

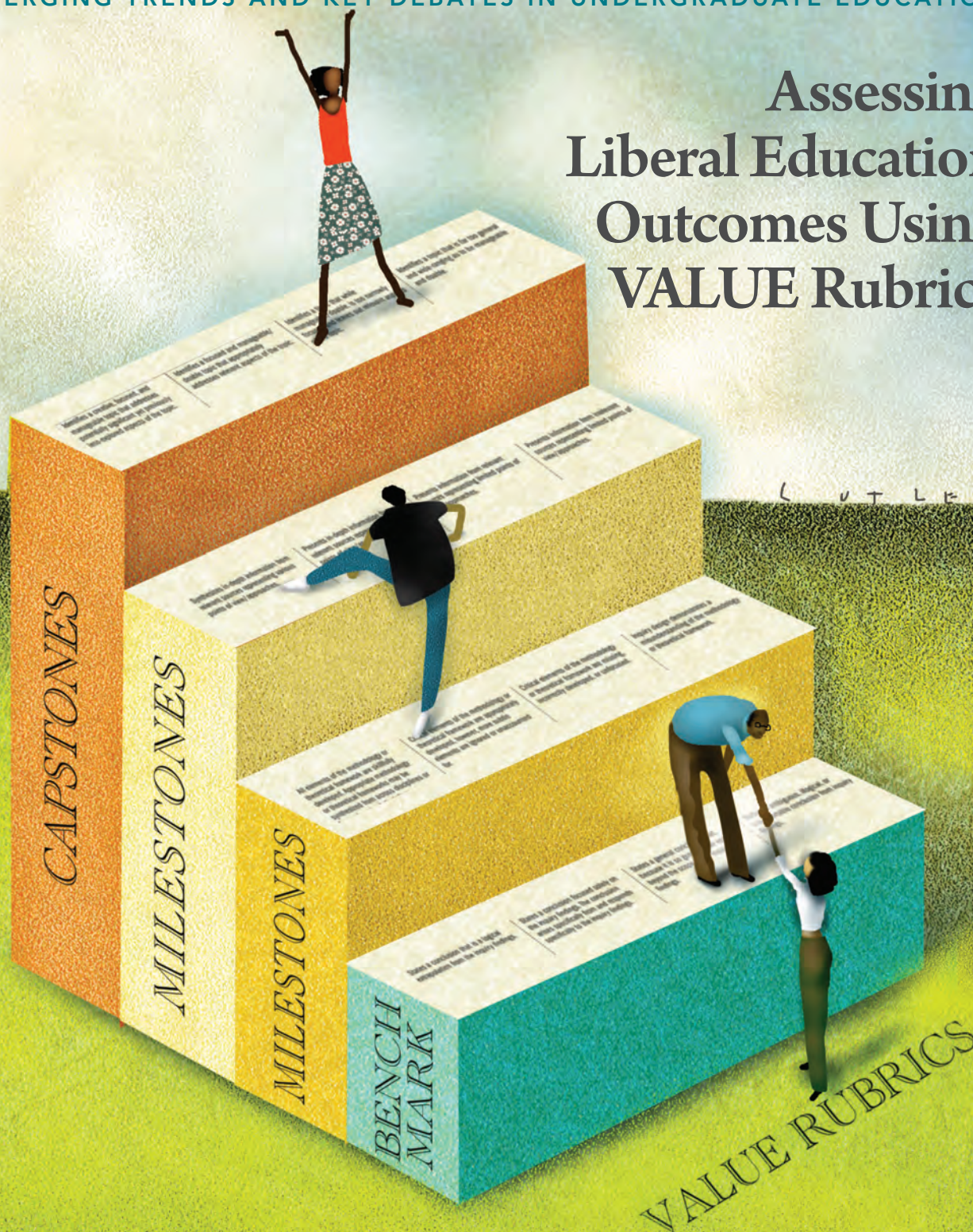
VOL. 13, NO. 4/ VOL. 14, NO. 1

FALL 2011/WINTER 2012

peerReview

EMERGING TRENDS AND KEY DEBATES IN UNDERGRADUATE EDUCATION

Assessing Liberal Education Outcomes Using VALUE Rubrics



A PUBLICATION OF THE ASSOCIATION OF AMERICAN COLLEGES AND UNIVERSITIES



CONTENTS

Vice President for Communications and Public Affairs

Debra Humphreys

Editor

Shelley Johnson Carey

Associate Editor

Wilson Peden

Design & Production

Darbi Bossman

Cover Illustration

David Cutler Studios

Editorial Advisory Board

James A. Anderson

Fayetteville State University

Randy Bass

Georgetown University

David A. Berry

Community College Humanities Association

Norman Coombs

Rochester Institute of Technology

Peter Ewell

National Center for Higher Education Management Systems

Ann S. Ferren

Association of American Colleges and Universities

Mildred García

California State University–Dominguez Hills

Richard Guarasci

Wagner College

Elise B. Jorgens

College of Charleston

Adrianna J. Kezar

University of Southern California

Ann Leffler

University of Maine

Donna Maeda

Occidental College

David E. Maxwell

Drake University

Catherine Hurt Middlecamp

University of Wisconsin–Madison

Chandra Talpade Mohanty

Syracuse University

John P. Nichols

Saint Joseph's College

G. Roger Sell

Missouri State University

Joan Straumanis

Association of American Colleges and Universities

Beverly Daniel Tatum

Spelman College

peerReview

EMERGING TRENDS AND KEY DEBATES
IN UNDERGRADUATE EDUCATION

VOL. 13, NO. 4/VOL. 14, NO. 1 | FALL 2011/WINTER 2012

Published by the Association of
American Colleges and Universities ©2012
1818 R Street, NW • Washington, DC 20009
202.387.3760 • www.aacu.org

ISSN: 1541-1389

FALL 2011/WINTER 2012

From the Editor 3

ANALYSIS

Emerging Evidence for Positive Change

Terrel L. Rhodes, Association of American Colleges and Universities 4

PRACTICE

Measuring Student Progress with E-Portfolios

J. Elizabeth Clark and Bret Eynon, LaGuardia Community College 6

Assessing General Education Learning Outcomes

Linda Siefert, University of North Carolina Wilmington 9

Assessing Early Integrative Learning

Bernice Braid and Gladys Palma de Schrynmakers, Long Island University,
Brooklyn Campus; Alan W. Grose, The Washington Center for Internships
and Academic Seminars 12

Fostering Integrative Knowledge and Lifelong Learning

Melissa R. Peet, University of Michigan–Ann Arbor; Laura Reynolds-Keefer, University of
Michigan–Dearborn; Patricia Gurin, University of Michigan–Ann Arbor; and
Steve Lonn, University of Michigan–Ann Arbor 15

Staying on Track with Rubric Assessment

Megan Oakleaf, Syracuse University 18

Using VALUE Rubrics to Evaluate Collaborative Course Design

Andrea Greenhoot and Dan Bernstein, University of Kansas 22

Increasing the Validity of Outcomes Assessment

Kathleen Langan Pusecker, Manuel Roberto Torres, Iain Crawford, Delphis Levia,
Donald Lehman, Gordana Copic; all of the University of Delaware 27

RESEARCH

How Reliable Are the VALUE Rubrics?

Ashley P. Finley, Association of American Colleges and Universities 31

REALITY CHECK

Our Primitive Art of Measurement

Trudy Banta, IUPUI 35

Annual subscription rates are \$35 for individuals and \$45 for libraries. Single issues are \$8/\$10; bulk discounts are available. For additional information or to place an order, visit us online or call 800.297.3775.

www.aacu.org/peerreview

Peer Review (ISSN-1541-1389) is published quarterly by the Association of American Colleges and Universities, 1818 R Street, NW, Washington, DC 20009-1604. Annual rates are \$35 for individual subscriber and \$45 for libraries. Periodicals postage paid at Washington, DC, and at additional mailing offices. Copyright 2012. All rights reserved. POSTMASTER: Send address changes to: *Peer Review*, 1818 R Street, NW, Washington, DC 20009-1604.



As institutions are asked to document the quality of student learning, the VALUE (Valid Assessment of Learning in Undergraduate Education) project is helping them define, document, assess, and strengthen student achievement of the essential learning outcomes that stand at the center of the Association of American Colleges and Universities' Liberal Education and America's Promise (LEAP) initiative. Recognizing that there are no standardized tests for many of the essential outcomes of an undergraduate education, the VALUE project has developed rubrics for faculty and institutions to collect convincing evidence of student learning.

While the word rubric has recently been associated with education and assessment, the term has an interesting etymology. According to Roman technical writers Columella, Vitruvius, and Pliny, in ancient times, *rubrica* was the name given to the red clay line that carpenters marked on wood as guides before making cuts. In medieval times, the term held particular meaning regarding religious texts and calendars. On those calendars, Saint's days, feasts, and other holy days were hand lettered by monks in red ink and called rubrics—the origin of the expression “red letter days.” It is in the spirit of these two definitions that the term rubric is now used in an educational context—both acknowledging rubrics as significant measurement guides for teaching and learning and signaling the importance of this assessment tool.

In this issue of *Peer Review*, we feature articles from a range of campuses across the country on their use of the VALUE rubrics to assess student learning and to engage both faculty and students in a broad range of strategies for powerful assessment.

- At the University of North Carolina–Wilmington, the use of course-embedded assignments and faculty leadership are the linchpins of a learning assessment process begun in the general education program. Four VALUE rubrics are being used initially across the curriculum. The attention to producing reliable results has provided useful information on student learning for improved pedagogy as well as evidence that supports the integral role of faculty in meaningful assessment.
- The Brooklyn Campus of the Long Island University, a private, primarily minority-serving institution, has found the Integrative Learning VALUE rubric to be particularly useful in helping students connect learning in formal courses with community-based

(and life) learning, commencing when students enter the university. Further, the rubric has allowed faculty and curriculum development to become a positive, collaborative endeavor that is creating a more coherent curriculum and learning experience for students.

- The University of Michigan presents an example of how a large, multicampus university used the VALUE rubrics and adapted them to their own mission and purposes to develop an assessment instrument that is being used on two campuses and across multiple colleges within Michigan, and in both undergraduate and graduate programs. The initial pilot has resulted in the award of a FIPSE grant that will allow further development and testing through a six-campus consortium.
- Syracuse University is leading another consortium of five campuses (later to be expanded to ten campuses) linking faculty, librarians, and other information literacy professionals in enhancing and assessing student information literacy.
- The University of Kansas has engaged in an extensive research initiative to improve students' written communication and critical thinking skills. Concentrating on large enrollment courses, faculty intentionally designed assignments and courses for students to demonstrate their learning that could then be evaluated through multiple measures—the use of grades, a standardized exam (the Collegiate Learning Assessment), and the VALUE rubrics.
- Trudy Banta reminds us of the need for multiple measures of learning that can also provide guidance for improvement in curriculum and instruction, and the limitations of our proclivity to engage in institutional comparisons based on inadequate and inappropriate measures.

Since the VALUE rubrics were made available for download on the AAC&U website starting in 2010, nearly 11,000 first-time visitors have accessed them. As such, the collection of articles in this issue provides just a glimpse of the rich work being done on campuses with these rubrics. These efforts are helping institutions give students a much-needed compass by assessing essential learning outcomes and providing meaningful feedback that will foster student success throughout their undergraduate educations and beyond.

—SHELLEY JOHNSON CAREY



Emerging Evidence on Using Rubrics

► **Terrel L. Rhodes**, vice president for quality, curriculum, and assessment, Association of American Colleges and Universities

In 2009, the Association of American Colleges and Universities (AAC&U) publicly released a set of rubrics for evaluating achievement of a wide array of cross-cutting learning outcomes. These rubrics were developed as part of AAC&U's VALUE (Valid Assessment of Learning in Undergraduate Education) project—part of AAC&U's Liberal Education and America's Promise (LEAP) initiative. The VALUE rubrics were developed as an alternative to the snapshot standardized tests used by many to assess student learning. Since September 2010 when we started tracking the data, well over three thousand different organizations have downloaded one or more of the VALUE rubrics and are using them. Campuses are now beginning to report findings based on their use of these rubrics to assess student learning across a set of fifteen outcomes, all essential for student success in the twenty-first century.

AAC&U developed the VALUE initiative to test the proposition that cross-cutting rubrics could be used as a viable alternate approach to the existing tests, (e.g., ETS Profile, the Collegiate Assessment of Academic Proficiency, and the College Learning Assessment) for accountability reporting purposes. More importantly, there could be approaches to assessing student learning that would yield important information about the quality of student performance over time that faculty and others could use to improve pedagogy and practice in the classroom and beyond.

CORE EXPECTATIONS FOR LEARNING

What the VALUE initiative established was that faculty members actually share core expectations for learning, regardless of type of institution or region of the country. Further, VALUE showed that these shared expectations exist for a broad array of student learning outcomes; that the work we ask students to do through the curriculum and cocurriculum is the best representation of our students' learning; and that faculty need to be directly involved in assessing the quality of the learning.

With the emergence of the Degree Qualifications Profile in January 2011 (Lumina 2011) and its focus on the integrity of the academic degree, the importance of specified levels of performance across five areas of learning—specialized and broad/integrative knowledge, intellectual skills, applied learning, and civic learning—has surfaced as a core requirement for student success in higher education in this country. The VALUE rubrics are providing a means by which campuses and their faculty can create common standards for evaluating the quality of performance expected for attainment of specified degree levels, e.g. Associate or Baccalaureate degrees. In essence, what has emerged is a framework of quality standards without standardization.

The work we ask students to do through the curriculum and cocurriculum is the best representation of our students' learning; and that faculty need to be directly involved in assessing the quality of the learning.

The VALUE rubrics were developed as broad, institutional level, or “meta” rubrics, and we recognize that rubric use at program, disciplinary, or even classroom levels requires adapting the rubrics with more specific language and purposes. For example, using rubrics to grade assignments does not necessitate the need for precise, psychometric methods. In developing and testing the rubrics, the faculty and professional staff development teams anchored the demonstrated learning for students at the entry and exit points for degree attainment. To provide signposts that learning was moving in the desired direction, milestones were developed between the benchmark—where rubric developers found their students' current performance on average—and the capstone—where rubric developers hoped their students would be to attain a baccalaureate degree.

By having faculty on over one hundred campuses test the rubrics with student work from a variety of classes and disciplines, the



rubrics were refined based on the feedback and found to be useful across the curriculum and cocurriculum for assessing student progress. What was found through the development process was that the commonly shared performance levels for learning in the rubrics reflected what faculty were looking for in student work and that the rubrics were sufficiently generic and flexible to be successfully used across the disciplines, hence articulating usable standards for attainment of learning for the outcomes.

The validity of the VALUE rubric approach is attested to by the rapidity with which campuses have been adopting and adapting the rubrics for use in courses, programs, and whole institutions—the rubrics resonate with faculty. All types of colleges and universities—both two year and four year—report that the use of the VALUE rubrics has catalyzed rich conversations around student learning and assessment. Some campuses have modified the language of the rubrics to reflect their own culture and mission, some have added dimensions to the rubrics, and others are using them “off the shelf.” Many campuses use the rubrics to assess collections of student work, often choosing to use an e-portfolio for students to present work to be assessed. Most of the leading e-portfolio vendors have incorporated the VALUE rubrics into their software as an option for campuses to utilize for evaluation and reporting on their students’ learning.

TESTING RUBRIC RELIABILITY

Issues of reliability are frequently raised when choosing to use rubrics to assess learning. In the initial VALUE development, a national panel of thirty individuals who had not seen the rubrics or been involved with their development—administrators, faculty, student affairs professionals, local community members, and teachers—was assembled for three days to test use of the rubrics in assessing work in a set of student

e-portfolios. The inter-rater reliability results exceeded the 0.8 standard routinely used for such measures. In a more recent pilot project that relied upon a virtual, truncated inter-rater reliability approach, a group of forty faculty members from a variety of campuses across the country assessed a sample of student work using three VALUE rubrics. From this pilot, a composite reliability score was calculated for all faculty scorers across four broad disciplinary areas: humanities, social sciences, natural sciences, and professional and applied sciences (Finley 2012).

EXAMPLES FROM THE FIELD

This issue of *Peer Review* focuses on campus-level use of the VALUE rubrics and the findings that have resulted from that use. These examples of how the VALUE rubrics are being used on campuses across the country to assess student learning and to engage faculty, students, and others in evaluating learning represent a broad range of strategies for powerful learning assessment. Three of the examples involve consortia of campuses that are using the rubrics to connect their individual campus efforts, and these articles define the benefits of those collaborative endeavors that contrast the ways institutions currently are ranked or compared. The other articles detail the benefits and challenges of using rubrics on different types of campuses and the ways in which the VALUE rubrics have helped campuses enhance student learning.

The Connect to Learning project involves twenty-two public and private, community college, research, comprehensive, and liberal arts colleges and universities focused on connecting traditional indicators of student success, such as retention and graduation, with more nuanced assessment of learning through student work and reflective practice. This three-year FIPSE-funded project is in partnership with the Association of Authentic, Experiential and Evidence-Based Learning

and utilizes e-portfolios of student work and the VALUE rubrics. This national network of campuses is a collective, recursive knowledge-generation project linking teaching strategies with evaluation of student achievement levels.

Through a collection of examples in this issue of *Peer Review*, we see that assessment can be and is a high-impact practice capable of improving learning and pedagogy on our campuses. Through the use of VALUE rubrics, campuses are demonstrating that students benefit from knowing what is expected of them; that faculty leadership emerges through the development of rubrics and e-portfolios; that there is an important relationship between intentionality, assignments, and learning; that there is value in basing assessment on student work; and that there is an emerging culture shift among faculty from “my work” to “our work.”

