susan Albertaine note in the Foreword, AAC&U has long advocated “inclusive excellence” in higher education. When it comes to high-impact practices, this means “providing a liberal education that offers not only equitable access. . .but also equitable achievement of outcomes” (p. v). This study advances that agenda by developing a multimethod approach

that individual campuses can replicate and supplement with additional assessment. Part I draws on National Survey of Student Engagement (NSSE) data from thirty-eight state institutions in California, Oregon, and Wisconsin to examine rates of participation in high-impact practices, effects of participation on students’ perceptions of their learning, and effects of participation in multiple high-impact practices on these perceptions. Part II adds data gathered from student focus groups, designed “to explore the educational experiences of underserved students” (p. 19) at nine comprehensive public institutions.

Some of the findings will not be startling to educators who have followed the discussion of high-impact practices: NSSE data show that underserved students report lower participation in these practices, but greater learning gains when they do participate, than traditionally advantaged students. More surprising perhaps is that participating in multiple high-impact practices results in even larger boosts to underserved students’ perceptions of their learning, and begins to close the gap between these students and traditional students in perceived educational achievement. The focus groups yielded helpful insights about barriers to these students’ access to high-impact experiences and gaps in the advising and guidance they receive. Consistent themes of the focus groups included “the value of being in educational environments that encouraged [underserved students] to interact with others, explore differing opinions, apply knowledge in real-world settings,

incorporate their lived experiences into their learning, and participate in support networks” (p. 29).

The monograph closes with recommendations for improving equity in high-impact practices on individual campuses, including a reminder to supplement the indirect methods used in this study with direct assessment of student work. Especially valuable here is the addition, in Appendix A, of an “Assessing Equity in High-Impact Practices Toolkit,” developed by the University of Southern California’s Center for Urban Education. The Toolkit provides a framework—focused on equal representation, access, and impact—and a six-step process that can guide campus practitioners in becoming “more equity-minded in their use of high-impact practices” (p. 36). An emphasis here is on avoiding “deficit-minded explanations or existing assumptions” that focus “on student characteristics as the cause of the inequity” (p. 40).

As educators, we bear responsibility for the success of all of our students. *Assessing Underserved Students' Engagement with High-Impact Practices* offers us valuable and practical guidance for ensuring such success. ■

References

Kuh, G. 2008. *High-Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter*. Washington, DC: Association of American Colleges and Universities.

Susan Kahn is director of institutional effectiveness at Indiana University—Purdue University Indianapolis.

Strengthening Foundations for Assessment Initiatives through Professional Development

(continued from page 6)

discussion was discipline-specific and couched in accreditation, faculty appeared less resistant to change (Ewell, Paulson, and Kinzie 2011).

Professional Development Examples: Institute for Student Learning Assessment. With the goal of providing advanced training in student learning assessment, data analysis, and record keeping, in 2008 we initiated our own assessment conference, currently titled “The Institute for Student Learning Assessment.” Our conferences have involved a keynote speaker on a topic relevant to issues selected by the Office of Assessment and two sets of break-out sessions led by representatives from programs exhibiting good practice in assessment. In 2013, the Institute was expanded into a state conference, with session leaders from colleges and universities across Kansas and a national keynote speaker. In 2014 a keynote speaker from the Lumina Foundation introduced the Degree Qualifications Profile.

This year’s state-wide Institute included four sets of break-out sessions: Best Practices of Assessment for University Level Issues; Best Practices of Assessment for Faculty; Best Practices of Assessment for Program Assessment Coordinators; and Best Practices of Assessment for Student Life Units. The purpose for expanding to a state-wide conference was twofold: to bring new ideas to expand understanding of student learning assessment at Kansas State, and to bring together assessment coordinators from other Kansas institutions to learn from each other through discussion of common issues and solutions.

Professional Development Examples: Assessment Showcase. The Assessment Showcase is an opportunity to recognize

outstanding efforts by programs to enhance student learning assessment and to close the loop by making program improvements. The Director of the Office of Assessment uses two sources in selecting award recipients: recommendations from each college’s assessment coordinator and observations drawn from the annual program progress reports. Awards recognize efficient assessment processes, improvements made to assessment processes, program improvements implemented in response to assessment results, and clarity in assessment reporting. The provost presents seven to ten assessment awards each year to individuals or programs. These awards are often displayed in the offices of recipients, indicating their perceived value. This year we are asking each recipient to present a short description of his or her efforts as a means of sharing good practice.

