

Demonstrating student learning gains as part of an annual performance report:
Efforts by different states as a result of *Measuring Up 2000* and *2002*

Thomas P. Marcy, Ph.D.

Lakeland College

[August 14, 2008, Revised May 28, 2009]

In *Measuring Up 2000* and *Measuring Up 2002*, the state-by-state report cards on higher education performance, all 50 states received an Incomplete in the category of learning.¹ Although *Measuring Up* evaluated, compared, and graded the states in other key categories of higher education performance (including preparation for college, participation, completion, and affordability), the report card found that "there is no information available to make state-by-state comparisons" of higher education's most important outcome, learning. Subsequently, The Pew Charitable Trusts decided to sponsor [through a grant to the Institute for Educational Leadership (IEL)] an investigation into how to generate grades in the category of learning. The National Forum on College-Level Learning was created, a five-state demonstration project to develop a model of college-level learning for the states. Peter Ewell, vice president of the National Center for Higher Education Management Systems (NCHEMS), was the senior consultant to the project, although Dr. Ewell was just one of nine advisors. Margaret Miller, director of the Center for the Study of Higher Education at the Curry School of Education, University of Virginia, and editor of *Change* magazine, became director of the project. Between 2002 and 2004, the demonstration project team assembled information on the National Adult Literacy Survey (NALS) and on graduate-admission and licensure tests for each demonstration state. Meanwhile, the states administered general intellectual skills tests to a random sample of students at a representative sample of public and private two- and four-year institutions within their borders. The four-year institutions also attempted to collect information about their alumni's perceptions of their own intellectual skills. Also, both two- and four-year institutions in each state administered surveys aimed at gauging students' engagement with their collegiate experience, since research suggests that engagement is associated with learning. The engagement measures were subsequently dropped from the model, since they are not direct measures of learning. Experience with the demonstration project suggested that it was feasible to extend the approach to other states and eventually to create a nationwide benchmark for learning. For the learning category of the report card, indicators that reflected various dimensions of state performance were grouped under several overall themes, or clusters, as follows, and each was weighted:

- **Literacy Levels of the State Population (25%)** – This was originally based on the 1992 NALS for residents ages 25 to 64, updated using the 2000 census, and then was replaced by the National Assessment of Adult Literacy (NAAL) in 2004, in order to capture the literacy levels of the college-educated population rather than of the state population as a whole. This indicator addresses the question, "What are the abilities of the state population (using the NALS), or of the college-educated population (later, using the NAAL)?" States can determine the value added by a college education with respect to the literacy levels of the population.
- **Graduates Ready for Advanced Practice (25%)** – These indicators examine the proportion of the state's college graduates (from both two- and four-year institutions) who are ready for advanced practice in the form of vocational/professional licensure or graduate study [licensure examinations such as nursing, clinical pathology, physical therapy, respiratory

therapy, radiology, and physician's assistant; earning nationally competitive scores on competitive admissions exams such as the Graduate Record Examination (GRE), the Graduate Management Admissions Test (GMAT), the Medical College Admissions Test (MCAT), the Law School Admissions Test (LSAT), and the Pharmacy College Admissions Test (PCAT); teacher preparation measures such as passing a teacher licensure exam in the state in which they graduated]. This group of indicators addresses the question, "To what extent do colleges and universities educate students to be capable of contributing to the workforce?"

- **Performance of the College Educated (50%)** – At two-year institutions, the ACT WorkKeys assessments are used; at four-year institutions, the Collegiate Learning Assessment (CLA) is used. These indicators address the question, "How effectively can students who are about to graduate from two- and four-year colleges and universities communicate and solve problems?"

National or state benchmarks may be created by use of the above indicators. However, problems that arise as a result of using such indicators are (a) small sample sizes (particularly in smaller states), (b) the ways of determining common standards or norms by which to compare a state's results, (c) whether or not all states taken together should be used to determine national norms, or (d) whether norms should be determined by geographic region or by state population size and educational expenditures. As evidence of just one of these problems [sampling], the National Forum endorsed the recommendation made in the report of the National Commission on Accountability in Higher Education, sponsored by the State Higher Education Executive Officers (SHEEO), that the size of the adult literacy survey be increased.² There are two challenges the nation faces in the assessment of collegiate learning – a notable gap in the performance of white students and students of color on the direct measures of learning, and the uneven performance of states in preparing future teachers. But, in spite of these challenges, by use of consistent statewide information about learning outcomes in this way—to identify the specifics of a collective policy challenge that all institutions can help address – we are able to have comparable information across states that can help policymakers identify best practices and track progress. The model described produces information that is valid and useful for state policy, supplements existing accountability approaches and campus-based assessment efforts, leads to informed discussions about a public agenda for higher education, and can help provide motivation toward achieving that public agenda.³