The validity and reliability of this on-going, developmental rubric/portfolio approach to assessment has demonstrated the usefulness, the need, and the reality of meaningful faculty-led methods and approaches to assessing student learning—approaches that represent the integrity of a student’s particular degree not solely in terms of completion, but more importantly in terms of the quality of the learning. ■

NOTE

Copies of the VALUE Rubrics can be found at <http://www.aacu.org/value>.

REFERENCES

- Adelman, C., P. Ewell, P. Gaston, and C. G. Schneider. 2011. *The Degree Qualifications Profile*. Indianapolis: Lumina Foundation.
- Finley, A. Forthcoming. *Making Progress? What We Know about the Achievement of Liberal Education Outcomes*. Washington, DC: Association of American Colleges and Universities.
- Lumina Foundation. 2011. *Degree Qualifications Profile*. Indianapolis, IN: Lumina Foundation. http://www.luminafoundation.org/publications/The_Degree_Qualifications_Profile.pdf.



Measuring Student Progress with E-portfolios

► **J. Elizabeth Clark**, professor of English, LaGuardia Community College

Bret Eynon, assistant dean, Center for Teaching and Learning, LaGuardia Community College

In his “An Essay on Criticism,” Alexander Pope defines wit thusly: “True wit is nature to advantage dressed, /What oft was said, but ne’er so well expressed/Something whose truth convinced at sight we find, /That gives us back the image of our mind.”

In higher education today, “What oft was said, but ne’er so well expressed” serves as a compelling metaphor for the progressive uses of institutional rubrics. Are our institutions doing what we say they are? Institutions often tout student progress and achievement based on new initiatives and proven programs and approaches built over time. What happens, however, when we connect these claims to authentic student work? Do they align? Can we talk about student work and achievements across institutions? Why would we want to do so?

Faculty, students, and administrators are engaged in exciting new conversations about definitions of educational progress. While many faculty have long used rubrics to measure student achievement in a course or on a specific assignment, institutions are increasingly using rubrics to assess authentic student learning based on student-produced work connected to larger institutional outcomes. Because of the national, collaborative work in creating the Association of American Colleges and Universities’ VALUE rubrics, these rubrics lend themselves to extending the conversation about work in individual courses, in programs and majors, in institutions, and across institutions.

ACADEMIC PATHWAYS THROUGH HIGHER EDUCATION

The flexibility of the VALUE rubrics allows them to be used in different segments of higher education and adapted to serve multiple purposes. Students are the ultimate stakeholders and beneficiaries, and so a key national issue in higher education today is helping students to see a clear pathway to graduation and to achieving their personal academic and vocational goals.

Too often, students see their overall education as a disconnected series of courses and experiences. Situating a wide range of skills

and areas of knowledge over a four-year scale, institutions can use the VALUE rubrics to help students see how their skills and knowledge build sequentially over time, from course to course, and sometimes, from institution to institution.

Jason C. Evans, chair of the English department at Prairie State College, believes the VALUE rubrics are helpful for engaging students in a common language around academic expectations. “Since they’re widely used in two- and four-year settings, we can feel comfortable knowing we’re more or less in line with other colleges. They provide a helpful common language for our students to understand what they’re learning and why.”

Natalie McKnight explains that in Boston University’s College of General Studies (CGS) “The rubric is important because it makes faculty and students more conscious of what we are trying to accomplish; knowing and discussing our goals enables us to reach them more successfully. Plus students, faculty, and administrators can use the rubric as a gauge to assess the impact of our program, something we can all benefit from knowing.” She adds, “Articulating goals and then evaluating how well one has achieved those goals are essential to progress and learning in any field. The VALUE rubrics have played a key role in that process for us, as have e-portfolios.”

For students, the four-year scale of the VALUE rubrics allows them to understand their present learning in connection with future learning at their present or future institutions. Dean Susan Solberg of Prairie State College believes the power of the VALUE rubrics lies in this inter-institutional credibility. “I especially like the fact that we can say that we are adopting standards that have national credibility because many college and universities have worked on them collaboratively.” Sarah Burns-Feyl, Joy Tatusko, and Linda Anstendig of Pace University note, “The VALUE rubrics provide a gold standard for assessing learning outcomes and are grounded in theory and practice. After all, these rubrics were developed by teams of faculty over an eighteen-month period, and have been tested in various settings.” Knowledge building becomes



more than situational at a particular college or university, communicated instead as a cultural value and a benchmark of higher education.

ADAPTING THE RUBRICS TO MEET LOCAL NEEDS

A key component of VALUE rubrics is their cross-institutional construction and transferability to local campuses as institutions transform the rubrics to make them relevant for local student learning.

Alison S. Carson explains how Manhattanville College adapted the rubrics to find new ways to talk with students and faculty about evaluating students through their distinctive portfolio program, transforming from paper to e-portfolios: “We began with the recognition that the rubric had to work for our unique portfolio requirements and learning outcomes, as well as be user-friendly, which meant not too long.”

Burns-Feyl, Tatsuko, and Anstendig of Pace University also believe that adapting nationally normed rubrics for a specific, local educational mission allows them to think differently about student learning.

At Pace, we have had various initiatives and pilots to assess learning outcomes (especially related to our last Middle States accreditation evaluation), but now as our ePortfolio program evolves, we have the opportunity to pursue assessment in a much more systematic way, and the VALUE rubrics (as well as our adaptations of them) are a crucial part of this process.

ENGAGING STUDENTS IN SELF-ASSESSMENT

Another important emerging trend is using the rubrics with students. Rather than just “assessing” students, some colleges and universities are asking students to engage in self-reflection and self-analysis. How do the rubrics allow them to understand their own growth and development over time?

Boston University’s CGS introduces students to a modified VALUE rubric in their freshman year. At the end of the year, students write a self-assessment essay or a video analyzing their own development.

At Stony Brook University, Nancy Wozniak engages her student e-portfolio consultants in the use of the VALUE rubrics focused on measuring “intensity” in student e-portfolios over time.

UNDERSTANDING STUDENT LEARNING

The VALUE rubrics have also provided an important benchmark for emerging work in e-portfolio research in individual courses, in college-wide e-portfolio programs, and across emerging national networks.

At Manhattanville College, Carson explains that the VALUE rubrics provide a way to establish a “more consistent, objective and transparent way of evaluating Portfolios.” The college has decided to use “the integrative and written communication VALUE rubrics to align with the college’s learning outcomes for their Portfolio system.”

Virginia Tech is an early leader in the use of e-portfolio rubrics to measure student learning. Marc Zaldivar explains the university’s pilot of three different rubrics to evaluate learning gains in first-year experience and the way the institution then hopes to use the emerging data:

The VALUE rubrics are a constant resource for Virginia Tech. As we work with new programs that are exploring assessment or learning with e-portfolio frameworks, we use these as a starter guide to defining specific learning outcomes (from the LEAP initiative) and how to assess those outcomes with the VALUE rubrics. Our largest project to adopt them is our growing first-year experiences, which is a combined, cross-campus effort for our Quality-Enhancement Plan for our recent SACS

accreditation. In this, we are collecting artifacts and reflections from students on three specific outcomes: problem solving, integration of learning, and inquiry. For each of these outcomes, a modified version of the VALUE rubric will be employed to assess student learning in those three domains. The academic year 2010–2011 was our first year of this pilot effort, with six participating programs and roughly one thousand students. Assessment of those materials will begin in the next few months. The next academic year will expand to [assessment of] twelve different programs and approximately two thousand students. Our target is to reach 100 percent of the entering students, or roughly five thousand students, within the next three years. Other programs at Virginia Tech, such as the SERVE Living Community, are also exploring ways that other VALUE rubrics, such as the civic engagement and critical thinking rubrics, may be used to help them assess the effectiveness of their programs.

Like Manhattanville College and Virginia Tech University, LaGuardia Community College has also begun to use the VALUE rubrics to assess student learning. The college modified the information literacy rubric for its local research and information literacy (RIL) competency.

Recently LaGuardia conducted college-wide Benchmark Assessment Readings to assist the college in preparing for its Middle States Review comparing student work from students with twenty-five credits or less and forty-five credits or more. The college believes that the data gleaned from student artifacts will provide an additional understanding of the extent of student learning of general education outcomes. In an examination of student work submitted for the RIL competency, an interdisciplinary group of faculty scorers



found that students made a gain of +1.45 on a six-point rubric, demonstrating key gains between an earlier and later point in students' tenures at the college.

For LaGuardia, this gain indicates that the college is effectively helping students to make gains in research and information literacy throughout the curriculum cumulatively over time. As students prepare to enter four-year colleges after graduation, they are better prepared for the rigors of research in the junior year. These findings are also helping the college to shape its conversation about student learning over time and preparing students for transfer to four-year institutions.

What emerged as one of the most significant findings of this evaluation process at LaGuardia, however, was the importance of reading and scoring student work as part of the college's faculty development. Faculty found this process most meaningful because of the way it helped them to understand their relationship to student learning, the college's core competencies, and the work of departments and disciplines. The assessment readings proved to be an enormously rich site of shared faculty exploration into the general education (core) competencies, and how those competencies play out in different courses. As a direct result of this process, several programs have already begun to revise assignments they use in order to ensure that what they ask students to do more effectively corresponds to the college's competencies and rubrics. More importantly, faculty were engaged in reading student artifacts, using the assessment interface, and defining the college's assessment outcomes.

INTEGRATIVE LEARNING AND E-PORTFOLIOS

Building on its work as a national e-portfolio leader, LaGuardia Community College recently launched the Connect to Learning project (C2L), a national three-year FIPSE-funded project coordinated by

LaGuardia's Making Connections National Resource Center in partnership with the Association for Authentic, Experiential and Evidence-Based Learning. The C2L project works with a national network of twenty-two campuses—community colleges, private colleges, and research universities—to engage in a collective and recursive knowledge-generation process. The project focuses e-portfolios on reflective pedagogy and student learning, correlating improvement on student success measures, such as retention, with more nuanced assessment of student work using the VALUE rubrics.

As LaGuardia Community College, Manhattanville College, Pace University, Boston University, Stony Brook University, Virginia Tech, Prairie State College, and other campuses in the C2L network expand their campus-based use of various VALUE rubrics, the C2L network as a whole will begin experimenting with VALUE as a tool for cross-institutional exchange. Highlighting the use of e-portfolio to support and surface students' longitudinal development across semesters and disciplines, the C2L network will consider ways to use the VALUE rubrics to examine student work and compare student learning in diverse institutional settings.

This past summer, the teams from twenty-two campuses sampled student e-portfolios in a structured conversation around integrative learning and e-portfolio at the C2L Summer Institute. The teams used the integrative learning VALUE rubric to explore the range of integration in student e-portfolios. Every campus contributed a sample student e-portfolio. The campus teams rated the student e-portfolios in teams and small groups and then discussed the characteristics of integrative learning.

What emerged in a large group discussion is the importance of intentional e-portfolio development on our campuses. Reflection and integration have to be

prompted and structured activities for students as they learn to be more reflective learners integrating their knowledge across disciplines. The group also concurred that asking students to integrate across disciplines can be particularly challenging when institutions do not model effective integration and cross-disciplinary practice for students. It is our collective hope that this will provide us with new ways to discuss e-portfolios as a pedagogical resource for higher education and point the way to future study.

BUILDING A RELATIONAL CULTURE IN HIGHER EDUCATION

Institutions of higher education have always been proficient at talking about what their campuses do well. The VALUE rubrics represent a radical shift, encouraging institutions to build a common discourse and to ground our exchange about pedagogy and innovation in a nuanced examination of student learning that these rubrics afford. Not meant to be prescriptive, the VALUE rubrics are broad and flexible enough to encourage campuses as diverse in their educational missions as Prairie State College, LaGuardia Community College, Manhattanville College, Pace University, Boston University, Virginia Tech, and Stony Brook University to assess student learning on their campuses in a relational way. They also present a unified view of the cultural, social, and economic importance of key skills like written communication and information literacy while also creating space for emerging concepts like integrative learning. This work allows us to talk concretely rather than abstractly about student learning and our shared vision for higher education in the United States. It "[G]ives us back the image of our mind," for what higher education has always aspired to: creating a community of educated individuals poised to meet tomorrow's challenges. ■



Assessing General Education Learning Outcomes

► **Linda Siefert**, director of assessment, University of North Carolina Wilmington

Much has been written about the current shift toward a knowledge economy and the resulting effects on our society and culture (Houghton and Sheehan 2000). Just as the practices of our business community are quickly evolving to keep pace with this shift, so is the way the education community thinks about assessing the learning of those who will function in this new knowledge economy. In the twentieth century, assessment relied on tests of *explicit* knowledge, or what we call content knowledge in education. Since content is now available quickly and inexpensively through electronic sources, simply knowing the correct answer no longer defines expertise. As educators, we need to prepare our students for success in life and in their careers by placing more emphasis on knowledge as transferable skills and abilities, such as the abilities to communicate thoughtfully and effectively, to think creatively and critically, and to access, evaluate, and use information to accomplish a purpose. To use the language of information and technology researchers, our focus is changing from assessing *codified knowledge* to assessing tacit knowledge (Stiles 2000). This requires more complex assessments that rely on authentic demonstrations and detailed and well-vetted rubrics. Whether these demonstrations are assessed individually from course-embedded assessments or in a purposeful offering like a portfolio, the processes are similar and the challenges remain the same.

FACULTY AS AN INTEGRAL PART OF THE PROCESS

The tacit knowledge and transferable skills that our faculty believe will prepare our students for the twenty-first-century workplace are reflected in the learning goals adopted by the University of North Carolina Wilmington (UNCW). Since curriculum and assessment are two sides of the same coin, our approach to assessing the UNCW Learning Goals places faculty at the center of the process. The faculty role begins within the general education courses. Each instructor of a course chosen for assessment selects a course assignment that is a regular part of the course and that addresses the specific learning goal. Since the assign-

ments are part of the course content and course grades, students are motivated to perform at their best. This also means that little additional effort is required on the part of course faculty. Another benefit is that there is a natural alignment often missing in standardized assessments, and results can be linked directly to the curriculum, making it straightforward to identify areas of strength and areas for reinforcement of learning.

Faculty members also do all of the scoring in the UNCW assessment process. The student work products sampled are scored independently from the instructor grades by trained scorers from across disciplines using common rubrics. A disadvantage to this approach (when compared to standardized tests) is that results cannot be compared to those from other institutions. We are mitigating this by using the AAC&U VALUE Rubrics (Rhodes 2010) for four of our learning goals. Our hope is that institutions begin to share their findings from the VALUE Rubrics so that cross-institutional comparisons can be made.

Preparing scorers is key to obtaining reliable results. Metarubrics, such as the VALUE rubrics, are constructed so that they can be used to score a variety of student artifacts across preparation levels, across disciplines, and across universities. However, the generality of a metarubric makes it more difficult to use than a rubric that is created for one specific assignment. At UNCW, faculty scoring volunteers initially attend a two-hour workshop on one rubric. At the workshop, they review the rubric in detail, are introduced to the assumptions we've adopted for applying the rubrics, and practice scoring benchmark papers. Subsequent discussion begins the process of developing a common understanding of the rubric. On the day of scoring, scorers work in groups of two or three. Scoring papers from an assignment begins with the members of the group independently scoring one piece of student work and discussing their scores. This discussion continues the norming before scorers proceed through scoring the assigned papers.

VALUE rubrics detail four levels of achievement—benchmark, two levels of milestones, and capstones. One of the most important



assumptions we make when using the VALUE rubrics is that we are comparing each separate work product to the characteristics we want the work of UNCW graduates to demonstrate—level 4 of the rubrics. So even when we are using the rubrics to assess work products from 100- and 200-level courses, we are comparing the work to our expectations for graduating seniors, not to other students in the course or even students of the same level. We have not, as yet, determined our exact expectations for scores on work from these lower-division courses. That is why the results presented here give the percent of work scored at or above each of the milestones, levels 2 and 3.

WHAT WE'VE LEARNED ABOUT STUDENTS' ABILITIES

UNCW has results from four of the VALUE Rubrics (written communication, inquiry, critical thinking, and information literacy) for our general education courses (mainly 100- and 200-level courses). We have discovered things about our students that are contrary to anecdotal information. For example, within the learning goal of written communication, students are not in general weaker in the control of syntax and mechanics than they are in other dimensions of writing, although this is an area often discussed by faculty as problematic. Instead, our results show that students struggle most with using sources to support ideas. Results also help point out relative strengths and weaknesses across the learning goals. Findings to date illustrate relative strengths in information literacy, followed by written communication and inquiry. Critical

thinking scores have shown the most need for improvement, and have also provided the most useful information for curriculum change. Details are provided for information literacy and critical thinking in tables 1 and 2.

INFORMATION LITERACY FINDINGS

We have assessed student work products on information literacy in one academic year, sampling seventy-eight work products from four sections of our culminating English composition class, one of two main general education courses that emphasize information literacy skills.

For this UNCW Learning Goal, the scores are fairly consistent across all dimensions of the rubric with respect to the percent of work products scored at or above a level 2. Relative strengths and weaknesses show up more clearly for the work scored at or above a level 3. At this milestone, we see, for example, that almost 60 percent of students in this course are able to use at least three of the information-use strategies that provide evidence of ethical and legal use of information (IL5). However, only 40 percent of the work products evaluated the information to the level of identifying the students' and others' assumptions when presenting a position (IL3). Almost half (49 percent scored at a level 2) of the students identified some assumptions, although they may have demonstrated more awareness of others' assumptions than their own. With 81 percent of the students in the sample in their first and second years, the findings indicate that students have a sound base from which to continue to practice their information literacy skills within their majors.