Professional Development Examples: In-Service Meetings Specific for Programs. One important element of developing a paradigm of assessment across campus is meeting each program and its faculty on their own terms. Disciplinary autonomy is one of the most important conditions for success of a university-wide assessment system, and with it comes respect for the academic value of each program’s curriculum and mission. Program assessment processes must reflect the unique ways that students demonstrate learning and not be forced into a one-size-fits-all assessment system (Ewell et al., 2011).

Dictating one mechanism for assessment or one way of representing learning of an outcome can cause programs to lose ownership of their assessment process and student learning to be misrepresented by nonauthentic assessments.

The Office of Assessment staff has met annually with individual program representatives to collaborate on assessment processes and to encourage programs to choose assessments that suit their curricula, accreditation requirements, and faculty culture. We have seen involvement in student learning assessment reporting increase from 60 percent to 100 percent of our programs as a result of individual attention and respect for programmatic needs.

Conclusion

The culture of assessment desired for higher education must be centered on providing evidence to document learning that is authentic for each discipline. This culture can be enhanced through focused professional development opportunities that, like courses in a well-planned curriculum, are mapped to specific outcomes. Well-planned professional development creates the foundation for effective assessment practice. ■

References

- Ewell, P., Paulson, K., and Kinzie, J. 2011. “Down and In: Assessment Practices at the Program Level.” <http://www.learningoutcomesassessment.org/documents/NILOAsurveyreport2011.pdf>
- Miller, M. 2012. “From Denial to Acceptance: The Stages of Assessment.” <http://www.learningoutcomesassessment.org/FacultyEmbracingLOA.htm>

Frederick Burrack is director of the Office of Assessment and professor of Music Education and Christopher Urban is assistant director of the Office of Assessment at Kansas State University.

Involving Undergraduates in Assessment

(continued from page 4)

valuable feedback. Collecting student feedback is SATAL students' main job, and they have experienced firsthand how much students struggle to provide constructive and useful feedback when they assess a course or program. Some students benefited from the presentation more than others by paying close attention and applying the rubric during the activities. The fact that FI was presented by peers added some extra value to the importance of providing constructive feedback. Overall, SATAL students

improves students' reflections on their learning. Results indicate that students could benefit from direct instruction on how to provide constructive feedback with the aid of a rubric, because most of the participants indicated they found the rubric useful and that they "will utilize it in the future." Although an anecdotal concern, FI does take away time from class; to address that issue, more research is needed to provide evidence of potential positive effects of FI on students' critical thinking skills.

The lines of evidence demonstrated that scaffolding students' comments by providing them a rubric and modeling how to improve their comments on the board were key activities noted by the students and instructors as very helpful.



recognize the importance of investing some class time to train students on how to give constructive feedback, and in particular for peer-review exchanges in writing classes.

Helpful Hints

The lines of evidence demonstrated that scaffolding students' comments by providing them a rubric and modeling how to improve their comments on the board were key activities noted by the students and instructors as very helpful. Students' engagement with FI was very high (89 percent). The top three most used criteria were (1) offer specific suggestions that model appropriate behavior, (2) focus on content rather than on the person, and (3) provide a balance of positive and negative feedback. Another gain identified in the study is that by providing and receiving better feedback, students ultimately become better writers. For course evaluations, some evidence suggests that feedback training

Applications

The specific context of a relatively new university that enrolls a majority-minority undergraduate population and routinely appoints a high percentage of new faculty each year requires a highly contextualized approach for SATAL assessment of teaching and learning. The FI provided predominantly freshman students with a rubric and strategies that develop feedback skills to navigate a variety of learning activities effectively. FI has proven to have a three-way win for the SATAL students, their peers in the class, and faculty. FI findings have been shared through SATAL presentations during faculty and staff meetings and symposia on campus. In particular, we have encouraged faculty teaching predominantly freshmen to review students' most-used rubric criteria in class and consider the benefits resulting from training students to provide constructive feedback for their day-to-day teaching practices as well as final course evalua-

tions. One of seven axioms of classroom assessment noted by Angelo (2008) is stated, "If an assessment is worth doing, it's worth teaching students how to do it well."

Acknowledgments

I would like to acknowledge UC Merced Merritt Writing Program directors Robert Ochsner and Anne Zanzucchi; assessment committee members and their chairs Angela Winek and Michele Toconis; and Cindy Chavez, Cheryl Finley, Pam Gingold, and Jane Wilson, who offered their courses and input in all phases of the project. I gratefully acknowledge the work of the SATAL students: Charlesice Hawkins, Michael Pham, Valorie Smart, Cherrylyn Cruzat, Greg Dachner, and Tatsiana Verstak for their dedication to the project, perspective on activities, and assistance with data collection. ■

References

- Angelo, T. A. 2008. "Doing Assessment as if Learning Matters Most: Simple, Practical Classroom and Course-level Approaches." The 2008 Assessment Institute in Indianapolis. Materials for a Workshop Session. Retrieved from https://www.units.miamioh.edu/celt/resources/strategies_documents/CATS_angelo1.pdf.
- Brinko, K. 1993. "The Practice of Giving Feedback to Improve Teaching: What is Effective?" *The Journal of Higher Education* 64 (5): 574–593.
- Hattie, J., and Temperley, H. 2007. "The Power of Feedback." *Review of Higher Education* 77 (1): 81–112.
- Panadero, E., and Jonsson, A. 2013. "The Use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review." *Educational Research Review* 9: 129–144.

