

**The Commission's Assessment Initiative:  
A Progress Report**

Cecilia L. López, Ph.D.  
Associate Director  
North Central Association of Colleges and Schools  
Commission on Institutions of Higher Education

Presented at the 103rd Annual Meeting of the NCA/CIHE  
March 1998  
Chicago, Illinois

Based on a Paper Presented at the  
Association of American Colleges and Universities  
February 1998  
Tampa, Florida

## Outline

- I. Introduction to the Study
  - A. The Sample
  - B. The Data Sources
- II. NCA's Commitment to the Assessment of Student Learning
- III. Documenting Programs to Assess Student Learning: The Evidence
- IV. Measures of Student Learning
  - A. Program Goals and Objectives: Links with Specific Measures
  - B. The Importance of Sound Methodology and Measures
  - C. Quantitative and Qualitative Measures
    - 1. The Survey or Questionnaire
    - 2. Examples of the Use of Commercially Produced Standardized Tests
    - 3. Examples of the Use of Locally Developed Tests
    - 4. Examples of the Use of Portfolios
- V. Three Levels of Implementation: Characteristics
  - A. Beginning
  - B. Some
  - C. On-Going
- VI. Challenges

### **Introduction to the Study**

My presentation today will report on the results of a study I recently made of the status of efforts to assess student learning. The study examined data provided by a random sample of 130 of the approximately 250 institutions that hosted comprehensive Evaluation Team visits for reaccreditation within the last two academic years. Three sources of information were analyzed: the Self-Study Reports of 130 institutions, the Team Reports that on-site Teams provided to them, and the progress reports that 20 institutions submitted to Staff as evidence that their assessment plans were being implemented.

This progress report is primarily focused on the assessment of student learning in only one type of academic program, namely the general education program at a variety of 2-year, 4-year, and doctoral institutions, public and private. There are two main reasons for this focus: (1) the study was completed last month and was presented at the meeting of the Association of American Colleges and Universities; (2) the majority of Self-Study Reports and Team Reports that comprise the sample for this study suggest that assessing student learning across the general education curriculum is for many institutions a major challenge. However, most of what I will report today holds true for assessing student learning in any undergraduate major and any graduate or professional degree program.

### **NCA's Commitment to the Assessment of Student Learning**

NCA has made and continues to make a major commitment to the assessment of student learning. The Commission's "Statement on Assessment of Student Academic Achievement" indicates its expectation that an effective assessment program will be structured, systematic, on-going, and implemented (*Handbook of Accreditation*, 1997). The fourth paragraph of the policy statement reads:

"The program to assess student learning should emerge from and be sustained by a faculty and administrative commitment of 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, it is ongoing, and it is related to other planning and budgeting processes."

The Commission's commitment proceeds from its belief that assessment of student academic achievement is key (1) to improving student learning, (2) to enabling an institution to verify that it is being accountable to its internal and external constituents, and (3) to documenting to the general public and interested parties the value of investing in higher education.

This study reveals that NCA's emphasis upon the assessment of student learning is starting to have a major impact on all types of institutions, even before their programs are fully implemented. In preparation for starting the assessment program, a number of institutions have revisited their mission and purposes statement and have revised the goals for their academic programs to make them more explicit, measurable, and focused on student performance.

Many institutions that did not start to implement their assessment programs until after their assessment plans had been approved by the Commission in 1995, are still in the early phases of implementing the plan. However, a growing number of institutions have progressed far enough in the implementation of the assessment program to report changes that have been recommended by faculty. These institutions are starting to develop clear feedback loops to ensure that the results of assessment are documented, are regularly disseminated to appropriate constituencies, and are used by faculty to improve student learning, teaching pedagogy, and the academic curriculum. What the evidence suggests is that the culture of these institutions is becoming more focused on the improvement of student learning.