CRITICAL THINKING FINDINGS

We have two years of results for critical thinking, from 302 student work products (187 in year 1 and 115 in year 2) sampled from fourteen sections of history, music, psychology, and sociology introductory courses.

Although not at all surprising, results from student work scored on critical thinking are the lowest across all learning goals. Within the dimensions of critical thinking, student performances were scored highest on explaining the issues (CT1), with over a third of the students able to clearly describe and clarify the issue to be considered (scores of 3 and 4), and another third of the students able to describe the issue, although with some information omitted (scores of 2). Students had the most difficulty identifying context and assumptions when presenting a position (CT3), and tying conclusions to a range of information, including opposing viewpoints and identifying consequences and implications (CT5).

First- and second-year students accounted for 77.8 percent of the work products scored in this sample. The results indicate that much practice is needed for these students in subsequent years. It also suggests that general education courses, in addition to courses in the majors, will likely need to stress critical thinking more in order for future graduates to attain capstone, or level 4, scores in critical thinking before graduation.

It is important to mention that we made a small change to the critical thinking rubric between years 1 and 2. According to feedback we received from faculty scorers after the first round of using the VALUE critical thinking rubric, the second dimension, Evidence, was difficult to apply. This dimension contains two statements, one addressing the level of interpretation and development of analysis, and the other focused on questioning the viewpoints of experts. Based on this feedback, we piloted a change to the rubric in which the two statements were applied independently. When we did this, the scores on the first

TABLE 1. INFORMATION LITERACY RESULTS

| DIMENSION | PERCENT OF WORK PRODUCTS SCORED 2 OR HIGHER | PERCENT OF WORK PRODUCTS SCORED 3 OR HIGHER |
|--|---|---|
| IL1 Determine Information Needed | 87.2% | 46.2% |
| IL2 Access Needed Information | 89.6% | 46.8% |
| IL3 Evaluate Information and Sources | 88.5% | 39.7% |
| IL4 Use Information Effectively | 85.9% | 43.6% |
| IL5 Access and Use Information Ethically | 93.6% | 59.0% |



part, interpreting the evidence and developing an analysis (CT2.1), are the highest of all dimensions, and the scores on the second part, questioning the viewpoints of the experts (CT 2.2), are the lowest of all dimensions. The information found from dissecting the dimension is quite important, as it suggests that students need to be instructed on the importance of including an author's viewpoint in critical analysis.

RELIABILITY OF THE SCORES

While the reliability of the scores used in our analysis is vital, a complete discussion of this topic is beyond the scope of this article. As already mentioned, reliability starts with norming the scorers, helping them come to a common understanding of the rubric. Measuring interrater reliability (IRR) tells us how well our norming process is working. We compute reliability statistics that are based on a subsample of work products that are scored by multiple scorers, usually between 20 and 30 percent of the total number of work products. The statistics we report include agreement measures—percent agreement and Krippendorff's alpha—and a consistency measure—Spearman's Rho. Benchmarks were determined based on the work of Krippendorff (2004) and Nunnally (1978).

We have met our benchmarks on some dimensions of the rubrics, but we have much more work to do. While our process was designed to help scorers navigate the difficul-

ties and standardize their application of the rubric, we are using our inter-rater reliability results to enhance our scorer norming procedures. Among other things, work needs to be done to set standards for aligning the dimensions of the metarubric to the requirements of an assignment.

CHALLENGES OVERCOME AND STILL TO BE ADDRESSED

Based on the year of work by a committee convened on general education assessment, we were able to anticipate many challenges as we began our work. The committee recommended a structure for accomplishing the assessment activities and for recommending changes based on those results. Among other structures, funds were allocated for faculty stipends for participating in the norming and scoring activities, and the Learning Assessment Council was charged with making recommendations to the faculty senate committee in charge of the general education curriculum.

Other challenges are arising as we implement general education assessment full scale. Our general education curriculum has grown from 214 to 247 courses with the introduction of our new University Studies curriculum, and this number will continue to increase as we implement the additional components of this new four-year curriculum. Appropriate sampling from the curriculum is key to being able to generalize

our findings to our student body, and planning that sampling over time is an integral part of our work. We are also working on additional avenues for disseminating the findings more directly to the faculty responsible for courses in the general education curriculum. To this end, we are providing a series of workshops through the Center for Teaching Excellence in which instructional and assessment experts provide the findings and discuss best practices in teaching and assessing each learning goal.

Looking ahead, we have begun assessing senior-level courses in the majors using the same set of VALUE rubrics. With these data, we will have information on our students' ability levels in their early years at the university and near graduation. This process will help determine where student skills are growing the most, and which skills may need additional emphasis over time.

We have built an assessment process around the skills that will matter most in the twenty-first century, one in which faculty participation and feedback is central. We will continue to look closely at our work, refining both the process and the rubrics based on evidence, in order to make sure our assessment results most closely reflect what our students know and can do. And we look forward to comparing our results to those of other universities. ■

REFERENCES

- Houghton, J., and P. Sheehan. 2000. *A Primer on the Knowledge Economy*. Victoria, Australia: Victoria University, Centre for Strategic Economic Studies Working Paper No. 18. Retrieved from <http://www.cses.com/documents/Wp18.pdf>.
- Krippendorff, K. 2004. *Content Analysis: An Introduction to Its Methodology*. (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Nunnally, J. C. 1978. *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.
- Rhodes, T. L. ed. 2010. *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*. Washington, DC: Association of American Colleges and Universities.
- Stiles, M. J. 2000. "Developing Tacit and Codified Knowledge and Subject Culture within a Virtual Learning Environment." *International Journal of Electrical Engineering Education* 37, 13–25.

TABLE 2. CRITICAL THINKING RESULTS

| DIMENSION | PERCENT OF WORK PRODUCTS SCORED 2 OR HIGHER | PERCENT OF WORK PRODUCTS SCORED 3 OR HIGHER |
|--|---|---|
| CT1 Explanation of Issues | 68.3% | 35.5% |
| CT2 Evidence Year 1 | 65.0% | 28.2% |
| CT2 Evidence Year 2* | | |
| Interpreting and Analysis | 72.8% | 38.6% |
| Questioning viewpoint | 40.9% | 13.6% |
| CT3 Influence of Context and Assumptions | 48.8% | 21.2% |
| CT4 Student's Position | 54.5% | 24.0% |
| CT5 Conclusions and Related Outcomes | 47.7% | 17.0% |

*In Year 2, CT2 was scored as two separate statements. See discussion below.



Assessing Early Integrative Learning

- **Bernice Braid**, director emeritus, University Honors Program, Long Island University Brooklyn
Gladys Palma de Schrynemakers, associate provost, Long Island University Brooklyn
Alan W. Grose, director of academic affairs, The Washington Center for Internships and Academic Seminars

In fall 2003, the Brooklyn Campus of Long Island University implemented a revised core curriculum that created a more collaborative learning environment, one built on a foundational course: Core Seminar (COS), *The Idea of the Human*. This course introduces students early in their academic careers to the methodology of open-ended inquiry in pursuing questions and answers that transcend any particular department or discipline. It not only makes liberal learning a shared activity for students, but also provides faculty with unique interdisciplinary development opportunities and an engaging teaching laboratory. This article traces how AAC&U's integrative learning rubric in particular has helped the campus navigate the intricacies of involving faculty in assessing experiential learning effectively in the larger context of understanding how students think. The course provides faculty an opportunity to shape assignments and design assessments through a portfolio review process. Students also use the integrative learning rubric to frame their own final self-reflection essay.

The mission of this campus is, in part, to “open the doors of the city and the world to men and women of all ethnic and socioeconomic backgrounds who wish to achieve the satisfaction of the educated life and to serve the public good.” The Long Island University student body is 71 percent female and 20 percent male and composed of 39 percent black, 27 percent white, 22 percent Asian and Pacific Island, and 14 percent Latina/o students. One of its greatest strengths is the diversity of its students and faculty, mostly residents of surrounding urban communities. Our core curriculum builds on the multiplicity of languages and cultures students bring with them. It situates them in the context of a unique metropolitan area, one that mirrors their breadth with exceptional cultural resources. The core curriculum review and redesign process has helped contextualize general education by using the city itself as a primary laboratory for learning and for fuller student development.

CORE SEMINAR

The second of three required writing-intensive courses, COS falls between first-year English and a departmentally designated writing/research course in the major. Its objectives include writing with clarity, information literacy, critical analysis, posing new questions, and understanding process. Above all, it aims to provoke students to see themselves in relationship to others and to the world. At the heart of this course is a challenge: to think differently about familiar topics, and to think deeply about unfamiliar, even uncomfortable ones. The course reader provides material from disparate fields, assembled into three sections: social commentary, scientific inquiry, and artistic expression. All selections raise questions about language, insight, and the construction of meaning around an organizing theme, the idea of the human.

Students bring their own distinctive experience into the mix, and because of the seminar format—cross-section joint sessions, small-team deliberation, and explorations off campus—they are encouraged to think critically throughout the course. Students explore one another's neighborhoods to see and hear with fresh eyes; they read philosophical and theoretical works so as to question their world and themselves; and they write so as to sharpen their thinking. Small-team structured field experiences lead them into new sites and settings. Presentations of their findings, often in workshops with students from other sections, hone their listening skills and presentational prowess. In library sessions students explore their own preliminary questions, and learn to tailor them into topics worthy of in-depth consideration. They learn to measure their impressions and their findings, to pose more complex questions, and to refine their own thoughts.

Because students write about the entire process (in journals, response papers, and research projects) and compile their own portfolios, they accumulate a compendium unusually rich in knowledge, insight, and reflection. A final essay, in which they examine their own evolving thought process, is a prime occasion



to consider interconnections of all sorts, including between theories of the human condition and the immediate dramas and demands of life they experience and represent in their classroom dialogue. COS provides undergraduates an intellectual compass that points to how the lives and selves they construct are the outcomes of fashioning a deep, implicit model of life (Brunner 1990, 138), and does so within their first two years of college. Core Seminar engages students to make creative contributions in the classroom by providing learning strategies to help them connect intellectual pursuits with personal queries focused on who they are and who they are becoming. A requirement for all students, this course is also an integrative learning opportunity for faculty. Instructors from all disciplines, from theoretical to practical domains, may teach it. As Carol Geary Schneider noted (2008), when faculty move toward “engaging students in the implications of knowledge,” they are “pulled toward more integrative designs for learning and the equal interest in getting students out in the field to test these skills against real problems” (3).

COS FACULTY AND THE VALUE RUBRICS

Faculty accustomed to teaching within their own disciplinary frameworks are unused to shifting perspectives to cofacilitate with others an open-ended inquiry into problem-based discourse. COS faculty interested in teaching agreed to participate in a semester-long faculty seminar. Readings from many intellectual viewpoints are the substantive texts for what amounts to a parallel “course” they take. Seminars, focused on how to ask questions, on designing and sequencing assignments, and on cross-disciplinary inquiry into abstract problems, become the ‘training ground’ where faculty experiment with modalities unfamiliar to them and create new communities of discourse

that later function as collaborators once they begin to teach COS.

The COS field explorations students undertake yield alternative insights which challenge their assumptions; they review, analyze, and reflect not only on their own observations but on information derived from informal interviews with neighborhood residents and workers beyond their classroom reading. Creating the assignments that send students out into the world to pursue these “street laboratories” requires careful and intentional conversation, a skill faculty hone during their own extended seminars and workshops.

The use of VALUE rubrics at every stage of this process has sharpened faculty debate, and made more acute the questions they raise about provocative or contradictory readings, including assignments that might generate thoughtful reflections and help students explore alternative ways of thinking. Intensified faculty intention-

ality results in the dynamic development of greater integrative capacities among students, as early results from our use of the integrated learning rubric suggest.

The continuing focus on VALUE rubrics in preparation and planning phases of faculty discourse has begun to develop both a shared language of inquiry and a sense of the architectural elements of assignment design that help produce results the rubrics then measure. An example of this interconnection can be discerned in the box below.

The course we have described is complex. It calls for assessment strategies that measure experiential learning, as well as levels of reading comprehension, clarity of communication, and analytical thinking. Embracing rubrics for written communication, critical thinking, reading, and integrative learning has proven a powerful catalyst for students and faculty alike. Instructors use the rubrics to frame

Assignment for COS that Promotes Integrative Learning— (Biology Faculty)

A popular gambit that my colleague (Nursing Faculty) and I use to promote integrative learning is a field exercise for our last joint session entitled “Power Dynamics and Social Interactions.” The assignment’s ostensible focus is to observe how humans interact with each other and/or with animals in a particular setting and to explore the concept of ‘power’ within the context of those interactions. The ultimate goal, however, is to encourage students to make connections between observations culled from their fieldwork and (1) a series of short readings distributed at the joint session; (2) longer readings and COS class discussions spanning the entire semester; (3) material covered in other courses at LIU or elsewhere; and (4) students’ own personal experiences—including those involving family or friends—related to issues of power and vulnerability within varied social situations. We try to provide the impetus to move students beyond simple and more obvious direct association to ones of greater complexity and sophistication.

Students must choose one of the following venues and observe an interaction between two or more individuals for a period of three to fifteen minutes: park or playground; restaurant, department store, or shopping mall; subway, bus, or other public transportation; dog run; or any local zoo. Prior to undertaking fieldwork students receive a series of comprehensive questions to which they must submit written responses on the day of the joint session. The questions are wide-ranging and demand factual information, e.g. demographic data like number, age, and gender of participants, as well as more qualitative and impressionistic judgments regarding body language, dominance status, and levels of emotion. In the joint session the combined classes are divided into several smaller groups, each one responsible for choosing a single member’s observations that most closely relate to a short assigned reading (i.e. a selection from Thomas Hobbes’ *Leviathan* or from Steven R. Covey’s *The 7 Habits of Highly Effective People*). Each group appoints a representative who presents key findings to the constituents of the entire assembly, who pose new (unanticipated) questions to the group and compete to make further connections.



planning workshops, in norming sessions with participants from diverse disciplines, and in workshops focused on designing assignments. In all of these sessions faculty have utilized VALUE rubrics to build a vocabulary of critical consensus, which they then translate into critical terms for students to use as guides in peer editing labs and discussions. COS Faculty Surveys indicate that instructors find increased comfort with discursive teaching and greater agility when designing courses that invite students to integrate their learning directly applicable to the disciplinary courses they teach. There is evidence that this rubric process has already resulted in transference of critical approaches in a variety of disciplines emanating from calibration sessions.

Our portfolio review session was a two-day workshop analyzing random samples from the previous year. Nine raters from seven disciplines read fifty-four portfolios—10 percent of those submitted. Participants identified which elements of the course each rubric measured. Faculty planned field activities and joint sessions, sequenced with specific reading and writing assignments. What emerged from these deliberations were faculty creating student centered assignments. Students' work has become much more creative, collaborative, and responsive to the world in which they live. It has begun to reflect a deeper understanding of the "Idea of the Human."

COS students are mostly in their second or third term, admitted with fairly common literacy challenges that require intervention. Therefore, it was particularly surprising to us when we reviewed end-of-term data using the integrated learning rubric to learn that students performed on a higher than anticipated level. On the criterion of transference, the work reviewed showed that nearly 17 percent of students placed into level 3: "Adapts and applies skills, abilities, theories, or methodologies

gained in one situation to new situations." In this student group, 35 percent reached level 2: "Uses skills, abilities, theories, or methodologies gained in one situation in a new situation to contribute to understanding of problems or issues."

On the criterion of students making connections between "relevant experiences and academic learning," 50 percent of the portfolio work reviewed was clustered at level 1: "Compares life experiences and academic knowledge to infer differences, as well as similarities, and acknowledge perspectives other than own".

These rubric results provide COS faculty and other stakeholders with evidence that the goals of the course, which are aligned with the campus's core curriculum goals, are congruent with nationally articulated learning outcomes (Rhodes 2010). The results of this initial assessment are early benchmarks important for all institutions that expect integrative learning to emerge among students' upper-level accomplishments.

These results help us see how students incorporate unmediated experiential learning into deeper transference and comprehension. In fact, students welcome the opportunity to integrate life experience, course work, and texts early in their undergraduate experience, if invited to do so. Howard Gardner, speaking about the importance of integrative knowledge in his work *Five Minds for the Future* (2008), believes that "in the 21st century, the most valued mind will be the synthesizing mind: the mind that can survey a wide range of sources, decide what is important and worth paying attention to, and then put this information together in ways that make sense to [itself] and, ultimately, to others as well" (18).

The significance of integrative learning is that it reflects multidimensional understanding. The use of the integrated learning rubric permits us to measure student growth from experiential learning.

Until now, however, we have not had an instrument to help us effectively measure adaptation to new situations, transference to solve problems, and insight into perspectives other than their own; the VALUE integrated learning rubric is such an instrument.

THE VALUE INTEGRATIVE LEARNING RUBRIC AND MAKING CONNECTIONS

For some time now faculty have been using field explorations and cross-disciplinary discussions about which students have written reflective essays. Their writing has clearly indicated how many connections they are capable of making. The VALUE integrative learning rubric seems to provoke in students a higher degree of consciousness of their own learning.