Adriana Signorini is assistant director of the Center for Research on Teaching Excellence at University of California, Merced.

Table 1. The senior thesis rubric				
	Level 1	Level 2	Level 3	Level 4
Focal question or hypothesis	Not clearly stated	Stated but unfocused, too broad, or too simplistic	Clearly stated, at an appropriate level of complexity	Significant and/or creative focus
Rationale/ motivation	No clear rationale or a weak rationale for the project	Some rationale presented, begins to motivate the work	Rationale makes clear why topic is worth investigating	Persuasive and creative rationale
Approach/ methodology	Not clear what was done or why, or an inappropriate method	Approach is generally appropriate for the topic	Clearly described and justified, well-chosen and properly executed	Creative and sophisticated methods
Scholarly context	Author overrelies on too few sources, or refers to published work without citations	Author demonstrates some awareness of a range of relevant literature	Author demonstrates broad awareness of the literature, including works presenting other perspectives	Author situates own work in a way that makes a contribution or identifies a new direction for investigation
Position	Does not take a clear or defensible position	States and/or critiques a position that may already be in the literature	Effectively supports, tests, extends, or critiques a position that may already be in the literature	Develops a clear and defensible position of his/her own
Argument	Weak, invalid, or no argument, perhaps a simple assertion	Some arguments valid and well supported, some not	Main arguments valid, systematic, and well supported	Arguments both well supported and genuinely compared to conflicting explanations
Use of evidence	Mostly relies on assertions or opinions rather than evidence, or evidence not clearly presented	Some appropriate use of evidence but uneven	Feasible evidence appropriately selected and not over-interpreted	Fully exploits the richness of the data/ evidence/ideas, and is sufficiently persuasive
Analytical insight	Treats related ideas or data as unrelated, or draws weak or unfounded connections	Begins to establish connections and perceive implications of the material	Brings together related data or ideas in productive ways, discusses implications of material	Develops insightful connections and patterns that require intellectual creativity
Writing mechanics				
Grammar, spelling, usage	Significantly impairs readability	Frequent or serious errors	Some minor errors	Virtually no errors
Organization	Needs significant reorganization	Structure is of inconsistent quality, may have choppy transitions, redundancies, or discontinuities	Structure supports the argument, clearly ordered sections fit together well	Structure enhances the argument; strong sections, seamless flow
Clarity, style, readability	Gets in the way of reading for content	Style is inconsistent or uneven	Effective prose style, follows relevant scholarly conventions, emergence of voice	Mastery of the genre, including elegant style, established voice
If this were a thesis at my institution, I would give it a grade of: A+, A, A–, B+, B, B–, C+, C, C–, D, F				

A Cross-Disciplinary Approach to Evaluating the Senior Thesis

(continued from page 2)

To explore the reliability of the rubric in the context of the original norming session, certain theses were read independently by two evaluators. During that session, all eleven categories showed an equal range of variability in scores assigned, as did papers in all four disciplinary areas (natural science, cognitive science, social science, humanities). We were also interested in the reliability of the rubric in the absence of norming. In the second evaluation exercise mentioned above—which did not include a norming session—both rubric scores and grades assigned to a single thesis varied to a considerable degree. For example,

As a next step, we decided to carry out a large-scale evaluation project to improve the robustness of our analysis. This time, in spring 2009, nineteen faculty members from seven institutions and ten disciplines used the rubric to evaluate 320 senior theses (with each evaluator reading theses from all institutions except his or her own). The project ended in spring 2013, when six of those institutions held sessions at which their own faculty members used the rubric to evaluate theses written by their own institution's students (but not by their own advisees). Sessions in both years began with norming exercises and included

cent received a B+, and only 8 percent received a B or below. One simple and pleasant explanation for this grade compression would be that by the time students write a senior thesis, most are doing excellent work. However, the rubric scores contradict this impression. When read by independent evaluators, theses that had received As from faculty advisers received rubric total scores ranging from 44 (the maximum points possible) down to 15 (just above the minimum possible score of 11).