### **Documenting Programs to Assess Student Learning: The Evidence**

Based on Team Reports reviewed for this study, patterns of evidence that demonstrate useful implementation of assessment plans include:

1. clearly stated goals for all academic programs;
2. explicit objectives for student learning derived from those goals that are publicly stated and linked to specific measures;
3. the use of multiple instruments that include direct measures of student learning;
4. minutes of Assessment Committee meetings that document how assessment information has been collected and interpreted;
5. a feedback loop, that is, documentation of how the information derived from interpretation of the data has been disseminated, to what constituents, how often, and to what end.
6. documentation of changes proposed and made in teaching, curricula, and and/or academic support services as a direct result of information derived from the analysis of the data;
7. evidence of any improvement in student achievement following the changes; and
8. evidence that the assessment program itself is being periodically evaluated.

Many institutions in this study are investing substantial amounts of time and resources in providing professional development opportunities to assist faculty in carrying on programs for the assessment of student academic achievement. Faculty and academic administrators typically have needed considerable encouragement, training, and assistance in each of the following components of an assessment program:

- (1) in developing measurable objectives of student learning;

- (2) in identifying and utilizing multiple measures of student learning to assess those objectives;
- (3) in collecting and interpreting the data from those instruments;
- (4) in disseminating the information about the results of assessment in structured feedback loops that provide faculty and administration with timely and useful information upon which to base recommended changes; and
- (5) in using the information derived from the assessment of student learning to ascertain desirable pedagogical and curricular changes and to introduce and evaluate the effectiveness of the changes.

For those institutions that have started to implement their assessment plans, certain characteristics associated with the effective administration and implementation of assessment programs are now emerging in Team Reports on all types of colleges and universities. Teams and institutional administration and faculty are clear that a successful program to assess student learning must be faculty owned, developed, and implemented. Almost all have determined that accountable individual, such as the chief academic officer, should have administrative oversight or directly administer the assessment program.

Many institutions have determined that in order for assessment to become an institutional priority, faculty efforts need to be recognized and rewarded, and senior leadership needs to publicly express strong support for those efforts. Other institutions have found that establishing feedback loops that regularly communicate the results of assessment activities are critical because feedback loops generate a great deal of interest and debate on student learning across departments and the campus.

A growing number of institutions are now placing information about their assessment programs on the World Wide Web, making the process and for some, the results, available to internal as well as external constituents. Still others have instituted bi-annual Assessment Days or even an Assessment Week as a time to engage intensively in a variety of assessment activities, to display student work, and to celebrate student achievement.

However, because many institutions in this study have only recently introduced the programs called for by their assessment plans, a considerable amount of the information provided by institutions and Evaluation Teams was related to identifying and developing goals and objectives for student learning and to selecting direct measures that will produce results relevant to the specific objectives. This report, therefore, focuses primarily on those two aspects of assessment.

### **Program Goals and Objectives for Student Learning**

Having clearly stated program goals and specific, measurable objectives to attain each of the program goals is critical to conducting an effective assessment program. Teams often commented negatively on assessment programs that *describe* experiences that students will be *exposed* to but that do not *assess* what students will be able to *do* as a result of that exposure. As many Teams noted, to be useful, an objective should be explicit and

measurable. Objectives should state what students will be able to do, not describe what opportunities or services the curriculum, faculty, or the institution will provide.

### **The Importance of Sound Methodology and Measures**

Both qualitative and quantitative measures are useful sources of direct and indirect information about student performance. But a number of Teams have noted that scores or results of these measures are only useful to the extent that the test or measure selected has been found to meet standards for validity and reliability. That is, the instrument selected should consistently measure what the faculty has stated its students are to know or do (reliability). The instrument should also yield scores from which faculty may derive reasonable and logical inferences about student learning (validity).

### **Quantitative and Qualitative Measures of Student Learning**

The types of measures most often mentioned in institutional Self-Studies and in Team Reports were: surveys and questionnaires, commercially produced standardized tests, locally developed tests, and portfolios. Examples drawn from specific institutions will illustrate how each type of instrument is being used by institutions to measure knowledge, skills and values.