The fact that almost one thousand students now take this course each year, in an undergraduate population fewer than six thousand, means that there is a growing community of discourse among students themselves that is a powerful incentive for involvement in experiential and cross-disciplinary inquiry. Our use in particular of the VALUE integrative learning rubric has produced a faculty conscious of the parameters of open inquiry, whose deliberately designed assignments have resulted in students' synthesizing disparate material relatively early in their undergraduate work. ■

REFERENCES

- Bruner, J. 1990. *Acts of Meaning*. Cambridge, Massachusetts: Harvard University Press.
- Gardner, H. 2008. "The Five Minds for the Future." *Schools: Studies in Education* 5 (12): 17–24.
- Rhodes, T. L., ed. 2010. *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*. Washington, DC: Association of American Colleges and Universities.
- Schneider, C. G. 2008. "From the President." *Peer Review* 10 (4): 3.



Fostering Integrative Knowledge and Lifelong Learning

- **Melissa R. Peet**, academic director for the Integrative Learning and MPortfolio Initiative, University of Michigan–Ann Arbor
Laura Reynolds-Keefer, associate professor of education, University of Michigan–Dearborn
Patricia Gurin, professor, Nancy Cantor Distinguished University Professor Emerita, University of Michigan–Ann Arbor
Steve Lonn, research specialist, University of Michigan–Ann Arbor

Like many other institutions, we have been working to develop new ways to educate our students to be flexible lifelong learners and leaders capable of responding to a world of constant change. We want to provide a foundation for intentional, critically engaged lifelong learners, people who can identify what they are learning and understand why it is relevant to their lives. We recognize that in order to learn for life, students will need to know how to consciously learn from their life experiences. They must learn how to pay attention to subtle “a-ha” moments, recognizing the insights and dissonance that often accompanies new learning. They will need to know how to work effectively within diverse teams and groups, balancing the needs and views of others while also staying engaged with their own intentions and sources of curiosity. To do this, they will need to be able to reflect critically on their decisions and actions, recognize the strengths and limitations of their own and others’ perspectives, and continually seek feedback from others and the environment. This seems to be a tall order.

Although the mandate for integrative and lifelong learning extends to nearly all fields and professions, we have found very little is actually known about how to facilitate this type of learning. The literature is not yet clear, for example, on how the term “integrative” applies to different types of learning environments. Reward systems for faculty and university staff, still firmly rooted within disciplinary and institutional silos, make it difficult for them to work across their differences to create more integrative experiences. Moreover, one of the biggest challenges to educating integrative and lifelong learners is the fact that much of the knowledge, skills, and capacities people gain through informal learning is tacit and therefore unconscious and invisible. Research shows that the more a person becomes competent or expert in a given task or area, the more likely it is that her knowledge will recede into a

tacit or unconscious realm (Polyani 1966). Although tacit knowledge is directly linked to the development of effective leaders, experts, innovators, and reflective practitioners, this type of knowledge is largely ignored in most higher education curricula (Sternberg and Horvath 1999).

Research conducted with University of Michigan (UM) students in 2005–2006 demonstrated the need to address these gaps. Focus groups with UM undergraduate student leaders showed that although most reported having “extraordinary” learning experiences at UM, they could not describe what they had learned, why it was valuable to them, or how they might apply the knowledge and/or skills they had gained once they left the university (Leskes and Miller 2006). Although these students felt quite positive about their UM experiences, they could neither articulate nor share what they had learned with others. In response to these and other gaps, we have spent several years developing, testing, and validating several methods for fostering integrative and lifelong learning through the use of integrative teaching methods supported by e-portfolios. To evaluate these methods, we have also created a multidimensional assessment instrument using criteria from a number of the VALUE rubrics, which were developed by the Association of American Colleges and Universities (AAC&U).

THE MPORTFOLIOS PROJECT

This work has been part of the larger campus-wide Mportfolio Project, which was formally established in 2008 as a way to coordinate various integrative learning and e-portfolio efforts across several UM campuses. As a joint effort between the University of Michigan–Dearborn School of Education and the division of student affairs and UM Libraries in Ann Arbor, the Mportfolio project seeks to create pedagogy, technology, and assessment methods that support students



in knowing and demonstrating what they are learning within both formal and informal learning environments, applying their knowledge to the world around them, and integrating learning with their own passions, goals, and sources of curiosity. Mportfolio is now imbedded in an ever widening variety of diverse learning environments on both the Dearborn and Ann Arbor campuses including schools, departments, and co-curricular programs serving diverse undergraduate, professional, and graduate students in fields including business, social work, health care, honors undergraduate education programs, and engineering.

Two integrative learning methods developed on our campuses thus far are referred to as The Integrative Knowledge Portfolio Process and Generative Knowledge Interviewing (Peet et. al 2011). These methods consist of a series of learning modules, interactive exercises, and reflective writing prompts that help students to integrate their learning within the context of a larger institutional culture that strongly emphasizes peer-to-peer learning, leadership development, and deep engagement with the world beyond the classroom. These methods support students to work together in learning how to

- identify, retrieve, and document their tacit knowledge—the unconscious insights, frames of reference, and intuitions they’ve gained from key learning experiences beyond the classroom;

- connect their tacit knowledge to explicit knowledge—the theories, methods, and concepts they are learning in their academic courses;
- integrate their academic learning with their personal passions, goals, and sources of curiosity;
- identify how they can use their knowledge, skills, and capacities to address problems of the world around them; and,
- create an integrative knowledge MPortfolio—a dynamic illustration of how their knowledge, skills, purpose, and goals are continually expressed and re-created in everyday life (for an example, visit <http://tinyurl.com/zellersPortfolio>).

USING THE VALUE INTEGRATIVE LEARNING RUBRIC

The AAC&U VALUE rubrics have played an essential role in our efforts to assess the multiple dimensions of integrative and lifelong learning that are facilitated through these methods. In 2008, several members of the UM Mportfolio research team participated in the AAC&U VALUE initiative. Our role was to test, validate, and provide feedback on the initial integrative learning rubric created by other campus teams. We applied the integrative learning rubric to students’ integrative knowledge e-portfolios to assess how well the rubric’s criteria fit their work.

The process of using the integrated learning rubric proved extremely valuable. It helped us recognize and articulate the tacit

(unspoken) criteria and assumptions we were using in facilitating the Integrative Knowledge Portfolio Process, allowing us to clarify how, why, and when we were prompting students to connect and synthesize their learning. We immediately recognized how our interpretation of integrative learning was both similar to, and yet different from, that of other campuses. Like others, we were using “integrative” to mean helping our students to synthesize learning from various experiences and understand the relevance and potential application of that learning to the world beyond the classroom.

The areas where our students’ portfolios did not align with the integrated learning rubric also proved valuable. For instance, our students’ portfolios demonstrated several areas of learning that were not represented in the initial integrative learning rubric (e.g. that they were able to recognize tacit knowledge and seek out and synthesize feedback from peers and others), and revealed that these differences were representative of the culture of our institution. This inspired us to create an assessment instrument that would capture the unique dimensions of integrative and lifelong learning emerging on our campuses. To do this, we turned to several other VALUE rubrics, particularly those focusing on civic engagement, creative thinking, team learning, and lifelong learning (see Rhodes 2010).

The result was an instrument that combines UM’s unique definition of integrative knowledge and lifelong learning and empha-

FIGURE 1. DIFFERENCES IN PRE/POST COMPOSITE SCORES FOR SIX FACTORS FOR INTEGRATIVE LEARNING

| MEASURE | N | PRE-SURVEY | | POST-SURVEY | |
|--|-----|------------|--------------------|-------------|--------------------|
| | | Mean | Standard Deviation | Mean | Standard Deviation |
| Demonstrate knowledge gained within and across specific contexts | 620 | 3.88 | 0.67 | 4.26* | 0.56 |
| Recognize and adapt to differences | 620 | 4.42 | 0.45 | 4.49* | 0.49 |
| Understand and direct oneself as a learner | 620 | 4.25 | 0.48 | 4.42* | 0.47 |
| Become a reflexive, accountable and relational learner | 607 | 4.10 | 0.53 | 4.31* | 0.52 |
| Identify and discern my own and others’ ethics and perspectives | 620 | 4.30 | 0.50 | 4.45* | 0.53 |
| Develop a professional digital identity | 609 | 3.49 | 0.86 | 4.09* | 0.78 |

Note: * = $p < .001$



sizes students' developing the capacities to reflect critically, tacitly share knowledge, give and receive feedback, and work with others for social change (Peet et. al 2011, 26–28). It uses language from at least four of the AAC&U VALUE rubrics. The instrument is a thirty-seven item pre/post self-assessment survey that addresses students'

- recognition of their strengths and challenges as learners;
- identification of their values and beliefs;
- understanding of their social identities and perspectives;
- skills in working across social/cultural differences;
- awareness of how one gains different types of knowledge;
- adaption of knowledge/skills to new contexts;
- evaluation of their work; ability to listen and seek feedback;
- recognition of one's own passions and sources of curiosity;
- development of a professional identity;
- ability to work with others to make a difference;
- understanding of how one's actions/decisions affect others.

This instrument was pilot tested during the 2009–2010 and 2010–2011 academic years (usually at the beginning and end of a term in courses/programs) in learning environments using the Integrative Knowledge Portfolio Process and/or Generative Knowledge Interviewing. Most recently, the data generated from this instrument demonstrated significant gains on six dimensions of integrative learning for 620 diverse students (including both traditional and nontraditional) from fourteen different academic and cocurricular learning environments across both the Dearborn and Ann Arbor campuses (for a complete description of this research see Peet et. al. 2011). These results validated our conceptual model of integrative knowledge and lifelong learning (see fig. 1), which includes students learning how to

1. Identify, demonstrate, and adapt knowledge gained within/across different contexts—the ability to recognize the tacit and explicit knowledge gained in specific learning experiences and the capacity to adapt that knowledge to new situations;
2. Adapt to differences in order to create solutions—the ability to identify and adapt to different people, situations, etc., while working with others to create positive change;
3. Understand and direct oneself as a learner—the ability to identify one's prior knowledge, recognize one's strengths and gaps as a learner, and know how one is motivated to learn;
4. Become a reflexive, accountable, and relational learner—the ability to reflect on one's practices and clarify expectations within oneself while also seeking feedback from others;
5. Identify and discern one's own and others' perspectives—the ability to recognize the limitations of one's perspective and seek out and value the perspectives of others;
6. Develop a professional digital identity—the ability to imagine how one will use current knowledge and skills in future roles and how one will create an intentional digital identity.

As is shown in figure 1, the gains were statistically significant for all six dimensions of the integrative knowledge and lifelong learning conceptual model. (Peet et al. 2011).

Given the fact that there is very little literature to date that fully operationalizes “integrative” or “lifelong” learning, or explicitly connects how different types of integrative teaching methods lead to the development of particular types of integrative capacities in students, we hope that this work will be useful to others.

Moving ahead, the instrument continues to be updated though feedback from UM educators, as well as educators

who are using the instrument at other institutions (e.g., Norwalk Community College and Long Island University). In 2010, UM was awarded a FIPSE grant to support the Lifelong Learning Curriculum Transformation Project, the purpose of which is to build a collaboration between UM–Dearborn and UM–Ann Arbor, as well as six other innovative institutions—Boston University, Clemson University, DePaul University, Mercy College (New York City), Oberlin College, and Portland State University. These institutions are working together to create and identify best practices and research that supports integrative and lifelong learning across a wide range of fields, disciplines, and professions in higher education. The AAC&U VALUE rubrics have provided a solid foundation for the University of Michigan campuses and their partner institutions to begin to think about ways to accurately assess student learning processes and changing capacities. The collaborative, with the continuing guidance of AAC&U will continue to explore how higher education can assist learners in becoming integrative lifelong learners, and how that learning can best be assessed. ■

REFERENCES

- Leskes, A., and R. Miller. 2006. *Purposeful Pathways: Helping Students Achieve Key Learning Outcomes*. Washington, DC: Association of American Colleges and Universities.
- Peet, M. 2010. “The Integrative Knowledge Portfolio Process: A Program Guide for Educating Reflective Practitioners and Lifelong Learners.” On-line journal *MedEdportal* <https://www.mededportal.org/publication/7892> (free access through site registration).
- Peet, M., S. Lonn, P. Gurin, M. Matney, T. Marra, S. Himbeault-Taylor, A. Daley. 2011. “Fostering Integrative Knowledge Through ePortfolios.” *International Journal of ePortfolios*, 1(1):11–31. <http://www.theiejep.com/current.cfm>.
- Polanyi, M. 1966. *The Tacit Dimension*. Chicago: University of Chicago Press.
- Rhodes, T. L., ed. 2010. *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*. Washington, DC: Association of American Colleges and Universities.
- Sternberg, R., and J. Horvath, eds. 1999. *Tacit Knowledge in Professional Practice*. Mahwah, NJ: Lawrence Erlbaum Associates.



Staying On Track With Rubric Assessment: Five Institutions Investigate Information Literacy Learning

► **Megan Oakleaf**, assistant professor, Syracuse University

Like many AAC&U essential learning outcomes, information literacy—the ability to locate, evaluate, and use information—is a concept integrated into curricula and cocurricula campuswide. Throughout college, students are asked to define their information needs in order to complete their academic work and make decisions affecting their personal and social development. They are expected to obtain information to meet those needs, evaluate the applicability of the information they locate, and revise their searches accordingly. They also are required to apply new information ethically and responsibly. These information literacy outcomes are integral to a successful college experience.

Outcomes with such broad applicability should be assessed often and well. In reality, information literacy is frequently omitted from higher education assessment efforts for several reasons. First, faculty and cocurricular professionals traditionally expect students to acquire information literacy skills prior to college; consequently, they typically do not teach and assess them as a focus of their courses and activities. Second, librarians—who do focus on teaching information literacy—often do not have the access to students in ways or settings most conducive to meaningful assessment activities. Third, because information literacy is often not taught in courses or cocurricular activities, existing information literacy assessments are frequently limited to survey and test formats that can be administered by librarians remotely. However, information literacy is an especially complex and context-dependent concept that is not easily assessed using these common fixed-choice methods.

THE INFORMATION LITERACY VALUE RUBRIC

The development of an information literacy VALUE (Valid Assessment of Learning in Undergraduate Education) rubric was a significant advance in information literacy assessment because

it addressed these obstacles to information literacy assessment. First, naming information literacy an “essential” learning outcome, the ability to locate, evaluate, and use information was elevated to a position valued by faculty, cocurricular professionals, and, of course, librarians. Second, the information literacy VALUE rubric established an expectation that this outcome will be assessed in the context of existing student learning activities and assignments—activities and assignments that are both complex in nature and steeped in authentic contexts.

RAILS

Recently, the Institute of Museum and Library Services awarded \$400,000+ in funding to support a three-year project called RAILS (Rubric Assessment of Information Literacy Skills). The RAILS project is designed to support the rubric assessment of information literacy outcomes at institutions nationwide. During the 2010–2011 academic year, five institutions participated in RAILS: a branch campus of a public university (2,800 FTE); a private, faith-based liberal-arts college (4,500 FTE); a private, liberal-arts college (6,400 FTE); a public, land-grant, high-research activity university (29,000 FTE); and a public college that focuses on workforce development and offers high-school completion, certificates, and associates degrees (30,000 FTE). To begin, librarians from each institution took part in intensive rubric training. As a part of their training, librarians learned to customize the information literacy VALUE rubric to fit the unique needs of their institutions and formed plans to test their rubrics. Using the rubric, they gathered 100+ artifacts of student learning to assess and selected ten participants (faculty, cocurricular professionals, and other librarians) to serve as raters. Intensive rubric revision, norming, and scoring sessions for all raters were then scheduled. During the scoring sessions, each of the ten participants rated all one hundred student



artifacts and input rubric scores for each student into an assessment management system that facilitates rubric usage. These scores are currently under analysis in order to learn about student information literacy skill levels and explore factors that contribute to inter-rater reliability. Results will be disseminated as they become available via a variety of venues, including the project website at www.railsontrack.info. During the 2011–12 academic year, this cycle will be repeated at five additional institutions. Combined assessment results from all ten RAILS institutions will be released in summer 2012.

Student Performance on RAILS Rubrics

In order to meet the needs of individual institutions, each rubric used during the first year of RAILS was slightly different in scope (see figures 1–3 below for examples). Each RAILS rubric was based on the information literacy VALUE rubric; however, individual institutions decided to “break out” different criteria and describe them in more specific terms. Figure 1 demonstrates how one VALUE criterion (“Use Information Effectively to Accomplish a Specific Purpose”) can be divided into three parts using the key verbs from the VALUE rubric: communicate, organize, and synthesize.

The first year of applying RAILS rubrics also revealed varying levels of student skills. For some institutions, student performance on the rubrics was “as expected.” Other institutions found more surprising results; in some cases, the surprise was how many students fell into unsatisfactory performance levels in specific skill areas. For example, one institution found that students in the RAILS sample did not achieve expected levels of ethical and legal information use (see figure 2). At this institution, almost 80 percent of students either did not follow style guide conventions when citing information sources (13 percent) or committed frequent errors (65 percent). These results provided the concrete evidence needed to significantly revise instruction to address skill weaknesses. At other institutions, interesting patterns of satisfactory performance emerged. For instance, one institution found that students were far more likely to evaluate an information source at an “accomplished” level based on currency (68 percent) or authorship (46 percent) than the source’s reliability (23 percent), accuracy (21 percent), or point of view (27 percent) (see figure 3). After receiving these results, faculty and librarians developed a shared understanding of information evaluation criteria, adjusted

assignment requirements, and improved in-class instruction.