Collegiate Learning Assessment Since the Collegiate Learning Assessment (CLA) carries so much weight as an assessment indicator for the 4-year institutions, this assessment indicator, its advantages, and perhaps its foibles, should be addressed in more detail. In July 2005, representatives from 35 colleges and universities, from across the United States, met for the first time to participate in the Council of Independent Colleges (CIC)/Collegiate Learning Assessment (CLA) Consortium, supported by a grants from the Carnegie Corporation of New York and the Teagle Foundation. Since that time, the consortium of institutions has grown to at least 47, and representatives of the institutions have gathered yearly to share the results of their testing and to develop continuous improvement strategies that build on their understanding of the CLA results. The project is intended to enable participating colleges and universities to test the effectiveness of CLA as an indicator of the "value added" to students' learning in their college years. Consortium

members began to administer the CLA, to groups of freshmen and seniors for three academic years, starting fall 2005.

The CLA has been offered as one way to provide tools that individual institutions can use for institutional improvement, and it can help to make the collective case for the effectiveness of independent colleges and universities. It may be particularly useful for institutions developing fundraising campaigns, institutional self-studies, or integrated marketing strategies.

So, what is the CLA? The CLA is an innovative examination that goes beyond multiple-choice testing by posing real-world tasks that students are asked to understand and solve. For example, students could be asked to draw scientific conclusions from a body of evidence in biology or examine historical conclusions based on original documents. Or they might be asked to write a persuasive essay, and analyze and then refute a written argument with logic and evidence. The CLA was designed to measure the cognitive growth of students and attempts to define the “value-added” contributed by the institution to student learning. The CLA is intended to provide direct evidence of student learning occurring over a typical four-year college program. The CLA provides a “value added” scoring of the institution’s unique contribution to the intellectual growth of students on aspects of cognitive development considered important in the liberal arts – critical thinking, analytic reasoning, problem solving, and written communication. The CLA test battery consists of two types of assessments—a set of four authentic tasks and a set of two writing prompts drawn from the Graduate Record Examination (GRE).⁴ An institutional score report from the CLA testing informs a college or university as to whether it is performing at, above, or below expected levels, given the entry characteristics of its students. Furthermore, because scores are standardized, the CLA can be a useful benchmarking tool to assess progress on student learning relative to the intellectual growth of students at other institutions. It is possible, for example, for a college with entering students who scored lower on the ACT or SAT to outperform more selective institutions, if the students at the less-selective college demonstrate greater gains on the CLA measures. Again, all of the CLA measures are open-ended, instead of being multiple-choice. Gain scores are calculated using SAT scores as a control, reported at the institutional level, and presented with comparisons to other institutions with similar characteristics.

The institution, not the student, is what is being measured with the CLA. There is no fixed standard of proficiency with the CLA. Progress is relative, depending on the characteristics of the students enrolled at the institution. As efforts of the CIC/CLA consortium progressed, some institutions have chosen to sample additional students with the CLA for comparisons between specific groups of students. Consortium institutions designate three-person teams to guide the institutional assessment efforts, agree to share assessment data with other Consortium members, collaborate on developing successful approaches to using the data, and participate in annual summer conferences with other Consortium members.

Margaret Miller, project director of the National Forum, noted that the CLA is “the first viable attempt to bring accountability and improvement together” in an assessment approach. Miller also recommended linking CLA results with other campus assessment data, including graduate admission and licensure exams, and surveys of student experience, such as the CIRP Freshman Survey and the National Survey of Student Engagement (NSSE), campus-based learning portfolios, or classroom measures.^{5,6}

As one example of institutional use of the CLA, the University of Wisconsin (UW)—LaCrosse had representative samples of freshmen and seniors taking the test during 2007-2008, approximately 100 freshmen in fall 2007 and approximately 100 graduating seniors in spring 2008.

Challenges for State Implementation Challenges to the states have begun to be addressed through collaborative work between the central coordinating boards and the colleges and universities. For example, in the State of Missouri, the Missouri Department of Higher Education (MDHE) and a number of Missouri colleges and universities established the Missouri Consortium for Measuring Value-Added Student Learning (MVASL), a two-year commitment focusing on the use of the CLA in participating institutions. [Moreover, the Missouri Assessment Consortium provides a discussion forum, along with its *Guiding Principles on Assessment*, for Missouri assessment coordinators.] Similar efforts have taken place in other states.