- **The Survey or Questionnaire**

Surveys are indirect measures of student learning. Almost every institution in the sample is using some kind of survey or questionnaire to probe student attitudes about their experience with the curriculum. However, Evaluation Teams were consistently critical of heavy or exclusive use of indirect measures such as surveys.

The principal problem with over-reliance on survey instruments (e.g., student, alumni, employer surveys) is that they yield self-report data; that is, they provide participants' *opinion* on what they have learned and on subjects not directly related to their learning, such as the quality of instruction or the 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. While the survey or questionnaire does have value as an indirect source of information about factors that often contribute to or detract from student learning, it cannot be considered a direct measure or substitute for a direct measure of student learning.

- **Examples of the Use of Commercially Produced Standardized Tests**

Teams found that some institutions had chosen a standardized examination solely for superficial reasons: it was nationally normed, and it was easy to score. Those facts alone did not always mean that the choice had been wise.

The disadvantages of standardized tests are well known, and Teams found a number of them in institutions' choices. A specific test may not reflect areas of emphasis within the core curriculum; or it may not provide students an opportunity to sufficiently demonstrate skills in problem solving tasks; or it may not adequately measure higher level thinking skills, the practical application of knowledge, or the development of values. Teams noted that institutions that had selected a standardized test without giving careful attention to its

actual content in order to determine whether the measure would adequately address the specific learning objectives established by the faculty for general education were disappointed with the results.

On the other hand, Evaluation Teams reported that a number of institutions had expressed some degree of satisfaction with the use of certain standardized tests for assessing student learning. They are being used for pre-and post-testing to determine value-added in regard to learning in specified domains.

Danville Area Community College (IL) is one example of an institution where this is being done. The College uses the *Cornell Critical Thinking, Pre- and Post-Test* to determine value-added student gains across the general education curriculum in such disciplines as the sciences, social sciences, humanities, and human health well-being. The pre-test is administered to new students, the post-test to students upon completion of 44 earned credit hours across the general education requirements.

Chadron State College reviews the scores from four different standardized tests to assess student learning across its general education program in the four areas for which the faculty had established objectives. The *Watson-Glaser Critical Thinking Appraisal* is used to assess students' ability to reason critically, analyze and solve problems objectively and think creatively. ACT's *College Outcome Measures Program Test (COMP)* and the *Pre-Professional Skills Test (PPST)* are used to assess proficiency in written and oral communication and the language and symbols of mathematics. The *Defining Issues Test (DID)* is used to determine students' ability to understand personal values and the ethical and moral implications of that knowledge.

Although not universally embraced by all of Chadron's faculty, annual assessment reports suggest that as the accumulation of longitudinal information from the assessment data and feedback expands, a growing number of faculty talk of the increasing impact data collection and analysis is having on identified areas of weakness and strength in the general education program. The Team judged that the assessment activities were making a positive impact on program reform and currency as well as on instructional delivery.

A large number of institutions are using the *Academic Profile* for longitudinal studies of the growth of their students' writing, math, and critical thinking skills across the general education program. For example, Shawnee State University (OH) and Taylor University (IN) use scores from the *Academic Profile* to provide baseline data that will facilitate determination of minimum competency levels in students' thinking and writing skills and value-added changes in students' development in writing, math, and critical thinking in the humanities, social sciences, and natural sciences from the beginning of their first year of study to the conclusion of their education at the institution.

Other colleges (e.g., Hastings College, NE; Estrella Mountain Community College, AZ) 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) that is being used by both of these colleges. Students are presented with a problem to be addressed in a

combination of short answer and essay format. The required skills measured include gathering the information needed to solve the problem, analyzing and evaluating that information, organizing ideas, and presenting a solution to the problem. Most importantly, faculty can score the *Tasks* test themselves, thus enhancing faculty buy-in to the results.

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.”

- **Examples of the Use of Tests Locally Developed by Institutions**

Typically, faculty decide to develop their own tests because they want more accurate information about their students' knowledge, abilities/skills, and values than that provided by commercially produced tests. Locally developed tests focus on the curriculum, goals, and objectives of a specific institution and thus permit more accurate data to be obtained about student performance or behaviors across an entire program and beyond the level of the classroom. If well constructed, measures of this type can meet standards of validity and reliability quite adequately.

