

ASSESSMENT OF STUDENT LEARNING

BY: CECILIA L. LOPEZ

Commitment to assessment THE NORTH CENTRAL ASSOCIATION (NCA) is committed to assessment of student learning. Accrediting associations cannot ignore the national imperative for institutional accountability and student performance. All six regional accrediting associations require that the institutions they accredit assess student learning across all their academic programs. In opening comments for the NCA/Commission on Institutions of Higher Education, Executive Director Steven Crow articulated the commission's continuing commitment to assessment: "My basic message about assessment today is two-fold: 1) assessment of student academic achievement remains one of the commission's highest priorities, and 2) the commission is in this venture for the long haul."

The Handbook of Accreditation, Second Edition, contains the commission's Statement on Assessment of Student Academic Achievement and the commission's policy relative to its philosophy and expectations for the implementation of effective assessment programs:

The program to assess student learning should emerge from and be sustained by a faculty and administrative commitment to excellent teaching and effective learning; provide explicit and public statements regarding the institution's expectations for student learning; and use the information gained from the systematic collection and examination of assessment data both to document and improve student learning. A strong assessment program is founded on a plan that is widely accepted and routinely updated, is ongoing, and is related to other planning and budgeting processes.

The focus here is on the status of the assessment of student learning at a random sample of 130 institutions (ninety-two baccalaureate and thirty-eight associate degree granting) that hosted comprehensive visits for reaccreditation by the Evaluation Team from January 1995 to December 1997. The information is drawn from Self-Study Reports of these institutions and the team reports that on-site Evaluation Teams provided for them, and from twenty progress reports providing evidence of implementation of assessment plans submitted to NCA staff for review and analysis.

Implementing assessment programs What are institutions reporting about their assessment programs? What are teams validating as common or good practice? Specifically, 1 focus on the implementation of programs that assess student learning across the general education program at a variety of two-year, four-year, and doctoral institutions, public and private.

Documentation of implementation is important. Evaluation teams request that institutions provide such evidence as:

- explicit objectives for student learning, publicly stated and linked to specific and direct measures of student learning;
- minutes of Assessment Committee meetings documenting how assessment information has been collected and interpreted;
- documentation of how the information derived from that interpretation of the data has been disseminated, to what constituents, and how often (i.e., the feedback loop);
- documentation of changes made as a direct result of information derived from the analysis of the data.

Teams advise institutions to avoid providing anecdote as a substitute for evidence. As one team wrote: We heard wonderful success stories; however, the college needs to collect and maintain documentation regarding the performance of the students across the stated goals for its general education program."

Explicitly stated objectives Having measurable objectives derived from broad program goals is an

important factor in improving assessment programs. Teams often commented negatively on assessment programs describing experiences students will be exposed to, but not assessing what students will be able to do as a result of that exposure. Many teams note, "The ability of an institution to document successful accomplishment of its mission and purposes depends entirely on its ability to define appropriate outcomes, directly measure behaviors relevant to those outcomes, and appropriately analyze and interpret the findings of its measures."

To be useful, an objective should be clearly expressed, measurable, and state explicitly what students will be able to do, not describe an internal state.

Evaluation teams questioned the usefulness of having learning outcomes for general education goals that were explicitly stated but that were not being measured appropriately. For example, they raised concerns about the practice of asking graduates whether or to what extent they believe they have achieved a competency and assuming responses constituted an objective measure of it. As one team stated, "Just as it is very important to select the best outcomes to define success, it is even more vital to select direct measures of such success."

Quantitative and qualitative measures There are differences among direct measures, indirect measures, and non-measures of student learning. Here I describe the types of instruments most often mentioned in the team reports I read. First, let me emphasize that both quantitative and qualitative measures were found acceptable.

