

ASSESSMENT AND ACCOUNTABILITY: UNVEILING VALUE ADDED ASSESSMENT IN HIGHER EDUCATION

**A Presentation to the AAHE National Assessment Conference
Denver-June 15, 2004**

Richard H. Hersh, Senior Fellow

CAE/RAND

Introduction

As I begin this presentation let me note that what I am about to say is hardly my thinking alone. Roger Benjamin, President of the Council for Aid to Education, and I have been working closely on this for four years and our many conversations and joint writing are reflected in what I am about to say. I have taken much of the material that we and our colleagues have published together or separately as substance for this talk and I have prepared a short listing of those sources that will be available at the conclusion of this session.

I want to speak today about the opportunity and challenge before this nation—to improve higher education at a time when we are struggling to fulfill the educational, economic, and civic promise college and university affords. Increased access to colleges and universities has raised the hopes and dreams of so many once left at its gates but higher education is increasingly being asked to become more accountable. “What difference do you make and how do you know?”

The Inextricable Relationship Between Assessment and Accountability

The key to answering that question and to improving our quality requires far better assessment of student learning. Let me be clear from the outset— I believe it is an educational, professional, and ethical duty of higher education to assess its impact on student learning in ways that promote our educational mission and at the same time improves accountability. Further, I will suggest that the best approach we can take in measuring student outcomes is to use value added assessment. Finally, I will describe the results of the Collegiate Learning Assessment Project (CLA) as an example of the promise of value added assessment.

Assessment

Why is assessment of student learning central to improving quality and in responding to the call for greater accountability?

First, because our primary responsibility is educational—student learning is our *raison d'être*—and we know that appropriate and timely feedback to students, which requires formative and summative assessment, increases student learning. Appropriately assessing student learning is the best way we as “learning organizations” can explain and defend to ourselves, our students, and others our curricula, pedagogies, and proposed changes.

Second, we in the academy *ought* to take the lead on assessment and accountability because we are professionals. Because of our training and professional status we are obligated and best equipped to assess learning. Might I also say that if we do not do it, others less capable will do it for us.

Third, whether private or public, we are all subsidized by the public, either directly through tax revenues or by the subsidy for privates through tax exemption. We thus have a responsibility to the public to be transparent in our assessment and accountability. Answering the question “what difference do you make?” can best be answered by including evidence of student learning.

The Access Revolution and the Issue of Quality

Given the global and economic realities of the 21st century, we now take for granted that higher education is virtually a necessity for all students. Indeed, fully 70% of high school students now enroll in college, a far cry, for example, from the early 1950s when that figure was under 25%. We have truly fashioned an access revolution whose American spirit began in the establishment of the Land-Grant Universities and echoed through the GI Bill, National Defense Loans, Higher Education Acts, and affirmative action. And with more access has come diversity with women now representing more than 56% of undergraduates and 50% of graduate students with African-American and Hispanic students in much larger numbers than ever before. This is not to say that the revolution is complete but only to say that we have made remarkable progress.

Now more than ever, economics is the driving force impelling access and fueling the critique of higher education. A college degree is “Sold” by those in and out of the academy as the necessary passport to future riches and that message has been bought by virtually everyone, so much so that a major complaint is that college has become unaffordable and threatens access. In turn, critics charge that higher education quality has suffered as we have become bigger, more impersonal, and money-grubbing to keep up with enrollment increases and the Harvard “Joneses.” We are charged with allowing students to leave college with little or no knowledge of history, literature, science, math, foreign languages and an inadequate ability to speak or write adequately. The pressure to demonstrate our “value” and justify the cost/benefit of our existence has thus increased.

The issues of quality, cost, and accountability have become intertwined and are now squarely on the public policy agenda.

The Pernicious “Guide” to College Quality

State legislatures and regional accrediting agencies have been focusing on the questions of cost, quality, and accountability in earnest for the past decade but getting few answers. Strange but true, with all the published information on reputation, endowments, graduation rates, and average SAT scores, no one can actually say whether one college or university produces better educated graduates than another.