A number of institutions have questioned the usefulness of data derived from the use of standardized measures. At Dakota Wesleyan University, for instance, faculty became dissatisfied when analyses of the results of successive years of pre- and post-testing showed significant gains in students' learning, but did not reveal useful patterns of evidence that suggested ways to improve the curriculum or student learning.

Some institutions have abandoned the use of commercially produced standardized tests for ones developed by faculty on campus. They gave as their reasons:

- (1) the material covered in the test is not sufficiently congruent with the program goals and objectives of the general education curriculum;
- (2) student interest and motivation to take the tests seriously are lacking;
- (3) cost and budgeting considerations make administering the test to large numbers of students prohibitive; and
- (4) the data derived from the tests do not provide useful information.

The faculty of some institutions decided from the beginning of their assessment efforts not to use standardized tests to assess general education skills/competencies. These institutions have typically developed a variety of their own measures in order to gather information faculty felt would be useful in assisting them to make informed curricular and pedagogical decisions.

Indiana University, Bloomington's, faculty developed six authentic, performance-based assessment measures to assess student proficiency in an important general education competency, problem solving. Bloomington, a highly decentralized Research I, with five

colleges, developed the *Student Performance Measure* to assess core abilities in reading and writing. Comprised of a series of six essay prompts, each with its own set of interdisciplinary readings, the *Student Performance Measure*, is a unique, locally developed test 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 specific and measurable objectives in terms of both the cognitive and behavioral domains. In addition, the *Student Performance Measure* uses six separate prompts that simulate a realistic situation that identifies the specific audience and then describes a writing task.

As a result of collecting and analyzing three years of data from administering the *Student Performance Measure*, the University has been able to document several positive reactions to the measure as well as some needed changes. First, the *Student Performance Measure* is perceived by both faculty and students as “academically meaningful and rigorous,” and therefore worth their time.

Second, because administration of the measure during the pilot stage was embedded in the curriculum of a one-credit introductory course, the sample was not random. Instead of using a cross-sectional sample, the Assessment Committee has recommended that longitudinal data be obtained across a sample of 3000 to 6000 first-year students and a sample of 1000 seniors.

As a result of the recommendation, each of the University's four largest colleges will provide a random sample of enrolled students. Two additional outcomes from the data have emerged. The four colleges are discussing how they can use these performance-based measures across their respective departments and, through task forces, are working on rethinking the general education curriculum.

#### **Example of the Use of Both Commercially Produced and Locally Developed Tests**

The majority of institutions in the sample use multiple measures, both commercially produced and locally developed, to assess student performance relative to explicit objectives for learning across the general education curriculum.

Mesa Community College (AZ) utilized a combination of commercial and locally-developed tests during its semi-annual Assessment Week. (logo: Show Me the Learning). Students were randomly assigned one of four measures: a locally developed test of student learning in general education, ACT's *COMP*, make a short oral presentation, or write an essay based on one of three prompts. Four faculty judges scored the written essays, with each essay being scored by at least two judges. The reliability coefficient of .45 reflected low internal consistency among the judges. This suggested the need for more rigorous training for faculty judges on how to use the rubric. Mesa's faculty are in the process of

comparing the data from these multiple measures to make decisions about the efficacy of the assessment instruments, procedures, and scoring rubrics.

Teams have noted that in order for faculty to recommend changes, it is essential that they have data which they can trust and that is useful. Data should be derived from valid and reliable instruments and sound assessment methodology. After faculty have pilot tested nationally-normed, standardized tests, locally-developed measures, or a combination of both, they should determine whether the measurement instruments, procedures, and scoring systems will need to be revised to ensure that they provide useful data to support faculty decisions about program or other changes needed to improve student learning and can be used on an ongoing basis.