Almost every institution is using some kind of survey or questionnaire that probes students' attitudes about their experience with the general education curriculum. However, evaluation teams consistently were critical of the heavy or exclusive use of indirect measures such as surveys. One problem with an over-reliance on survey instruments (e.g., student, alumni, employer surveys) is that they yield self-report data; providing only participants' opinions on what they have learned and on subjects not directly related to their learning, such as quality of instruction or competency of faculty. In other words, surveys do not focus on what the student has actually learned, nor do they in any sense measure it.

Commercially produced standardized tests Teams found that some institutions had chosen a standardized examination for superficial reasons: It is nationally normed, and it is affordable. The disadvantages of standardized tests are well known. For example, they do not adequately reflect areas of emphasis within the core curriculum, or they may not provide students an opportunity to sufficiently demonstrate skills in problem-solving tasks, or they may not adequately measure higher level thinking skills or the practical applications of knowledge. Not paying serious attention to the actual content of such tests until after the scores were in, institutions found the results were predictably disappointing. Students were unhappy with the test, and faculty rejected the results as trivial.

On the other hand, evaluation teams have reported that a number of institutions have experienced some degree of satisfaction with their use of certain standardized tests. One, Danville Area Community College, uses the Cornell Critical Thinking, Pre- and Post-Test. The pre-test is administered to new students, the post-test to students upon completion of a set number of earned credit hours across the general education requirements. Danville has used the test to determine value-added gains across the general education curriculum and has found it useful.

Evaluation teams comment positively on institutions that have adopted pre- and posttest designs that demonstrate value-added gains in student learning. Faculty from Adams State College and Dakota Wesleyan University determine value-added gains between matriculation and graduation by comparing, in order to estimate growth, the ACT test scores from entering students with the scores of graduating seniors on the ACTs College Outcome Measures Program Test (COMP).

Chadron State College compares the scores from four different standardized tests to assess student learning across its general education program: 1) the Watson-Glaser Critical Thinking Appraisal, assessing ability to reason critically, analyze, and solve problems objectively and think creatively; 2) ACTs COMP; 3) the Pre- Professional Skills Test (PPST) for proficiency in written and oral communication and language and symbols of mathematics; and 4) the Defining Issues Test (DID), to

determine ability to understand personal values and the ethical and moral implications of that knowledge.

Both the University of Nebraska-Omaha and Carthage College compare the writing skills of their students with nationally normed tests. They have chosen ETS's Academic Profile to provide baseline data that facilitates determination of minimum competency levels in students' thinking skills and value-added changes in students' development in writing, mathematics, and critical thinking.

Still other colleges (e.g., Hastings College, Estrella Mountain Community College) have adopted performance-based or constructed-response tests. An example of a commercially produced, performance-based test is Tasks in Critical Thinking, a standardized measure published by Educational Testing Service. Students are presented with a problem to be addressed in a combination of short answer and essay format. Required skills include gathering the information needed to solve the problem, analyzing and evaluating that information, organizing ideas, and presenting a solution to the problem.

Faculty may choose to score tests themselves or receive the scores from ETS, reported on groups of twenty to thirty students who have taken one of the ten Tasks. Each of the ten Tasks covers three skills areas—Inquiry, Analysis, and Communication—across multiple disciplines—humanities, social sciences, natural sciences. Hastings College reports that its faculty are willing to take the results of this instrument seriously, since it "is designed to measure the two skills deemed by the faculty to be the most important education goals of our core curriculum, namely, critical thinking and effective written communication."

Locally developed tests Chandler-Gilbert Community College and Saint Mary College are examples of institutions that have abandoned the use of commercially produced standardized tests, finding that 1) the material covered is not sufficiently congruent with their general education program goals and objectives, or 2) that the data derived from the tests do not provide useful information.

Typically, faculty decide to develop their own tests because they want more accurate information about their students' knowledge and ability. Locally developed tests focus on the curriculum, goals, and objectives of a specific institution, thus allowing better projections of student performance or behaviors beyond the level of the classroom. If such tests are well constructed, standards of validity and reliability can be adequate for these types of measures.