Potential Barriers to Rubric Assessment

After rating student work using RAILS rubrics, participants were asked to identify major barriers to their personal use of rubrics as well as their perceptions of potential barriers for their colleagues. More than half listed a lack of time and coordinated structures for assessment (e.g., an assessment “point person” or committee) as major barriers to their use of rubrics. More than a quarter of RAILS participants cited insufficient institutional financial resources, lack of staff, and uncertainty about their role in assessment. There were also concerns about possible inaccuracy of assessment tools and misuse of assessment data. In considering their colleagues, more than a third of participants stated that their colleagues would be discouraged by these barriers plus two more: their own lack of familiarity with rubric assessment in general and a lack of rewards for participating in assessment activities.

Initial RAILS Findings

Barriers notwithstanding, institutions have already reported a number of important findings resulting from RAILS participation including changes to teaching, assessment, and collaborative activities.

FIGURE 1. STUDENT PERFORMANCE ON RAILS RUBRIC “USE INFORMATION EFFECTIVELY TO ACCOMPLISH A SPECIFIC PURPOSE”

| CRITERION | 3 | 2 | 1 |
|---|---|---|--|
| Organizes Content Are the sources in the right places? | Consistently organizes cited information in a manner that supports the purposes and format of the product/performance. Students rated as 3: 35% | Inconsistently organizes cited information in a manner that supports the purposes and format of the product/performance. Students rated as 2: 45% | Does not organize cited information in a manner that supports the purposes and format of the product/performance. Students rated as 1: 20% |
| Synthesizes New and Prior Information Do the sources help to support new claims or make points? | Consistently connects new and prior information to create a product/performance. Students rated as 3: 27% | Inconsistently connects new and prior information to create a product/performance. Students rated as 2: 48% | Does not connect new and prior knowledge to create a product/performance. Students rated as 1: 25% |
| Communicates Information Do they have sources? | Consistently communicates information from sources via products/performances. Students rated as 3: 37% | Inconsistently communicates information from sources via products/performances. Students rated as 2: 50% | Does not communicate information from sources via products/performances. Students rated as 1: 13% |



Teaching

All institutions report substantive improvements in teaching information literacy concepts. Furthermore, librarians and faculty have revised student assignments, altered classroom pedagogy, and enacted changes in curriculum sequencing. Faculty reported a higher quality of student work and plan to use rubrics as one of their teaching tools. One RAILS participant reflected, “I learned that grading the assignments in the RAILS project was an empowering act for me. It will strengthen my teaching the next time because I now understand what the students really are not getting. This rubric creation and rating experience has facilitated valuable reflection on my teaching practice and I hope to weave what I now understand into my teaching the next time around.” According to another participant, RAILS “changed the way I teach... [the instructional] session has more structure, and the students seemed much more engaged.” A third participant shared a student remark about the increased level of hands-on engagement

that resulted from pedagogical changes: “The day that we went as a class to the library... was probably one of the most beneficial days of my semester.”

Assessment

In addition to teaching improvements, all institutions also documented increased assessment activity. Several institutions created or revised rubrics focused on information literacy and other essential learning outcomes. Faculty and librarians collected more artifacts of student learning for assessment purposes, and one institution regularized consent forms for collecting student work samples. In addition, more than a few institutions are contemplating the purchase of assessment management systems to improve the organization and reporting of their assessment data. Significantly, the librarians who participated in RAILS feel less disconnected from institutional assessment efforts and more committed to participating in program reviews and accreditation processes campuswide.

Collaboration

Beyond teaching and assessment improvements, rubric assessment has enhanced cross-campus collaboration. Faculty, teaching and learning center professionals, librarians, and writing center professionals have reported increased partnerships and future plans for co-created projects. A few have commenced rubric-related research projects; others have joined together to communicate rubric assessment results.

PRELIMINARY RESULTS

Although the statistical analysis of first year RAILS results is still in process, a few preliminary findings have emerged. First, faculty, cocurricular professionals, and librarians need to increase their awareness and knowledge of rubric assessment. Many of the participants involved in the RAILS project had no prior experience with rubrics, but all expressed interest in learning more about them. Second, if multiple participants plan to use the same rubric to score artifacts of student learning, norming is critical for establishing shared understanding of the rubric and achieving

FIGURE 2. STUDENT PERFORMANCE ON RAILS RUBRIC “USE INFORMATION ETHICALLY AND LEGALLY”

| CRITERION | ADVANCED | DEVELOPING | BEGINNING |
|--|---|--|---|
| | Applies Outcome Successfully; Many Strengths Are Present | Shows Skill in this Outcome; Improvement Needed | Evidence of the Outcome May Be Minimally or Not at All Present; Need for Improvement Outweighs Apparent Strengths |
| Style conventions | Follows style guide conventions with few errors. Students rated as Advanced: 22% | Follows style guide conventions with frequent errors. Students rated as Developing: 65% | Does not follow style guide conventions. Students rated as Beginning: 13% |
| Correspondence of bibliography and in-text citations | Bibliography and in-text citations correspond. Students rated as Advanced: 39% | Bibliography and in-text citations do not correspond. Students rated as Developing: 53% | Does not include a functional bibliography and/or in-text citations. Students rated as Beginning: 8% |
| Common knowledge and attribution of ideas | Consistently distinguishes between common knowledge and ideas requiring attribution. Students rated as Advanced: 33% | Inconsistently distinguishes between common knowledge and ideas requiring attribution. Students rated as Developing: 59% | Does not distinguish between common knowledge and ideas requiring attribution. Students rated as Beginning: 8% |
| Paraphrasing, summarizing, quoting | Summarizes, paraphrases, or quotes in order to integrate the work of others into their own. Students rated as Advanced: 43% | Summarizes, paraphrases, or quotes, but does not always select appropriate method for integrating the work of others into their own. Students rated as Developing: 53% | Does not summarize, paraphrase, or quote in order to integrate the work of others into their own. Students rated as Beginning: 4% |



greater inter-rater reliability. Third, analytical rubrics appear to be more practical for assessing student artifacts than holistic rubrics. The more specific the language and the narrower the scope of the rubric, the more confident participants seemed to be about the accuracy of their ratings. Fourth, the more analytical the rubric becomes, the more directed the rubric scorer must be in looking for evidence of learning. Thus, participants appeared to be more confident about their ratings when student artifacts under analysis were concrete, focused, and shorter in length. This may mean targeting a different type of student artifact or subdividing larger artifacts to facilitate scoring. Fifth, large scale analysis of rubric assessment results

is faster and more convenient when an appropriate assessment management system is a part of the process. The use of the assessment management system greatly facilitated the data recording, analysis, and reporting of RAILS rubric data.

Once statistical analysis of RAILS has been completed, more conclusions can be made. During the 2011–12 year, five additional institutions will participate in the RAILS process and benefit from the lessons learned from the first year of the grant.

CONCLUSION

The information literacy VALUE rubric in its original, holistic form is a significant step forward in the assessment of

students' ability to locate, evaluate, and use information, both in and out of the classroom. The RAILS project has already captured important information about student information literacy skills at several institutions and identified potential barriers to rubric assessment so they may be better understood and overcome. Perhaps most importantly, RAILS has demonstrated that VALUE rubrics, adapted for analytical, campus-specific purposes, can spur instructional improvements, increase assessment activity, and improve collaborations among faculty, cocurricular professionals, and librarians. In the coming year, RAILS will build on these successes to keep rubric assessment of information literacy "on track"! ■

FIGURE 3. STUDENT PERFORMANCE ON RAILS RUBRIC "EVALUATES INFORMATION AND ITS SOURCES CRITICALLY AND ACCESS THE NEEDED INFORMATION"

| CRITERION | ACCOMPLISHED | DEVELOPING | INADEQUATE |
|---------------------------------------|--|--|---|
| Evaluates authority | Student shows sufficient evidence of the author's credentials and qualifications. Students rated as Accomplished: 46% | Student briefly identifies the author's credentials and qualifications. Students rated as Developing: 35% | Student does not identify the author's credentials or qualifications. Students rated as Inadequate: 19% |
| Evaluates currency | Student comments on the source's publication year and retrieves the source that is published within the last five years. Students rated as Accomplished: 68% | Student either comments on the source's publication year or retrieves a source that is published in the last five years, but does not do both. Students rated as Developing: 26% | Student does not comment on the source's publication year and does not retrieve a source that is published in the last five years. Students rated as Inadequate: 6% |
| Evaluates reliability | Student shows adequate evidence of whether or not the source is trustworthy. Students rated as Accomplished: 23% | Student shows superficial evidence of whether or not the source is trustworthy. Students rated as Developing: 53% | Student does not show evidence of whether or not the source is trustworthy. Students rated as Inadequate: 24% |
| Evaluates accuracy | Student provides a thorough explanation of the accuracy of the source. Students rated as Accomplished: 21% | Student provides superficial explanation of the accuracy of the source. Students rated as Developing: 51% | Student does not explain the accuracy of the source. Students rated as Inadequate: 28% |
| Evaluates perspective | Student identifies the author's point of view in detail. Students rated as Accomplished: 27% | Student briefly identifies the author's point of view. Students rated as Developing: 53% | Student does not identify the author's point of view. Students rated as Inadequate: 20% |
| Evaluates reflection of source | Student explains in detail how the source contributes to his/her knowledge. Students rated as Accomplished: 29% | Student identifies how the source contributes to his/her knowledge. Students rated as Developing: 51% | Student does not identify how the source contributes to his/her knowledge. Students rated as Inadequate: 20% |
| Access the needed information | Student accesses information using effective, well-designed search strategies. Students rated as Accomplished: 27% | Student accesses information using simple strategies, including both search term(s) and tool(s). Students rated as Developing: 53% | Student does not specify strategy with both search term(s) and tool(s). Students rated as Inadequate: 20% |



Using VALUE Rubrics to Evaluate Collaborative Course Design

► **Andrea Greenhoot**, associate professor of psychology and faculty fellow, Center for Teaching Excellence, University of Kansas
Dan Bernstein, professor of psychology and director, Center for Teaching Excellence, University of Kansas

In recent years, the University of Kansas has made it a priority to improve the written communication and critical thinking skills of our undergraduate students. We especially have embraced the challenge of enhancing these skills in courses with larger enrollments—those with sixty to several hundred students. Although most faculty members presume that high-end student writing and critical analysis can only be properly taught or evaluated when class size is below thirty or forty students, a group of University of Kansas colleagues has redesigned large-enrollment courses around a cognitive apprenticeship model to specifically target these skills. Faculty members working with colleagues in the writing center and libraries developed and implemented staged assignments that allow students to acquire writing and critical thinking skills through multiple iterations with individual feedback.

In early versions of the project, five faculty members worked with a subject librarian and a writing consultant to design sequenced assignments that would develop students' skills in written communication, critical reading, and synthesis of research literature. Later, through a project funded by the Spencer and Teagle Foundations, we expanded to ten additional courses and we developed a graduate student fellowship program in which teaching assistants were prepared as consultants by specialists in the library and in the writing center. We provided supplemental financial support to graduate student fellows so they could give extra time and skill to the courses they are assisting.

We wanted to know whether collaborative course design was worth the investment of time and resources needed to scale up beyond a pilot program. To answer this question, we used

rubrics developed by the Association of American Colleges and Universities (AAC&U) to evaluate the written communication and critical thinking skills of students in these team-designed courses. The VALUE rubrics supplemented existing instructor evaluations of successful learning (i.e., grades on course-specific rubrics), allowing us to measure skills that are neither course nor discipline specific. We also liked the well-articulated goals and benchmarks of performance of the VALUE rubrics because they connect our expectations to those of a community of faculty

The VALUE rubrics supplemented existing instructor evaluations of successful learning (i.e., grades on course-specific rubrics), allowing us to measure skills that are neither course nor discipline specific.

colleagues from around the country. As a third measure, we had samples of students take the Collegiate Learning Assessment (CLA) to evaluate our course design model and to triangulate among three estimates of student learning (grades, VALUE rubrics, and CLA).

THE RATING PROCESS

To date, we have used the VALUE rubrics on written communication and critical thinking to score the written assignments of about one hundred students from the first year of our three-year team-design project. We began by gathering random samples of assignments from the two team-designed courses from the project's first year (political science and psychology) and two



traditional courses of similar size and curricular level in the same disciplines. We convened a three-hour session of four graduate students (the raters) and a few faculty members to iteratively rate and discuss five or six assignments and make minor adjustments to the rubric language until they came to a shared understanding of the rubric categories and criteria. Each assignment was then independently scored on both the written communication and critical thinking rubrics by two different raters. The raters were quite reliable with each other, providing scores that were identical or one category apart at least 90 percent of the time. At the end of this process, the raters met again to discuss scoring disagreements and were permitted, but not compelled, to change their ratings following the discussion.

RESULTS

Once the scoring was complete, our first step was to look for overlap across the individual dimensions of the rubrics as a possible way of simplifying our inspection and representation of student performance. Scores on all five dimensions of the written communication rubric were strongly associated with one another so that students who scored high on one dimension generally scored high on all dimensions. Therefore, we pooled the scores across all of the written communication dimensions in our evaluation of students' skills. In contrast, the critical thinking scores clustered together in two groups: the explanation of issues, student's position, and conclusions/related outcomes dimensions were highly correlated with one another, and the evidence and context/assumptions dimensions were highly correlated, but the two

groups were only weakly related to each other. For this reason, we aggregated the critical thinking scores into two sets, one that we refer to as "Issues, Analysis and Conclusions," and the other that we call "Evaluation of Sources and Evidence."

We represented students' skill attainment by looking at the proportion of ratings that fell into each performance category (i.e., not met, benchmark, milestone 1, milestone 2, and capstone) for students in the team-designed and traditional courses. Figure 1 presents the distributions of scores, generated by all raters, pooled across all five written communication dimensions. This convention allows us to see clearly and quickly the variability of skill among our students and also reveals meaningful differences between the two types of courses. Very few of the scores are at the bottom or top ends of the distribution, but the distribution is shifted to the right (i.e., higher scores) for students in the team-designed courses—and this shift is statistically significant.

We constructed two sets of distributions of critical thinking ratings (see the top chart in fig. 2) representing the Issues, Analysis, and Conclusions scores and the other representing the Evaluation of Sources and Evidence scores (see the bottom chart in fig. 2). One clear take-away point is that regardless of course type, almost none of our students are reaching the highest level in either set of critical thinking skills. Nevertheless, the Issues, Analysis, and Conclusions ratings show a significant advantage for students in the team-designed courses. In the traditional courses, 20 percent of the ratings were below the minimal level of skill attainment. In the redesigned courses, virtually all of the ratings showed benchmark or higher skill attainment, and there were many more scores in the milestone 1 and 2 categories than in the traditional courses.