Dordt College (IA) developed its own instrument to measure growth of Biblical faith, a construct that some educators have said cannot be assessed, but one that is fundamental to Dordt's Mission and to its general education curriculum. For colleges like Dordt, demonstration of "biblical faith" and an ability to articulate and apply a "biblical reformed world view" are important goals. Dordt College has succeeded in finding a satisfactory means to measure student learning across the general education curriculum that includes measuring student growth in values central to its Mission.

As is the case in effective assessment programs, Dordt College 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 to make 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*. Given the College's Mission and Goals, the *Social Challenges Essay* is the most interesting. Faculty developed the *Social Challenges Essay* primarily as a measure of a student's ability to articulate and apply the principles found in Dordt's Mission Statement. Student responses on the *Social Challenges Essay* 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-test and post-test design to determine value added, first year students and senior students are required to write for one hour in response to four broad prompts concerning current social problems.

Using a 7-point scoring rubric for the *Social Challenges Essay*, student essays are blind scored by faculty. Each essay is rated on the student's ability to demonstrate competencies as: (1) articulating an explicit, analytical, and critical perspective; (2) forming a moral judgment; (3) reflecting and affirming a world view; (4) and stating the historical, societal, or cultural context of societal problems.

The College has been documenting and reporting the results from the administration of all measures, including the *Social Challenges Essay* since 1995. Besides interpreting the data, the Assessment Committee has issued annual reports and has made specific recommendations to the Dordt community based on those data.

- **Examples of the Use of Portfolios**

Both the disadvantages and advantages of using portfolios are described in a number of Self-Study Reports and Team Reports that comprised the sample for this study. One disadvantage of using portfolios is that they are labor-intensive and thus may have substantial associated costs.

Teams suggest that there are other issues that institutions should address if portfolios are used as measures for assessment of student learning across a program such as general education. These include the need:

- (1) to assure that the sample is representative;
- (2) to develop appropriate protocols to deal with such procedures as the selection, and therefore consistency of “artifacts” that are placed in the portfolio;
- (3) to develop a reliable instrument and rating rubric to score the artifacts; and
- (4) to train faculty evaluators and reduce the likelihood of low inter-rater reliability among judges.

In spite of the challenges they present, characteristics associated with the effective use of portfolios as a direct measure of student learning are identified in a number of Team Reports. Evaluation Teams praise the use of portfolios as a measure when:

- (1) it is closely linked with goals and objectives for student learning;
- (2) it documents student learning and development across the student's academic career;
- (3) it contains a clearly stated rationale for the portfolio and specific faculty expectations for the learning students should demonstrate; and
- (4) when the results of portfolio review are used to make changes intended to improve student learning.

Because portfolios as a type of measure are perceived by faculty as being authentic and performance-based, portfolios are being used by a large number of different types of institutions.

Institutions as different from one another as Carthage College, Creighton University, GMI Engineering & Management Institute, Simpson College, and the Ohio State University all use the portfolio as a means of assessing some portion of their objectives for student learning (typically writing) across their core or distribution requirements for general education.

Johnson County Community College (KS) uses the portfolio to collect and review existing student work, called artifacts, which demonstrates student performance across the entire curriculum for each of the four major areas that faculty value.

Johnson County's philosophy is that the assessment of student outcomes should yield meaningful data from which decisions about curricular improvement can be made. Explicitly stated outcomes were developed by faculty divided into four sub-committees, each charged with operationally defining a specific component of the outcome statement.

Evaluation of the artifacts is conducted by three to four interdisciplinary faculty using holistic scoring rubrics and standards they developed for each of the stated general education outcomes. Results are reported in the aggregate, and each department addresses assessment results in its annual master planning document.

Another example of the use of the portfolio is the University of Detroit Mercy (MI). Its faculty created a manageable assessment process which evaluates learning over the general education core as a whole rather than measuring students' attainment of each of the six objectives and their various subdivisions established for the core.