A clear commitment of each state's coordinating board [or similar body] and its chief executive is a necessary condition for obtaining the campus cooperation that is crucial to implementing the model described above. Active involvement of the chief executive is particularly important – especially in the early stages of implementation, for generating political impetus and board support, creating buy-in from the campuses by convincing them of the value of this kind of assessment for various stakeholders, reducing the threat of inter-institutional comparison that cross-campus assessment might seem to pose, encouraging and supporting campus leadership, and using the results to help create a public agenda for higher education and enrich each state's accountability system. This individual needs to be able to (a) consult with senior campus leadership, (b) channel resources to the institutions, (c) keep campus personnel informed through the various phases of implementation, and (d) consult with them about the strategies and protocols for testing. Through periodic meetings and regular email communication, the state coordinator would need to provide campus leaders with crucial administrative and moral support. Such lines of communication are important, particularly timely information about the purposes and value of the project, the psychometric properties of the assessments to be used, and effective implementation strategies.

Campus leadership is as important as state leadership. Campus coordinators are the people most directly responsible for the success or failure of the effort. Campus coordinators should be involved early in the design of sampling and testing procedures, since they know best how to recruit students for local assessment efforts. The institutional president needs to communicate the purposes of the project to all campus constituents, ensure that the campus coordinators are provided with the resources (time, money) required, and ensure that the results are disseminated and used for benchmarking results of assessments. A huge difficulty is the incentive for students to take the tests seriously. Paying students, or similar incentives, to take tests may not always be the best answer, and it is a costly one. Embedding the tests, such as the CLA, into appropriate courses in the general education curriculum, for both pre-testing and post-testing, may be the best way to provide incentive for students to try to do their best on the tests. Regardless, this approach also includes the logistics of scheduling rooms and computers for testing, dealing with software problems, and accessing technical support, not to mention that the courses used would need to be those of at least 2-hour class periods.

Campus coordinators of 4-year institutions would also need to collect other needed information, such as GRE scores, IPEDS data, SAT scores, and so forth. All test scores except GREs can be obtained directly from national sources. GRE scores may be compiled by asking participating institutions in each state to request their scores from ETS, which can be done via a standard report, for a small fee. Adjustments of sub-score aggregations, standardizing scores, time period aggregations would need to be undertaken by the state coordinator. Sampling procedures for the

general intellectual tests (WorkKeys and CLA) require a total sample of some 1,200 test-takers for each of the two test batteries in a given state. This necessitates a cluster-sampling approach – first, a sample of institutions is drawn, and second, the sample of students to participate from each institution is selected. This sampling approach represents a compromise, based on the conflicting need to attain some degree of statewide representativeness and the desire to include enough test-takers at participating institutions to enable them to use the resulting data for local purposes. Thus, about 75 to 100 test-takers at 12 to 15 four-year institutions, and at an equivalent number of two-year institutions, would be required in any state undertaking this task. The applicable institutions (four-year public, four-year private, and two-year) should be divided into groups of roughly comparable institutions. Variables used to construct these groups should at a minimum include institutional size, type, disciplinary mix, selectivity, urban/rural location, full-time/part-time ratio, and racial/ethnic distribution, resulting in five to seven distinct groups of institutions within each category of institutions (public four-year, private four-year, and two-year). In the case of the CLA, test-takers should be asked to complete either one task or two GRE prompts. The total testing time for the CLA battery administered in this way would be just over two hours. Details of costs are found in *Measuring up on College-Level Learning*.³

Associated with these efforts by the state and campus coordinators, as outlined above, other necessary efforts in the states, though not directly related to statewide “value-added” initiatives but supportive of the curriculum offerings in the State, include program review, transfer and articulation policies, and curriculum alignment. Program review ensures high quality programs offered in the state, efficient use of resources, and collaboration between institutions, but also helps to prevent unnecessary duplication of academic programs within a service region. Transfer and articulation policies provide avenues for reviewing and making recommendations on transfer issues, study and develop transfer guidelines for traditional and non-traditional credits, and review and recommend resolutions on cases of appeal from institutions or students. The main purpose of the curriculum alignment is to identify competencies for entrance into and exit from beginning-level collegiate courses.

A final word on statewide efforts should include mention of the Voluntary System of Assessment (VSA) – an educational assessment being developed by American Association of State Colleges and Universities (AASCU) and the National Association of State Universities and Land-Grant Colleges (NASULGC), in order to demonstrate accountability and stewardship to the public and to identify effective educational practices.

¹ *Measuring Up 2000: The State-by-State Report Card for Higher Education* (San Jose: National Center for Public Policy and Higher Education, 2000).

² <http://www.sheeo.org/account/accountability.pdf>

³ October 2005. Miller, Mary A., Ewell, Peter T. *Measuring up on College-Level Learning*. The National Center for Public Policy and Higher Education (National Center Report #05-8).

⁴ www.cae.org/content/pdf/CLAOpportunityToParticipate.pdf

⁵ Summer 2005. CIC/CLA Consortium Participants Discuss Campus Goals.

⁶ Summer 2007. New Opportunity to Assess Student Learning.