Western Illinois University used a Delphi technique to develop a 45-item instrument representing curricular areas identified by the Human Well-Being General Education Assessment Team. Student knowledge of the principles of wellness for living a healthy and fit life was measured quantitatively through forty multiple choice items and qualitatively through five open-ended questions. After pilot testing the data collected from first year students and seniors, the faculty determined, using the Kuder-Richardson Formula, that the test's reliability coefficient was .73. The assessment team members are working to improve the instrument's internal consistency.

Mesa Community College in Arizona used a combination of commercial and locally developed tests during its semi-annual Assessment Week. Students were randomly assigned one of four measures: a locally developed test of student learning in general education, ACT's COMP, making a short oral presentation, or writing an essay based on one of three prompts.

Faculty scored the locally developed measure using scoring rubrics developed by "faculty cluster groups." All responses were scored independently by at least two faculty members. Since only three areas of the six that make up the COMP's subscores related to Mesa's general education goals (communication, solving problems, and using the arts), its faculty reports and uses subscores from just those three areas. The oral presentation was rated by three faculty judges using a rubric describing eight outcomes developed by faculty members on the Communications Cluster Team. Four faculty judges scored the written essays, with each essay being scored by at least two judges.

Mesa's faculty are in the process of comparing the data from these multiple measures to make

decisions about the efficacy of the instruments, procedures, and scoring rubrics. In order for faculty to recommend changes, it is imperative that they have data which they can trust, that is, data derived from valid and reliable instruments and sound assessment methodology. Whether faculty decide to use a nationally normed, standardized test, or a locally developed measure, they should consider whether the measurement instruments, procedures, and scoring systems will need revision to obtain data supporting their decisions about program changes or improvements for student learning.

Portfolios A large and varied group of institutions, such as Creighton University, GMI Engineering and Management Institute, Johnson County Community College, Ohio State University, and Simpson College, uses the portfolio as a means of assessing some portion of their objectives-typically, writing-for student learning across their core or distribution requirements.

The faculty at the University of Detroit Mercy created a manageable assessment process which would focus on the general education core as a whole and not on each of the six objectives and their various subdivisions for the core. They decided to use the portfolio as a means of assessing specified attributes in three areas: general knowledge (e.g., draw on multiple sources), skills (solve systems of simultaneous linear equations), and values (identify the larger vision in the material).

Each instructor in a core course submits three exhibits of student learning which, in the judgment of the instructor, represents a) excellent work, b) average work, and c) a low pass. These exhibits, or artifacts, constitute the portfolio indicating broad levels of student achievement. The faculty members from the Core Assessment Committee blind score the artifacts and produce reports for faculty evaluation and discussion that provide a snapshot of the knowledge, skills, and values exhibited in student work in the core.

Examples of assessment of general education skills Examples of Good Practice highlight four very different types of institutions. While none has a perfect assessment program, these four have done something different from any others in the sample: Dordt College (Baccalaureate II), Grinnell College (Baccalaureate 1), the University of Wisconsin-Stevens Point (Masters 1), and Indiana University-Bloomington (Research 1).

Dordt College faculty developed its own measure for a construct fundamental to its mission and general education curriculum. Dordt, affiliated with the Christian Reformed Church, attracts students who characterize themselves as conservative on most social and political issues. With a strong commitment to their faith, 78 percent describe themselves as born-again Christians. For colleges like Dordt, where demonstration of biblical faith and ability to articulate and implement a "biblical reformed world view" are important goals, assessment of student attainment of those goals can be a challenge. As in all effective assessment programs, Dordt employs multiple measures to assess the cognitive, behavioral, and affective domains across its general education program. It triangulates the results from three measures to assemble the information it needs for decisions that can enhance student learning and development. The measures are a locally developed test called the Social Challenges Essay; the Cooperative Institutional Research Project (CIRP) Questionnaire; and the Academic Profile. Faculty developed the Social Challenges Essay to measure a student's ability to articulate and apply principles in the mission statement. Responses are also evaluated for value-added gains in critical thinking and analysis.