FIGURE 1. DISTRIBUTIONS OF WRITTEN COMMUNICATION RATINGS

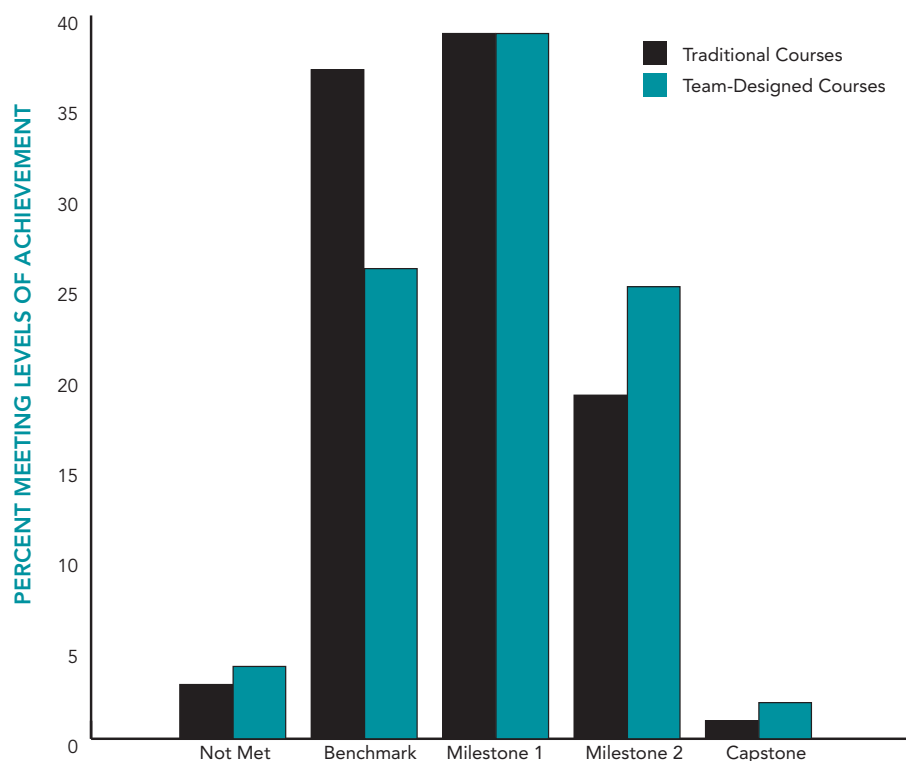




FIGURE 2A. CRITICAL THINKING: ISSUES, ANALYSIS AND CONCLUSIONS

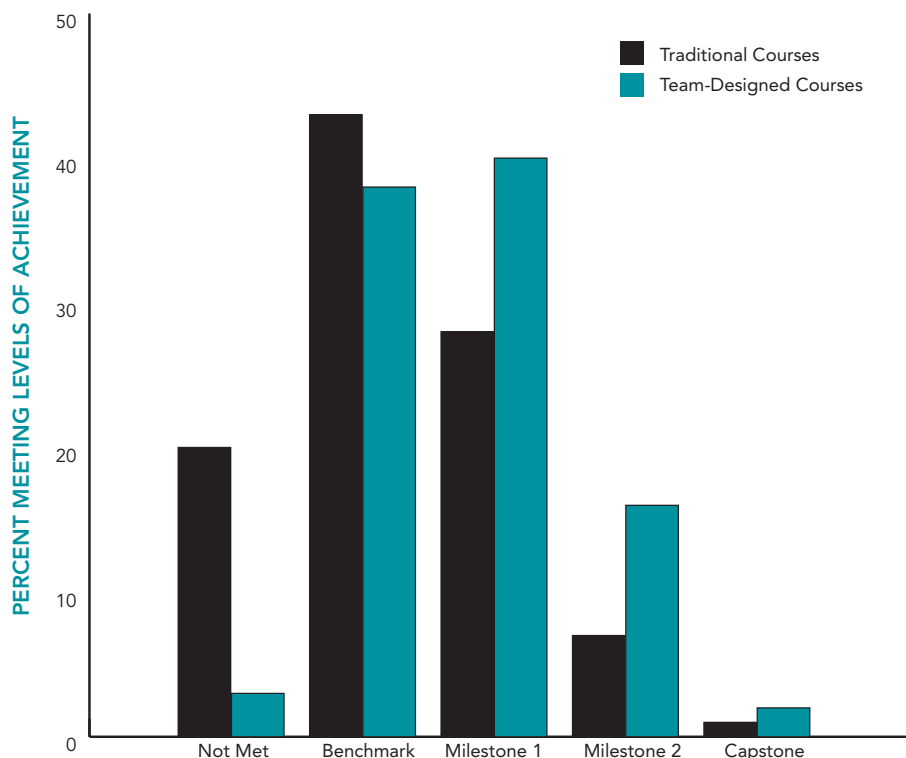
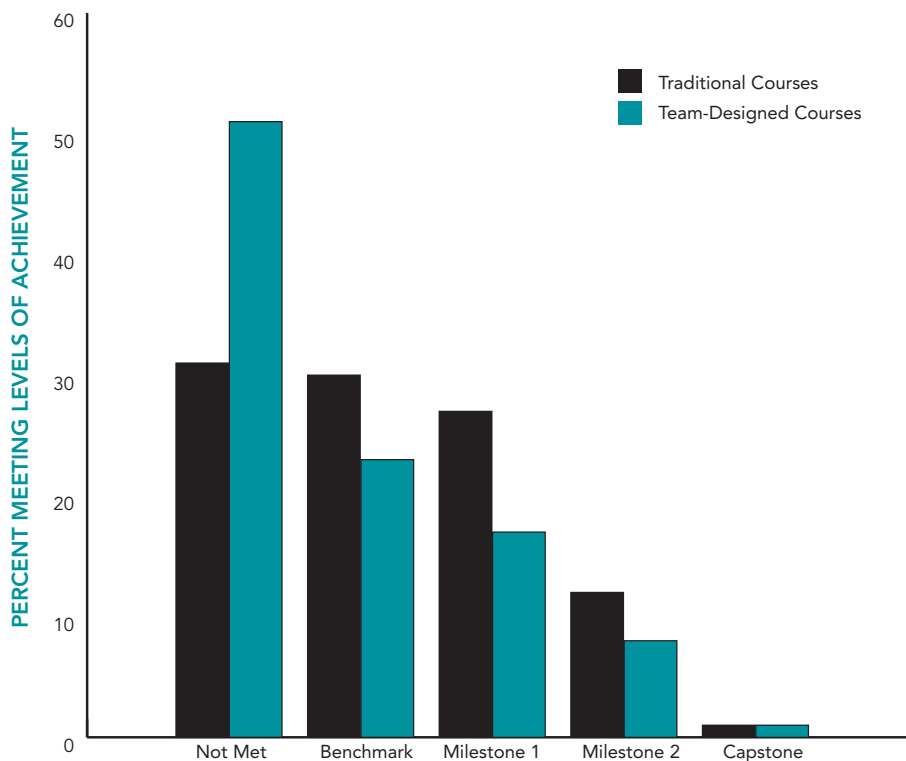


FIGURE 2B. CRITICAL THINKING: EVALUATION OF SOURCES AND EVIDENCE



Compared to the other skills we evaluated, a high proportion of the Evaluation of Sources and Evidence ratings for all classes were below benchmark, suggesting a weakness in our students' skill sets. There was also a surprising pattern: more than half of the ratings of student work in the redesigned courses did not meet the benchmark criterion, compared to about one-third in the traditional courses. We noted that the assignment instructions in both of the traditional courses called on students to evaluate their sources and evidence, whereas in the team-designed courses these skills were taught and already evaluated in an earlier stage of the assignment sequence. It appears that some skills will not be visible in student work unless the assignment explicitly calls upon students to use them. That difference in assignment requirements may also explain why our sample showed two unrelated subsets of dimensions within critical thinking.

We were also interested in how well our evaluations of student work via the VALUE rubrics corresponded with the two other measures of written communication and critical thinking in our project (i.e., grades on course-specific rubrics and CLA scores). Interestingly, the patterns that were visible in the VALUE rubric scores were not mirrored in the CLA scores; students in the team-designed and traditional courses performed no differently on the CLA. Students' performance on the CLA, moreover, was generally unrelated to the VALUE rubric ratings of their coursework, as well as the instructors' grading of the same coursework. In contrast, the latter two measures were highly correlated, suggesting that the VALUE rubrics capture qualities of critical thinking and writing that fit well with what faculty members value in their students' work.



LESSONS LEARNED AND NEXT STEPS

We represent the results of our student learning assessment as distributions of performance across categories, not as arithmetic averages of numerical values assigned to levels of skill. These graphs display the variability among our students and show the magnitude of the challenges we face in increasing student achievement. Although averages would illustrate the same changes associated with our course design intervention, the means mask much valuable information about the actual levels of skill that students are exhibiting. We suggest that graphical distributions of VALUE rubric scores are also an excellent way to track student growth across the first four years of their education. To illustrate, in figure 3 we present hypothetical distributions of

VALUE rubric scores for first-year and fourth-year students in the same graph. Such distributions could represent either cross-sectional comparisons of samples of first- and fourth-year students gathered in the same year, or longitudinal comparisons that represent distributions from a large sample of the same students as they change over their college careers. This is a very transparent and direct form of evidence for any conversation about the value added by higher education. It speaks to the most important form of institutional evaluation: how varied are our students and how are they transformed by their time with us? Institutional self-assessment can be exceptionally well-served by careful tracking of performance distributions on the VALUE rubrics.

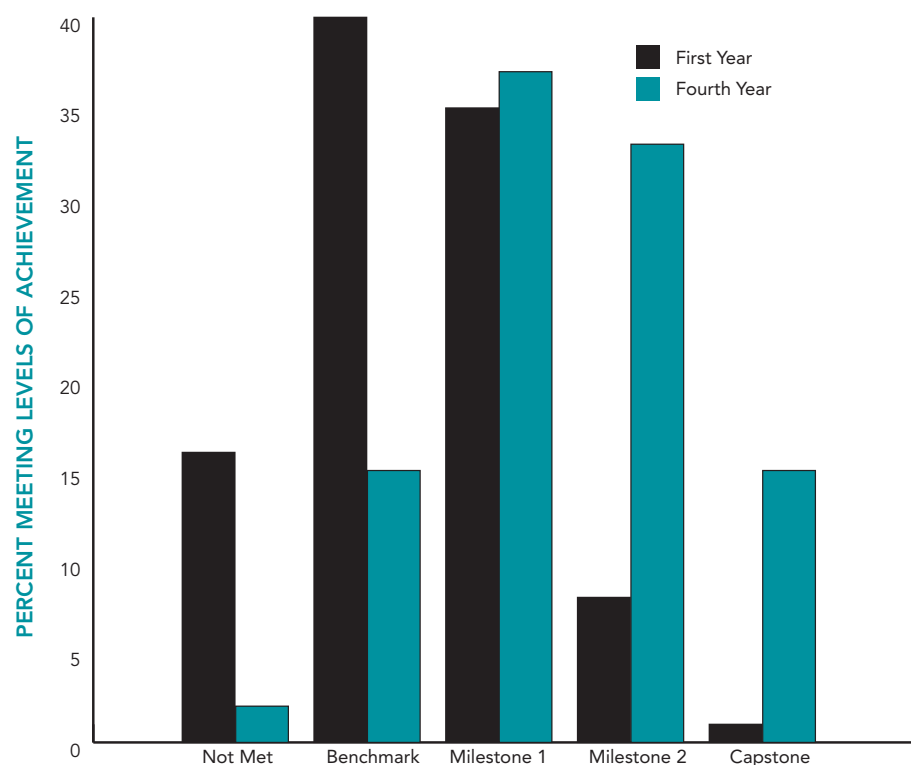
Based on the VALUE rubric data and changes in course-specific measures of

student skill, we will continue to expand the number of courses that take advantage of collaborative course design. It is essential that our analysis includes an independent reading that identifies the level of intellectual work based on criteria that come from outside our university. When we used the AAC&U category definitions, which were generated by two years of national conversation on the rubrics, we confirmed what our local grades suggested about the benefits of collaborative course design. We also learned that relatively few of the students in this initial set of four courses appeared to meet the definitions for the best two rating categories on several dimensions of their work. Because we had expectations and goals derived by a consensus among members of a higher education organization, we were able to learn how well our students met those broader expectations of intellectual work. We look forward to a time when comparable institutions would share their distributions on these rubrics to provide useful benchmarks of performance.

Our analysis of the data from the VALUE rubrics affirmed that a team approach to course design can improve, some forms of student writing and thinking more than traditional solo design. We also demonstrated that graduate teaching assistants could be readily prepared to serve as informed consultants on assignment design, so we have a model for scaling up the entire enterprise to wider range of large enrollment courses, especially if they have teaching assistants. Success with writing center and library consultants suggests that we can expand to include instructional technology and student support specialists as options in building a course design team.

We also saw that the rubrics work best when there is close alignment between the nature of the assignment and the dimensions of intellectual skill described

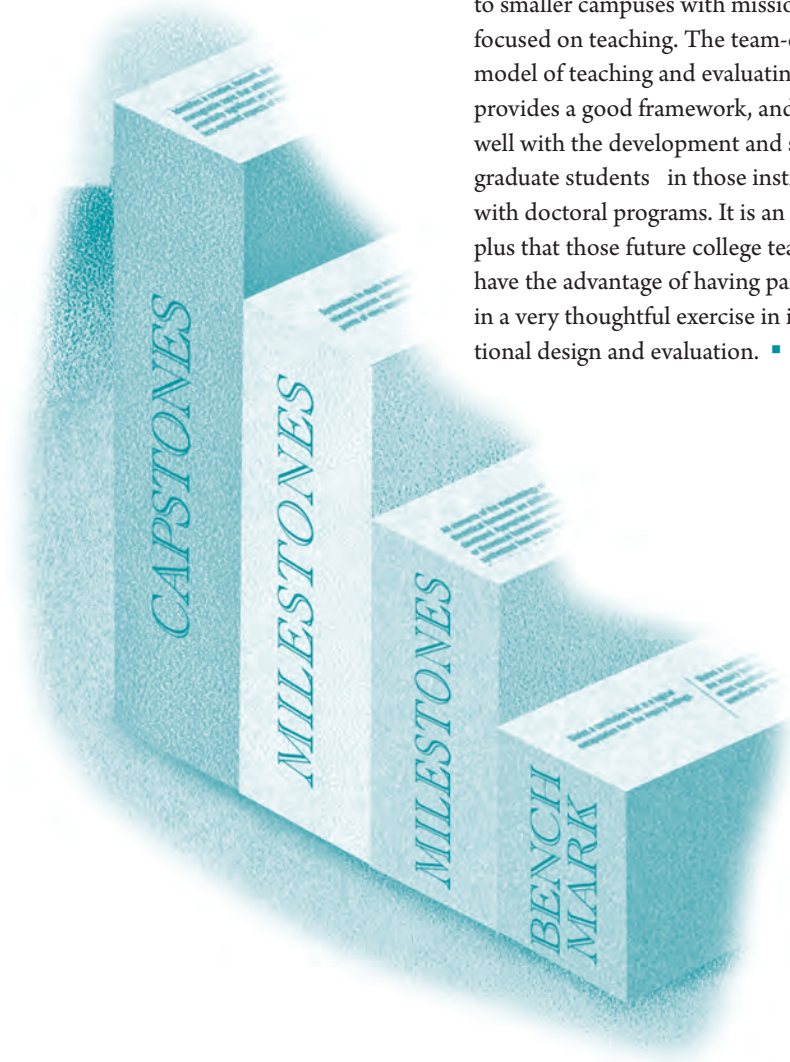
FIGURE 3. HYPOTHETICAL ILLUSTRATION OF “VALUE ADDED” IN FOUR YEARS, USING VALUE RUBRIC RATINGS





in the rubric. One of our surprise results is probably attributable to differences in how assignments were framed for students. One course presumed a level of skill in source evaluation, perhaps acquired from earlier courses, and students were expected to demonstrate it in their final assignment. Another course explicitly taught that skill in the early stage assignments, and the final assignment did not call for students to demonstrate it. In order for an independent reading to be effective as institutional feedback, we need to sample student work carefully, being sure that the assignment instructions are well aligned with the goals used by readers.

Finally, at a practical level, we are very encouraged that this process is manageable and sustainable. Faculty members who have used the VALUE rubrics to rate student work typically find the dimensions and category definitions sensible and meaningfully related to their existing views of educational goals. This affirms the lengthy process AAC&U followed in developing and refining these rubrics. We also believe the development of graduate students as partners in the enterprise is a useful model for campuses that have both teaching assistants and large populations of students. We see no reason that this form of evaluation should be limited only to smaller campuses with missions more focused on teaching. The team-designed model of teaching and evaluating teaching provides a good framework, and it fits well with the development and support of graduate students in those institutions with doctoral programs. It is an additional plus that those future college teachers will have the advantage of having participated in a very thoughtful exercise in instructional design and evaluation. ■



AAC&U MEETINGS

ANNUAL MEETING

**SHARED FUTURES /
DIFFICULT CHOICES:**
**Reclaiming a Democratic Vision
for College Learning, Global
Engagement, and Success**

January 25–28, 2012

Washington, DC

NETWORK MEETINGS

**GENERAL EDUCATION AND
ASSESSMENT:**
New Contexts, New Cultures

February 23–25, 2012

New Orleans, Louisiana

STUDENT SUCCESS:
Pushing Boundaries, Raising Bars

March 22–24, 2012

Seattle, Washington

**MODELING EQUITY, ENGAGING
DIFFERENCE:**
**New Frameworks for Diversity
and Learning**

October 18–20, 2012

Baltimore, Maryland

**NEXT GENERATION STEM
LEARNING:**
**Investigative, Integrative,
Innovative, and Inspiring**

November 8–10, 2012

Kansas City, Missouri

www.aacu.org/meetings



Increasing the Validity of Outcomes Assessment

- **Kathleen Langan Pusecker**, director, Office of Educational Assessment, University of Delaware
Manuel Roberto Torres, research analyst, Office of Educational Assessment, University of Delaware
Iain Crawford, acting chair of English, associate professor, University of Delaware
Delphis Levia, associate professor, geography, University of Delaware
Donald Lehman, associate professor, medical technology, University of Delaware
Gordana Copic, graduate student, University of Delaware

Increasing accountability in higher education has prompted institutions to develop methods to meet growing expectations that they will implement and demonstrate a commitment to the assessment of student learning outcomes. These responses range from ensuring institutional compliance with the mandatory requirements of accrediting bodies to the adoption of institution-specific assessment regimens designed to more closely align with local conditions and cultures. One such regimen is provided by the Voluntary System of Accountability (VSA), and this article explores how the University of Delaware (UD) implemented the Association of American Colleges and Universities' VALUE rubrics, as part of its ongoing initiative to create a campus-wide culture of assessment. As UD implemented the VALUE system for assessment, both the value of local adaptation and the limitations of adopting a national testing regimen have become increasingly apparent.

In 2007, the VSA was initiated by public four-year universities and two of the higher education associations—the Association of Public and Land-grant Universities and the Association of State Colleges and Universities to supply comparable information on the undergraduate student experience through the use of three standardized tests to assess core general education (Gen Ed) skills of critical thinking, reading, writing, and mathematics (quantitative reasoning). The argument for standardized tests is that they provide the best way to assess student learning through universal, unbiased measures of student and school performance. Critics claim, however, that these tests fail to assess accurately students' knowledge and school performance, may institute a bias against underserved populations, and are not reliable measures of institutional quality (Beaupré, Nathan, and Kaplan 2002). The VSA advocates the use of one of three standardized tests—UD chose to use the Educational Proficiency Profile (EPP). After an initial experience

with this test, UD decided to work with the VALUE project and its defined core learning goals and modifiable assessment rubrics. Choosing this path, institution leaders believed, would allow UD to examine student learning with greater sensitivity to local conditions and obtain more useful information on the quality and type(s) of student learning outcomes that are most challenging to evaluate via standardized tests.