They decided to use the portfolio to document the growth in students' learning across the core in three areas: General Knowledge (e.g., draw on multiple sources), Skills (e.g., solve systems of simultaneous linear equations), and Values (e.g., 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 excellent work, average work, and just passing work. The faculty members from the Core Assessment Committee blind score the artifacts and then produce reports of the results for the full faculty to evaluate and discuss. The Committees' reports provide the faculty with a "snapshot" of the knowledge, skills, and values exhibited in student work in the Core across one year.

### **Characteristics of Three Levels of Implementation**

The Self-Study Reports, Staff Analysis, and Team Reports reviewed for this study suggest that there are at least three levels of implementation of institution's assessment programs. I call them Level One, Beginning Implementation; Level Two, Some Implementation, and Level Three, On-going Implementation.

#### **Level One: Beginning Implementation**

Level One, Beginning Implementation, is typically characterized by the following evidence:

- Assessment plans are available for a number of academic departments or academic programs, but not all;
- Program goals and explicit objectives for learning have been developed for a number of academic departments or academic programs, but not all;

- Direct measures of student learning have been identified, and data above the level of the classroom have been collected for a number of academic departments or academic programs, but not all.

### **Level Two: Some Implementation**

Level Two, Some Implementation, is typically characterized by the following evidence:

- Senior Administration actively supports the assessment program and recognizes and rewards faculty's efforts to implement it;
- Results from measuring general education, the major, and graduate programs are being collected and interpreted from some but not all academic units or programs;
- Information about assessment is being disseminated to various constituencies;
- Evidence of specific strategies faculty are using to improve student learning is being gathered and documented;
- Evidence of specific strategies departments are using to assist faculty in improving student learning is being gathered and documented;
- Decisions about changing or refining methodology or measures are being made;
- Clear feedback loops are being developed to ensure that the results of assessment are used to improve student learning and teaching.

### **Level Three: On-going Implementation**

In addition to all of the characteristics of Level Two, Level Three, On-going Implementation, typically is characterized by the following evidence:

- The culture of the institution is now focused on student learning and how that learning can be improved;
- Proposed changes based on the analysis and faculty discussion of the assessment results are linked to departmental and/or institutional planning and budgeting processes;
- The results of assessing student learning are linked to academic program review;
- Changes in pedagogy, curriculum, and/or academic support services made as a result of assessment data are documented;
- The efficacy of the changes are evaluated and documented;

- The assessment plan is annually reviewed and updated;
- The entire assessment program and process is regularly evaluated and refined.

The institutions that make up this sample are at various levels in the implementation of their assessment programs. Virtually all are either in Level One (just starting various aspects of their assessment programs) or in Level Two (some implementation).

As Executive Director Steve Crow said in the Opening Session of the 103rd NCA Annual Meeting: "Institutions need to get serious about assessing student learning."

Too many institutions are still struggling to find the combination of methodology and measures to assess knowledge, skills, and competencies across the general education curriculum, the major, and graduate programs that will provide useful information about student learning to the faculty. And too few institutions have reported actual results or changes proposed or made as a consequence of analyzing those data. However, it is encouraging that so many of these institutions *do* report lively and productive debates and discussions about student learning across their departments and campuses.

For 1998-99, Evaluation Teams will expect institutions to have moved from Level One to Level Two. Additionally, institutions should be prepared to provide evidence that the assessment program is well on its way to progressing to Level Three.

The quality of a general education program is, I believe, the clear reflection of what a faculty values and is willing to be held accountable for. My challenge then is for all of you here today to provide leadership in documenting student learning across all of your academic programs, undergraduate and graduate.

Create useful feedback loops. Utilize the information derived from those measures to make a difference to improve student learning, to enhance teaching, and to strengthen the curriculum.

It will probably take a few more years before all institutions accredited by NCA can report, and Teams can validate, that changes made as a result of assessment efforts have, in fact, measurably improved student academic achievement. However, the potential of structured, on-going, systematic, and fully implemented assessment programs to transform all our institutions into student-centered, learning organizations and to bring about changes that will improve students' learning is real and exciting.