The Social Challenges Essay has an overall effective inter-rater reliability of .72, which is good for a subjective rating system Using a pre- and post-test design to determine value added, first year and senior students are required to write for one hour in response to four broad prompts concerning current social problems:

1. What are two significant challenges facing our society today?
2. Why are these challenges important?
3. What factors have contributed to the development of these problems?
4. How should Christians respond to these challenges and make a difference?

Faculty developed a scoring rubric, and the test is blind scored. The student essay is rated on the student's ability to: 1) demonstrate an explicit, analytical, and critical perspective; 2) form a moral

judgment; 3) reflect and affirm a world view; 4) articulate the importance or creational norms; 5) articulate personal involvement and responsibility for dealing with issues; and 6) state the historical, societal, or cultural context of societal problems.

Grinnell College, with its commitment to liberal education and specifically to one of its six goals, namely that of helping students to develop skills in analytical thinking and writing, requires all first-year students to take the Tutorial, a four-credit course with special attention to critical writing and thinking. Faculty from every department teach the core tutorial each year. To assess student learning in analytical thinking and writing, the faculty incorporated into their existing Faculty Summer Writing Seminars a new dimension for comparing firstyear tutorial papers with senior papers written by the same students. Senior papers comparable in scope and format to tutorial papers are chosen and blind scored, using a faculty-developed set of evaluation criteria.

The University of Wisconsin-Stevens Point has as its mission "To develop in students heightened intellectual, cultural, and humane sensitivities." The faculty have unequivocally stated that the general education program will provide students with a foundation in "fourteen skills and knowledges," and "they will learn to apply all fourteen." Data from multiple measures, both standardized and locally developed, have annually been collected and analyzed. As part of the feedback loop, faculty on the assessment committee have interpreted the data, drawn conclusions about their implication, and then reported the results to various constituencies. Annual assessment reports, submitted to the academic affairs committee of the Faculty Senate and the chief academic officer, are used as the basis for academic policy reforms affecting general education.

Indiana University-Bloomington, a decentralized Research 1 institution with five colleges, developed the "Student Performance Measure," an authentic, performance-based instrument measuring core abilities in reading and writing. With six essay prompts, each with its own set of interdisciplinary readings, the Student Performance Measure is unique because: 1) it demonstrates what students can do, not what information they know, and 2) it integrates multiple disciplines across the social sciences, physical sciences, and the humanities. The Student Performance Measure specifies exact and measurable objectives in terms of both the cognitive and behavioral domains. As a result of collecting and analyzing three years of data, the university has been able to document several outcomes and needed changes.

Challenge from the field All 130 institutions that make up this sample are at various stages of implementation. Most are still struggling to find the combination of methodology and measures to assess general education knowledge, skills, or competencies that will provide useful information to faculty about student learning. Few institutions have results that they report.

What is positive is that so many of NCAs accredited institutions do report heated and productive debates and discussions across their campuses about student learning, about what they really value and are therefore willing to evaluate and be held accountable for. The quality of a general education program is, I believe, what a faculty values and is willing to be held accountable for.

My experience of assessment in higher education leads me to recommend that colleagues:

- work to achieve a consensus on the institution's philosophy of general education and the image of the educated person they earnestly desire to graduate from their institution;
- fashion a curriculum that will express that philosophy;
- provide sorely needed leadership in systematically assessing student learning across the general education curriculum;
- seek agreement on the methodology, measures, and feedback loops to make the attainment of objectives real;
- collaborate with colleagues on how to utilize the information derived from those measures to make a difference in improving student learning, pedagogy, and the curriculum.

To the extent that we ignore the national and regional demands for quality assurance through the assessment and improvement of student learning, we invite federal and state intrusion.

[<< Return to Academic Assessment Mainpage](#)