VALIDITY OF EPP AND VALUE RESULTS

UD's Office of Educational Assessment (OEA) administered the abbreviated EPP to 196 first-year students and 121 seniors in fall 2010. ETS subsequently provided results in the form of scaled scores on the four core skill areas (reading, critical thinking, writing, and mathematics) as well as an aggregated individual score for each test taker. ETS stresses that institutions should not focus on individual results on the abbreviated EPP but instead concentrate on the aggregate scores. UD implemented the abbreviated version of the EPP, and ETS only provides mean scores from the long version of its Proficiency Profile Total Test. Since ETS did not provide mean comparative scores for the actual test UD students had taken, there was no precise basis for comparison, notwithstanding ETS's claim that the abbreviated test provides equivalent results. Thus, the actual scores obtained from ETS could offer little guidance in understanding how students at UD performed in comparison to test populations at other universities. Additionally, the lower number of questions in the abbreviated EPP created concern over content validity compared to results created with the longer version. There were, for example, only a total of nine questions designed to examine student proficiency in the area of quantitative reasoning.

In terms of the local testing environment, the methods required to obtain first-year students' and seniors' participation in this



Gen Ed study may well have affected the validity of the results obtained with the EPP. Freshmen who took the EPP were largely from First Year Experience (FYE) courses where faculty were asked to give up class time to offer students a chance to take this test. Seniors were mostly recruited from one class period in three different senior capstone courses. Finally,

as an additional strategy to recruit seniors, OEA staff placed open advertisements around campus in large classroom buildings requesting senior students to participate in the assessment, eventually offering a chance to win an iPad2 as an incentive to take the EPP.

These concerns over the validity of the EPP results motivated UD to explore a

different approach to assessing longitudinal learning outcomes. In summer 2011, the OEA hired three UD faculty members to evaluate Gen Ed competencies of critical thinking and quantitative reasoning. These Faculty Assessment Scholars (FASs) examined the artifacts provided by the same first-year students and seniors who also took the fall 2010 EPP. Since the student work samples assessed with the VALUE rubrics were provided by the same students who took the EPP, these data had the same bias challenges but, because the artifacts were selected by the students as their best work, this evaluation of authentic work samples did provide greater face validity. The FASs attended an hour-long training session to calibrate the two VALUE rubrics adapted by the OEA, one for assessing critical thinking skills (VALUE inquiry and analysis rubric) and

one for assessing quantitative reasoning skills. While this was a limited calibration exercise, it proved to be a successful model and one that could be readily enhanced in future assessment projects.

INTER-RATER RELIABILITY TESTING

Each criterion on the three separate rubrics (first-year critical thinking, senior critical thinking, and senior quantitative reasoning) assessed by the FASs was tested for inter-rater reliability. The inter-rater reliability among the three FASs was examined utilizing the intraclass correlation coefficient. The intraclass correlation coefficient is a measurement of agreement among multiple raters on the same set of criterion on the VALUE rubrics. For each criterion, ratings by all three FASs for all students' work samples were examined in a two-way mixed model to obtain intraclass correlation coefficients. The intraclass correlation coefficient, as a measure of inter-rater reliability, has a range of -1.0 indicating extreme heterogeneity to 1.0 indicating complete homogeneity (Kish 1965, 171). Ratings closer to 1.0 indicate less variation among the three FASs in evaluating student artifacts. Since there were six criteria and three rubrics, eighteen two-way models were produced.

Not all of the students who took the EPP chose to provide work samples; therefore, the OEA selected the artifacts to be evaluated based upon whether they contained enough information to be evaluated via the VALUE rubric before releasing them to the FASs. A total of ten first-year student critical thinking artifacts, ten senior critical thinking artifacts, and ten senior quantitative reasoning artifacts were evaluated by the FASs. The FASs utilized the quantitative reasoning rubric adapted by OEA to assess the ten artifacts from seniors; none of the first-year students provided an artifact. Table 1 illustrates the ratings given according to each criterion on this rubric. The ten quantitative reasoning

TABLE 1. QUANTITATIVE REASONING ASSESSMENT OF SENIORS

| QUANTITATIVE REASONING | RATING | SENIORS' PERCENTAGE |
|------------------------|--------|---------------------|
| Interpretation | 4 | 52% |
| | 3 | 43% |
| | 2 | 5% |
| | 1 | 0% |
| | | Means = 3.48 |
| Representation | 4 | 57% |
| | 3 | 29% |
| | 2 | 14% |
| | 1 | 0% |
| | | Means = 3.43 |
| Calculations | 4 | 75% |
| | 3 | 12.5% |
| | 2 | 12.5% |
| | 1 | 0% |
| | | Means = 3.63 |
| Application/Analysis | 4 | 58% |
| | 3 | 29% |
| | 2 | 13% |
| | 1 | 0% |
| | | Means = 3.46 |
| Assumptions | 4 | 50% |
| | 3 | 22% |
| | 2 | 22% |
| | 1 | 6% |
| | | Means = 3.17 |
| Communication | 4 | 52% |
| | 3 | 32% |
| | 2 | 16% |
| | 1 | 0% |
| | | Means = 3.36 |



artifacts were rated on a scale of 1 to 4 for the six criteria in this rubric. The table also illustrates the percentage of ratings received by seniors for possible rating on a scale from 1 to 4. The mean rating for seniors on each quantitative reasoning criterion is also provided.

The UD FASs utilized the VALUE inquiry and analysis rubric, adapted by OEA to assess the Gen Ed competency of critical thinking. FASs reviewed twenty artifacts from first-year students and seniors (ten in each group). Table 2 illustrates the percentage of ratings received by

first-year students and seniors for possible rating on a scale from 1 to 4 on the six criteria in the inquiry and analysis rubric.

Despite the fact that the FASs received minimal training to rate students' work, these findings indicate a high level of agreement.

TABLE 2. CRITICAL THINKING, INQUIRY, AND ANALYSIS ASSESSMENT OF FIRST-YEAR STUDENTS AND SENIORS

| CRITICAL THINKING CRITERION | RATING | FIRST-YEAR STUDENT PERCENTAGE | SENIORS PERCENTAGE | GAIN IN MEAN |
|------------------------------|--------|-------------------------------|--------------------|--------------|
| Topic Selection | 4 | 30% | 69% | |
| | 3 | 44% | 24% | |
| | 2 | 19% | 7% | |
| | 1 | 7% | 0% | |
| | | Mean = 2.96 | Mean = 3.55 | 0.59 |
| Existing Knowledge | 4 | 28% | 57% | |
| | 3 | 24% | 29% | |
| | 2 | 43% | 14% | |
| | 1 | 5% | 0% | |
| | | Means = 2.76 | Means = 3.43 | 0.67 |
| Design Process | 4 | 19% | 61% | |
| | 3 | 48% | 31% | |
| | 2 | 28% | 8% | |
| | 1 | 5% | 0% | |
| | | Mean = 2.8 | Mean = 3.58 | 0.78 |
| Analysis | 4 | 26% | 38% | |
| | 3 | 22% | 45% | |
| | 2 | 37% | 17% | |
| | 1 | 15% | 0% | |
| | | Mean = 2.60 | Means = 3.10 | 0.5 |
| Conclusions | 4 | 29% | 47% | |
| | 3 | 42% | 47% | |
| | 2 | 21% | 6% | |
| | 1 | 8% | 0% | |
| | | Mean = 2.92 | Mean = 3.40 | 0.48 |
| Limitations and implications | 4 | 30% | 41% | |
| | 3 | 40% | 48% | |
| | 2 | 25% | 11% | |
| | 1 | 5% | 0% | |
| | | Mean = 2.55 | Mean = 3.30 | 0.75 |
| Average Gain. 0.63 | | | | |

COSTS

Although the expense of using the Education Testing Service's EPP to examine Gen Ed competencies of UD undergraduates was not the determining factor in deciding on the VALUE rubrics as an alternative method of assessment, the costs were certainly substantial in relation to the benefits derived. By contrast, the direct and indirect expenses required to implement the locally-modified rubrics were far more modest. ETS charged \$7,000 for the scantron sheets, test booklets, and the processing and reporting of the test scores. Table 3 lists costs incurred for the EPP and costs necessary to utilize the VALUE rubric to assess Gen Ed competencies of undergrads independent of the EPP.

DISCUSSION

Implementing the abbreviated version of the EPP was time-consuming, expensive, and yielded results in which neither the UD FASs nor the OEA had great confidence. Because of the potential bias from students self-selecting to take the standardized test, there was little assurance that the results could be generalized to the student population at UD. Also, since the EPP is a low-stakes test and the abbreviated version provides a less intensive examination of skill levels than the more expensive and much longer version (40 minutes for the abbreviated versus 120 minutes for the standard EPP), its results could not be validly extrapolated as a measure of UD students' learning outcomes.

By contrast, using the VALUE rubrics allowed the creation of an assessment



regimen that was far more closely attuned to the actual learning experiences of UD students. It was possible, for instance, to obtain authentic samples of students' artifacts with minimal interruption to class time, and this model of data collection is readily expandable. An even more thorough representation of student work could be obtained, for example, through implementing e-portfolios with the first-year experience course: because this course is required for all UD students, faculty in charge of these courses could require that all freshmen students upload artifacts of their work so that a baseline measure could be obtained. A similar process could easily be implemented in senior capstone experience courses. Since almost all UD majors currently require a senior capstone experience, the artifact collection would be from the widest breadth of majors and consequently would allow the OEA comparative measures between first-year students' and seniors' Gen Ed competency levels. These methods of collecting student work would also eliminate

the costs of enticing volunteers with candy and expensive prizes, and they would save time by eliminating recruitment sessions and time required to administer standardized tests.

The use of a rubric tool, with its ability to isolate criteria for each goal, provided more useful and actionable information than simple levels of proficiency on large nebulous skills such as CT. For example, when UD FASs evaluated critical thinking skills, they observed that students' abilities to make assumptions and draw conclusions were their weakest skills. Results such as these allow the OEA to communicate to academic programs the need for students to be provided with further opportunities to develop this fundamental skill.

Good assessment should be cost effective and provide useful data with actionable results. By using adaptable VALUE rubrics and ensuring that the descriptors most closely match an institution's skill definitions, student learning can be assessed in the most meaningful

way. Students will complete graded assignments and these are most easily collected without an intrusion on class time. Students can be directed to upload assignments to a secure website where assessment faculty and staff can review authentic student work. A systematic assessment process that permits baseline data to be obtained for freshmen that can also be used to easily examine specific Gen Ed goals provides results that faculty can use to positively affect student learning outcomes. Therefore, as institutions continue to be pressured by external accreditation, they need to involve faculty in a cost-effective assessment process that allows them to connect assessment of Gen Ed to what is actually occurring in their courses and disseminate results to focus upon improving student learning. UD's experience using the VALUE rubrics and a team of Faculty Assessment Scholars created just this kind of involvement and can serve as a model of collaboration between academic faculty and administrative offices engaged in the institutional work of measuring—and improving—student learning outcomes. ■

TABLE 3. COSTS ASSOCIATED WITH IMPLEMENTING EACH TEST INDIVIDUALLY

| ITEM | EPP | VALUE RUBRIC |
|---|------------|--------------|
| Tests and processing fees charged by ETS | \$7,000 | — |
| Pencils for test completion | \$47 | — |
| OEA staff time to analyze data sent from ETS vs. Rubric Data | 60 hours | 25 hours |
| Registrar and OEA secretary time to secure classrooms for testing sessions | 10 hours | — |
| Candy for recruiting volunteers | \$41 | — |
| iPad2 | \$499 | — |
| OEA Staff time to recruit faculty in charge of the FYE and senior capstones to participate in open testing sessions | 30 hours | — |
| OEA staff time to attend all the FYE courses, capstone classes, and open sessions | 41.5 hours | — |
| Open sessions for senior volunteers: six senior "open periods" | 25 hours | — |
| Class periods taken from FYE courses: eleven periods | UC | — |
| Class periods taken from senior capstones: either periods | UC | — |
| Costs to hire 3 Faculty Assessment Scholars: \$705 x 3 | — | \$2,250 |

Hours = Indicates approximate time to implement this item

UC = Unspecified costs

REFERENCES

- Beaupré, D., S. B Nathan, and A. Kaplan. 2002. "Testing Our Schools: A Guide for Parents." <http://www.pbs.org/wgbh/pages/frontline/shows/schools/etc/guide.html> (accessed August 17, 2011).
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- Spirto, M., P. O'Mahony, and J. Malone. 2011. "Interrater reliability of the Melbourne Assessment of Unilateral Upper Limb Function for Children with Hemiplegic Cerebral Palsy." *American Journal of Occupational Therapy*, 65 (4): 378–383 <http://ajot.aotapress.net/content/65/4/378.full> (accessed August 15, 2011).
- Stattools.net n.d. "Intraclass Correlation for Parametric Data: Introduction and Explanation." http://www.stattools.net/ICC_Exp.php (accessed August 15, 2011).
- Yaffee, R. A. 1998. "Enhancement of Reliability Analysis: Application of Intraclass Correlations with SPSS/Windows v. 8" <http://www.nyu.edu/its/statistics/Docs/intraclass.html> (accessed August 15, 2011).



How Reliable Are the VALUE Rubrics?

► **Ashley P. Finley**, senior director of assessment, AAC&U

The VALUE reliability study was developed to gather data on the usability and transferability of rubrics both within and across institutions. This study was also designed to address the degree of reliability and consensus in scoring across faculty from different disciplinary backgrounds. Reliability data were gathered and analyzed for three of the fifteen existing VALUE rubrics—critical thinking, integrative learning, and civic engagement.

The effectiveness of assessment instruments is commonly evaluated by the degree to which validity and reliability can be established. Instruments should, therefore, both accurately capture the intended outcome (validity) and be able to do so consistently (reliability). Because validity is often harder to establish than reliability, it is preferable for assessments to contain multiple forms of validity. In important ways the rubric development process itself provided the VALUE rubrics with substantial degrees of two types of validity. First, because the VALUE rubrics were created nationally by teams of faculty, those people closest to student learning and outcomes assessment on campuses, the rubrics hold a high degree of face validity. The face validity of the rubrics is apparent in the scale of interest and circulation of the rubrics to date, as evidenced by the approximately eleven thousand people from over three thousand institutions and organizations, international and domestic, who have logged in on the AAC&U VALUE web page (<http://www.aacu.org/value/index.cfm>) to access the rubrics.

Second, the specific employment of faculty experts in particular outcome areas to populate the development teams provides the rubrics with additional content validity. Experts are commonly used to establish content validity to verify that “the measure covers the full range of the concept’s meaning” (Chambliss and Schutt 2003, 69).

The objectives for establishing national reliability estimates for the VALUE rubrics were two-fold. One, because the rubrics were created nationally and interdisciplinarily, we sought to emulate this procedure in order to establish a cross-disciplinary reliability score for each rubric. Two, we also sought to establish reliability scores within disciplines to examine the range of similarities and differences across faculty from different disciplinary backgrounds.

METHODS

Forty-four faculty members were recruited to participate in the study. Faculty members were evenly distributed across four broad disciplinary areas: humanities, natural sciences, social sciences, and professional and applied sciences. Each faculty member scored three samples of student work for each of the three rubrics. The initial round of scoring functioned as a “calibration round” through which scorers could familiarize themselves with the rubric and scores would be compared to previously determined calibration scores set by rubric experts. (Rubric experts employed to establish baseline scores for rubrics were national representatives already well-acquainted with rubric scoring and specifically with the VALUE rubrics.) If the faculty member’s scores were closely aligned with the pre-determined calibration scores, they were approved to go on to score two more work samples. If there was a divergence in scoring on particular criteria, the faculty member reviewed the discrepancy with the project manager and participated in a final round of calibration scoring before moving on to the additional work samples. The scoring process was conducted virtually. Scorers were given secure links to access samples of student work, rubrics, and the scoring sheet. Scores were entered online and uploaded to a secure spreadsheet for analysis.

The most common procedure for establishing reliability for rubrics is through inter-coder or inter-rater methods, by which two coders evaluate the same work sample, score it according to an approved rubric, and calculate a reliability score. To accommodate scoring from multiple raters, a multi-rater kappa reliability statistic was used. The multi-rater kappa statistic ranges from -1 to 1, where -1 indicates perfect disagreement beyond chance and +1 indicates perfect agreement beyond chance; a score of zero indicates perfect chance agreement (Randolph 2008). In terms of probability measures like reliability estimates, these scores indicate the degree to which raters’ scores are similar without simply occurring by happenstance alone, or put another way, due to dumb luck. The kappa reliability statistic, therefore, takes into account not just the amount of agreement on scores between rates, but also the probability that scorers will both agree and disagree with each other (see Viera and Garrett 2005). All



of this information is combined into the single, composite score (ranging between $-1 - +1$) described above.

A more straightforward measure of scoring consistency was also drawn using the overall percentage of agreement among scores. This statistic simply divides the total number of agreements among scorers by the total number of their agreements plus the total number of disagreements (see Araujo and Born 1985). Thus, the percentage of agreement score will often differ, sometimes markedly, from the score produced from the kappa statistic because the formula for the latter takes into account more information for determining agreement. The kappa statistic is based upon actual scores with additional information drawn from probability estimates, whereas the percentage of agreement calculation is based only on the actual scores and the probability of agreement by chance is not considered. Future work with these data will include additional reliability analyses that use other commonly

used measures of inter-rater reliability, such as Krippendorff's alpha, to provide further interpretation of these results.