As many faculty readers mentioned, grades usually include some evaluation of process, such as how the student went about working on the thesis, or how much they learned from the experience. Several individuals referred to the challenges involved in combining the assessment of product and process, as when a weak student works diligently and learns a lot, exceeding the adviser's expectations, but still writes a fairly weak end product.

The rubric is intended to evaluate the product exclusively, without the reader knowing anything about the student or the process she or he went through. Some faculty readers found this difficult, despite knowing that their own professional writing is evaluated in a similar environment (by journal editors and grant reviewers). Discussions in norming sessions revealed a tendency for some readers to make assumptions that gave the unknown student author the benefit of the doubt, resulting in scores that even the people giving them considered somewhat inflated with respect to the paper itself. Sometimes these assumptions took the form of giving the student credit for making a good effort, even though there was no way of knowing whether the unidentified student had worked hard or not. In another case, a reader made allowances for what she knew about how a particular subject was taught at her institution and what the student would have been encouraged to do. However, the rubric is intended to be applied in such a way that its results

Readers who were accustomed to stronger students gave lower grades and lower rubric scores than readers from institutions with weaker students.



one thesis received total rubric scores ranging between 19 and 26 (on a scale of 11–44), and grades ranging from 2.0 to 3.33 on a 4-point scale. We conclude that just using a rubric, in the absence of norming, does not result in good interrater reliability. Apparently, the same is true of grading.

To examine the nature of the variation in scores, we asked the evaluators to identify themselves by their home institution. The scores and grades they assigned were then compared to the average SAT score at their home institution. Readers who were accustomed to stronger students gave lower grades and lower rubric scores than readers from institutions with weaker students. In the absence of norming, the background of the reader will affect the scores given. Ideally, theses should be evaluated by readers who do not know the actual student author but who are familiar with students of similar academic ability.

double readings of a number of theses; most differences in scoring in these cases were minimal.

Readers in these years also assigned letter grades as a measure of overall quality. Mean scores for separate components ranged between 2.3 and 2.6, except for grammar, which had a mean of 3.2. Standard deviations averaged approximately 0.78. Excluding grammar, Level 1 scores made up 6–18 percent of each component and Level 4 made up 6–13 percent. Compared to rubric scores, grades were more skewed toward the top end of the scale, with 41 percent of theses receiving an A+, A, or A–, and 19 percent receiving a C+ or below.

We also collected the grades given to each thesis by each student's advisor. Across the five institutions that assign letter grades, the grades students received on their senior theses were highly compressed; 60 percent received an A and 22 percent received an A–. Ten per-

can be used to evaluate the effectiveness of the teaching students received, and that judgment can only be meaningful if readers do *not* make allowances for processes that might account for work not being of optimal quality.

One value of the rubric, then, is that it evaluates the work *without* evaluating the student. Some students may have to work very hard and learn a great deal to write a Level 2 paper, and they should probably get credit for that achievement in other contexts. But on the rubric their work should still be judged a Level 2. Use of the rubric makes possible a detached evaluation of the actual quality of student work. This knowledge permits an institution to consider how and what it teaches students to do and whether it does so effectively—that is the value of assessment as an activity distinct from grading and done by someone other than the student's teacher or adviser.

In addition to the summative assessment of the efficacy of the senior thesis at

a departmental or institutional level, there are other possible uses of the rubric. For example, it could serve as a vehicle for promoting conversations about learning outcomes among faculty, including during new faculty orientation. It can also be used for advising students who are in the process of writing senior theses, to help them understand what characteristics make a thesis of high quality.

Both the norming sessions and incidental discussions during the 2013 readings resulted in productive discussions of issues, such as whether a Level 1 thesis should pass, whether Level 2 as defined is too broad, and whether Levels 2 and 3 are too far apart. Some of the ideas raised in these discussions were used to edit the rubric; others could provide a basis for developing a local version that is better suited to the needs or interests of a particular institution.

The senior thesis rubric has by now been used and improved by readers from many institutions and many analytical

disciplines. It appears to work well; none of the 2013 readers had trouble with it. However, several of them made good suggestions for possible improvements. In particular, readers at several institutions recommended adding a category that evaluated the clarity with which the author had stated the paper's objective, whether it was a hypothesis or a focal question. The modified version of the rubric is presented in Table 1.

Acknowledgments

The authors would like to thank all the evaluators who participated in the development and testing of the rubric, and also the Teagle Foundation for financial support of the 2009 and 2013 reading sessions. ■

Carol Trosset is director of enrollment research at Bennington College in Vermont and Steven Weisler is the former provost at Dominican University of California.