ANALYSIS

The analysis of results from the scoring of student work samples were analyzed in three phases. The first phase of analysis assumes the objective is for raters to have exact agreement on scores. Scorers could give each rubric criterion a rating using a five-point scale, 0–4. Thus, the entire five-point scale was used in calculations (see column 1, below). However, in practicality, when working with faculty on campuses it is often not assumed that “perfect agreement” is necessary. It is assumed, rather, that close scores also count as agreement. To calculate “approximate agreement,” scores were reviewed prior to analysis and collapsed into four categories by calculating averages for each criterion and then combining close scores into one category. Additional measures of central tendency were also con-

sidered in determining which two response categories would be collapsed (i.e. median and mode) to assure that the fullest majority of data points were included among the two categories identified for consolidation. For example, if the average was 2.0 but the mode and median were 2.0 and 1.5, respectively, ratings of 1 and 2 would be collapsed into a single category. If the average was 3.2 on a criterion, all scorers who gave a rating of 3 or 4 were collapsed into the same category. Results from this analysis are displayed in column 2, below. In a few cases in which all three measures of central tendency were the same, all data points from the sample were scanned to determine the second most frequently identified response category for consolidation.

A third iteration of this procedure allowed for greater approximation of the average score by collapsing scores on either side of the average into one category. Therefore, using the previous example, if the average was 3.2, scorers who gave of rating of 2, 3, or 4 on the criterion were collapsed into the same category. This final procedure most closely resembles the process of calibration often used on campuses, where scores immediately adjacent to the most representative score are considered acceptable and only outlier scores are pursued for discussion. Results using this procedure are displayed in column 3, below. Standard deviations between disciplinary reliability scores are indicated in parentheses beneath each inter-disciplinary reliability result.

As would be expected, the percentage of agreement increases as fewer rating categories were considered in the calculations, thus limiting the possibility for difference, and the kappa scores also increased greatly (the standard for a “high” or “acceptable” Kappa score is 0.70). Across rubrics results indicate that there was greater agreement than disagreement among scorers (as indicated by positive numbers). The most conservative estimates of agreement and reliability, using the full range of scoring options on a

FIGURE 1. RESULTS FOR PERCENTAGE OF AGREEMENT AND RELIABILITY

| | COLUMN 1 Perfect Agreement | COLUMN 2 Approximate Agreement (Using 4 categories) | COLUMN 3 Approximate Agreement (Using 3 categories) |
|----------------------------------|----------------------------------|--|--|
| CRITICAL THINKING | | | |
| Percentage of Agreement | 36% (9%) | 64% (11%) | 89% (7%) |
| Kappa Score | 0.29 (0.11) | 0.52 (0.15) | 0.83 (0.11) |
| INTEGRATIVE LEARNING | | | |
| Percentage of Agreement | 28% (3%) | 49% (8%) | 72% (8%) |
| Kappa Score | 0.11 (0.04) | 0.31 (0.11) | 0.58 (0.11) |
| CIVIC ENGAGEMENT | | | |
| Percentage of Agreement | 32% (5%) | 58% (8%) | 78% (5%) |
| Kappa Score | 0.15 (0.06) | 0.44 (0.11) | 0.66 (0.08) |
| COMBINED AVERAGE FOR ALL RUBRICS | | | |
| Percentage of Agreement | 32% (6%) | 57% (9%) | 80% (7%) |
| Kappa Score | 0.18 (0.07) | 0.42 (0.12) | 0.69 (0.10) |



0–4 scale, indicated that nearly one-third of scorers had perfect agreement, though the overall kappa score was low. Allowing two categories to represent consensus, rather than one, provides a more compelling case for consistency across raters. Among the rubrics tested, the critical thinking rubric, which is also the learning outcome most often cited by faculty as a goal of their own teaching, had the highest degree of agreement and reliability; the integrative learning rubric, the newest learning outcome on the national scene, had the lowest scores on these measures. Relatively small standard deviations suggest that, overall, there was not great variation among scorers across disciplines. The critical thinking rubric had the largest amount of variation among scorers across the three analyses. Integrative learning and civic engagement have comparable levels of divergence among scorers.

ARE THE RUBRICS RELIABLE?

Consider the following scenarios:

Case 1. A group of faculty who have never met, who work on different and diverse campuses, and were trained in different disciplines are asked to score common samples of student work. This group engages in a single practice round and if issues arise they speak one-on-one with a rubric expert.

Case 2. A group of faculty from the same campus, who may be familiar with each other and perhaps work in the same department, are asked to score samples of student work. This group engages in an interactive practice round with a rubric expert present to discuss interpretation and application of the rubric before scoring work samples.

One could argue this study was conducted under circumstances that would most compromise the probability for finding reliability in scoring. We gathered a diverse sample of national faculty, who as a group had never participated in a common conversation about the structure of the rubrics or about their individual approaches to rubric

scoring. These faculty members also did not engage in a standard calibration process through which discrepancies in scoring could be reviewed as a group to better understand differences in interpretation and application of the rubric. Due to time and financial constraints, faculty discussed discrepancies with the project manager by phone without the benefit of hearing other perspectives and input from colleagues.

Given these challenges, the results produced from this study provide an encouraging starting point for larger scale national and campus reliability analyses of the VALUE rubrics. Of particular note is the relative convergence in scoring across raters, regardless of academic discipline. This provides support for the use of rubrics interdisciplinarily at the institutional level—the primary intent in their development. Additionally, the relative strength in the percentage of agreement and moderately high kappa scores, particularly in column 2 above, provide grounds for pursuing future studies that can be more comprehensive, rigorous, and—perhaps necessarily—campus-based.

We hope campuses will look toward case 2, above, as an ideal scenario for developing knowledge on the reliability of the VALUE rubrics. This study was undertaken, despite its caveats, to advance the discussion about the utility of rubrics nationwide and across faculty from different disciplines. It provides a model but not necessarily a guide. The collaborative work of campuses will be essential for carrying this work forward. ■

REFERENCES

- Araujo, John, and David G. Born. 1985. "Calculating Percentage of Agreement Correctly but Writing Its Formula Incorrectly." *The Behavior Analyst* 8 (2): 207–208.
- Chambliss, D. F., and R. K. Schutt. 2009. *Making Sense of a Social World: Methods of Investigation*. Thousand Oaks, CA: Pine Forge Press.
- Randolph, J. J. 2008. *Online Kappa Calculator*. Retrieved September 22, 2011, from <http://justus.randolph.name/kappa>.
- Viera, Anthony J. and Joanne M. Garrett. 2005. "Understanding Interobserver Agreement: The Kappa Statistic." *Family Medicine* 37 (5): 360–363.

Statement of Ownership, Management, and Circulation

(Required by 39 U.S.C. 3685)

(1) Title of publication: *Peer Review*. (2) Publication no. 3334. (3) Date of filing: 1/5/12. (4) Frequency of issue: quarterly. (5) Number of issues published annually: 4. (6) Annual subscription price: individual, \$35; libraries, \$45. (7) Complete mailing address of known office of publication: Association of American Colleges and Universities, 1818 R Street, NW, Washington, DC 20009. (8) Complete mailing address of the headquarters of the publisher: Association of American Colleges and Universities, 1818 R Street, NW, Washington, DC 20009. (9) Full names and complete mailing addresses of publisher, editor, and managing editor: Publisher: Association of American Colleges and Universities, 1818 R Street, NW, Washington, DC 20009; Editor: Shelley Johnson Carey, Association of American Colleges and Universities, 1818 R Street, NW, Washington, DC 20009; Managing Editor: none. (10) Owner: Association of American Colleges and Universities, 1818 R Street, NW, Washington, DC 20009. (11) Known bondholders, mortgages, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities: none. (12) The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes has not changed during preceding 12 months. (13) Publication title: *Peer Review*. (14) Issue date for circulation data below: Summer 2011. (15) Extent and nature of circulation (the first figure indicates average number of copies each issue during preceding 12 months; the second number indicates actual number of copies of single issue published nearest to filing date): (A) Total no. of copies: 8,941/9,200. (B) Paid and/or requested circulation: (1) Sales through dealers and carriers, street vendors, and counter sales: 415/398. (2) Mail subscription: 8,139/8,252. (C) Total paid and/or requested circulation: 8,554/8,650. (D) Free distribution by mail, carrier, or other means—samples, complimentary, and other free copies: 150/155. (E) Free distribution outside the mail—carriers or other means: 139/250. (F) Total free distribution: 289/405. (G) Total Distribution: 8,843/9,055. (H) Copies not distributed: (1) Office use, left over, unaccounted, spoiled after printing: 98/145. (2) Returns from news agents: 0; 0. (I) Total: 8,941/9,200. Percent paid and/or requested circulation: 95%/94%. (16) This statement of ownership will be printed in the Fall 2011/Winter 2012 issue of this publication. (17) I certify that all information furnished is true and complete (signature and title of editor, publisher, business manager, or owner): Shelley Johnson Carey, Editor.



AAC&U Assessment Resources



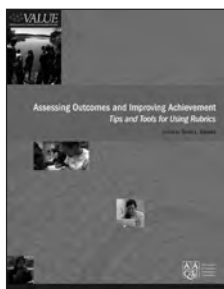
Valid Assessment of Learning in Undergraduate Education (VALUE)

VALUE is a campus-based project sponsored by AAC&U as part of its Liberal Education and America's Promise (LEAP) initiative. Supported with grants from the State Farm Companies Foundation and the Fund for Improvement of Post Secondary Education (FIPSE), VALUE supports and develops assessments that both evaluate and deepen student learning. VALUE advances an approach to assessment of learning outcomes that is based on examples of student work completed in the curriculum and cocurriculum and saved over time in e-portfolios. The project has also published fifteen rubrics developed and tested by faculty members on campuses across the country and that can be used to assess student work. VALUE rubrics have been tested on more than one hundred pilot campuses and include, among others, such outcomes as critical thinking, oral communication, intercultural knowledge and understanding, civic engagement, ethical reasoning, and integrative learning.

To learn more about VALUE, visit WWW.AACU.ORG/VALUE; to access the VALUE rubrics, visit WWW.AACU.ORG/VALUE/RUBRICS.

Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics

Edited by Terrel L. Rhodes



This publication provides practical advice on the development and effective use of rubrics to evaluate college student achievement at various levels. It includes rubrics for fifteen liberal learning outcomes, developed as part of AAC&U's VALUE project, and that can be adapted to reflect the missions, cultures, and practices of individual colleges and universities and their specific programs. The rubrics included in this publication were tested on more than one hundred pilot campuses and include

rubrics for outcomes often neglected in other assessment systems. These rubrics were developed by faculty members and academic professionals who articulated what important liberal learning outcomes look like at the beginning, intermediate, and advanced levels of accomplishment. (2010)

\$15 members/\$25 non-members

Electronic Portfolios and Student Success: Effectiveness, Efficiency, and Learning

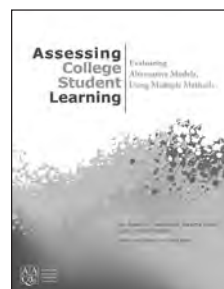
By Helen L. Chen and Tracy Penny Light



This publication presents an overview of electronic portfolios and ways individuals and campuses can implement e-portfolios to enhance and assess student learning, recognizing that learning occurs in many places, takes many forms, and is exhibited through many modes of representation. It is organized around eight issues central to implementing an e-portfolio approach: defining learning outcomes; understanding your learners; identifying stakeholders; designing learning

activities; including multiple forms of evidence; using rubrics to evaluate e-portfolios; anticipating external uses of evidence; and evaluating the impact of e-portfolios. This work is illustrated through multiple campus case study examples. (2010)

\$15 members/\$25 non-members



Assessing College Student Learning: Evaluating Alternative Models, Using Multiple Methods

By Robert J. Sternberg, Jeremy Penn, and Christie Hawkins, with Case Studies by Sally Reed

As colleges and universities respond to legitimate demands for accountability, it is essential that they clearly identify the student learning outcomes to be measured before they select an assessment tool to measure them. In *Assessing College Student Learning*, the authors seek to assist in this process, first, by exploring the tools currently available and, second, by examining the psychological theory of learning and achievement that underlies each of them. The publication also includes brief case studies exploring different comprehensive approaches to assessing student learning from St. Olaf College, Duke University, The University of Wisconsin-Madison, Miami Dade College, and California State University-Sacramento.

\$15 members/\$25 non-members



Our Primitive Art of Measurement

► Trudy Banta, professor of higher education and senior advisor to the chancellor, Indiana University–Purdue University Indianapolis

Every week brings news of another state legislature enacting ‘reforms’ of K–12 education, based on students’ test scores. Standardized test scores are being used to evaluate, compare, and fail teachers and schools. Will the “press to assess with a test” be focused next on postsecondary education? Will test scores be used to compare colleges and universities? Will public criticism, loss of public funds, and loss of students follow for some institutions? Will faculty feel pressure to raise students’ test scores, perhaps narrowing the curriculum to focus on the tested skills?

Certainly the 2006 report of the Commission on the Future of Higher Education, *A Test of Leadership: Charting the Future of U.S. Higher Education*, prompted moves in this direction with comments about the need for a simple way to compare institutions and public reporting of the results of learning assessments, including value-added measures. Today some one thousand colleges and universities are using one of three standardized measures of generic skills like writing and critical thinking to test first-year students and seniors; now value added can be measured, reported publicly, and compared among institutions.

Unfortunately, very little is being written about what these tests of generic skills are actually measuring and with what accuracy. Virtually nothing is

coming out about the validity of the value-added measure. We do know that the institution-level correlation between students’ scores on the tests of generic skills and their entering SAT/ACT scores is so high that prior learning accounts for at least two thirds of the variance in institutions’ scores. Out of that other one third, we must subtract the effects of age, gender, socioeconomic status, race/ethnicity, college major, sampling error, measurement error, test anxiety, and students’ motivation to perform conscientiously before we can examine the effects on learning of attending a particular college.

Institutional comparisons inevitably will be made on the basis of the negligible (even 1–2 percent) amount of variance that can be attributed to the contribution of any given college to students’ scores on these standardized tests of generic skills. We must argue for *multiple* measures of institutional effectiveness. Instead of relying on one assessment tool with serious limitations, we also must argue for the use of measures of learning that will provide specific guidance for *improving* curriculum and instruction.

Authentic assessment—using actual student work products that reveal their responses to the learning opportunities they are experiencing—is the best type of measurement for suggesting directions for improvement. For a decade I have argued that student electronic portfolios

evaluated with rubrics provide our best hope for assessing what students actually know and can do (validity). Portfolios can provide dramatic and convincing evidence of learning over time (value added) and there is evidence that multiple evaluators can achieve levels of agreement (reliability) exceeding 0.90.

We must continue to develop AAC&U’s VALUE rubrics and others that measure other outcomes to increase their reliability (through extensive training for raters) and to demonstrate their worth—or validity—in improving both individual students’ learning and the effectiveness of academic programs. But the very same concerns about using standardized test scores to *compare institutions* also apply to these authentic measures when they focus on generic skills. The current state of the art of measurement apparently is just too primitive to enable us to construct instruments that are valid for comparing the effectiveness of institutions in increasing student learning of generic skills. After all, the venerable SAT has undergone more than eighty years of continuous development, yet questions about its validity continue to exist. Measurement scholars, please don’t make us wait another eighty years for the ability to demonstrate convincingly to ourselves and to the public the variety and complexity of the student learning that is occurring in colleges and universities! ■



Association
of American
Colleges and
Universities

1818 R Street, NW
Washington, DC 20009

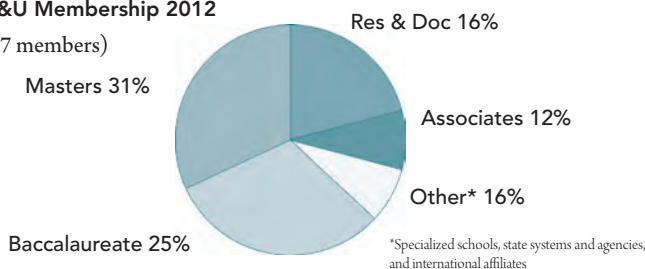
AAC&U is the leading national association concerned with the quality, vitality, and public standing of undergraduate liberal education. Its members are committed to extending the advantages of a liberal education to all students, regardless of academic specialization or intended career. Founded in 1915, AAC&U now comprises more than 1,200 member institutions—including accredited public and private colleges, community colleges, and universities of every type and size.

AAC&U functions as a catalyst and facilitator, forging links among presidents, administrators, and faculty members who are engaged in institutional and curricular planning. Its mission is to reinforce the collective commitment to liberal education at both the national and local levels and to help individual institutions keep the quality of student learning at the core of their work as they evolve to meet new economic and social challenges.

Information about AAC&U membership, programs, and publications can be found at www.aacu.org.

AAC&U Membership 2012

(1,287 members)



PERIODICALS
POSTAGE PAID

AAC&U's 2012 Summer Institutes



Institute on General Education and Assessment

June 2–6, 2012

The Hotel at Turf Valley
Ellicott City, Maryland

Deadline for application: February 17, 2012

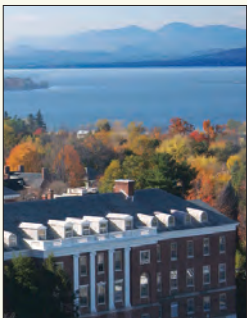


Institute on High-Impact Practices and Student Success

June 19–23, 2012

Portland State University
Portland, Oregon

Deadline for application: March 1, 2012

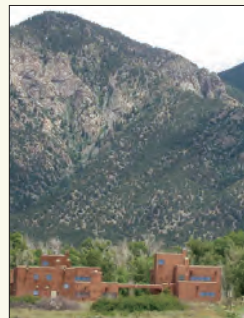


Institute on Integrative Learning and the Departments

July 11–15, 2012

The University of Vermont
Burlington, Vermont

Deadline for application: March 16, 2012



PKAL Summer Leadership Institutes for STEM Faculty

July 17–22/July 31–August 5, 2012

Baca Campus of Colorado College
Crestone, Colorado

Deadline for application: April 6, 2012