Even with college guides, U.S. News & World Report's rankings, guidance counselors and college admissions specialists, there is still no accurate and effective measure of how well individual colleges and universities actually prepare their students for life following graduation. The most pernicious of all, however, has been the U.S. NEWS & World Report's annual rankings. Why do I say that? Because they are using data and creating rankings under the false pretense that “input” variables are the equivalent of output quality. And worse, they know exactly what they are doing and it does sell. There have been numerous critiques of their methodology and a study several of us conducted in 1997 demonstrated that one of the variables they use in their equations, “endowment per student,” predicts almost 70% of all the other variables such as class size, alumni giving, faculty salaries, selectivity, etc. Wealth, and as it turns out, the age of a college or university, predicts very well the ranking in the top tiers of schools. What the data and rankings do not disclose, however, is the actual *quality* of the school in terms of what students learn and are able to do because they attended that school.

To be fair, one cannot blame U.S.NEWS alone for this rankings arms race. Campus administrations and Boards of Trustees blithely submit the annually requested data even while most complain bitterly that the rankings are unfair. But more fundamentally, U.S. NEWS and its ilk have filled a vacuum—currently there are no widely-used and generally acceptable ways of measuring student outcomes. This is why I said earlier that we in the academy have the educational and ethical obligation to take the leadership in developing student learning assessment strategies lest we fall prey to others in the media and legislatures who are increasingly interested in doing so.

What does it really mean to get an "excellent" education? How might colleges and universities, not to mention parents and legislators, increase their confidence that students have become appropriately educated after four years? It seems to me that we as educators need and have the opportunity to ask and answer those questions more publicly than ever before. The quest for those answers is part of the national public-policy conversation yet many in the academy argue that assessment and accountability cannot be mixed in fear of losing academic freedom. I maintain this is specious reasoning and want to make the argument that without academic leadership on the accountability issue we will do a disservice to ourselves and our students by allowing others to continue to step into that void and thus place academic freedom at far greater risk.

The Call for Accountability and Assessment of Student Outcomes

During the past twenty years the issues of access and cost have dominated the public discussion of higher education. But more recently, in the shadow of the K-12 high-stakes testing and “No Child Left Behind” movements, the issue of higher education quality has been moved to the front burner in two ways. The first has been the call for higher education to demonstrate its educational efficacy. More than forty states, for example, now have laws on the books requiring that public institutions provide evidence of student learning. Moreover, state legislators, students, and parents, concerned about rising tuition costs, have increasingly called for more public accountability.

The second has been the emphasis on assessment of higher education quality in terms of student learning outcomes. Each of the regional accreditation associations, the historic arbiters of higher education quality, now specifies assessment of student learning as the ultimate criterion of educational quality. The Association of American Colleges and Universities (AAC&U) in its national two-year study of higher educational quality, (*GREATER EXPECTATIONS: A New Vision of Learning as a Nation Goes to College*) calls for institutional accountability based on student learning. In a recent paper devoted exclusively to the subject of accountability, “Our Students’ Best Work,” AAC&U asserts that “too many institutions and programs still are unable to answer legitimate questions about what their students are learning in college.”

The corporate community weighs in as well. The Business-Higher Education Forum recently disseminated its report **PUBLIC ACCOUNTABILITY for STUDENT LEARNING in HIGHER EDUCATION: ISSUES and OPTIONS**, in which they argue strongly for measures of student learning as the central component of a higher education accountability system. “One of the most important public policy imperatives in higher education is to enhance institutional productivity by focusing on learning.....” (p.13).

These new calls for accountability, in particular the use of student learning as the key accountability variable, are particularly telling at a time in American history when public financial support for public institutions has greatly eroded. Once supported to the tune of 40-70% of the budget, state-institutions are now calling themselves “state-assisted” with many now receiving only 15-25% in state-funding. In the both the public and private spheres, tuition has risen precipitously to maintain “quality” and in lieu of any solid evidence to the contrary, the market has for the moment grudgingly borne the higher prices in the belief that the perceived future economic value of a degree is worth the cost.

Public Good versus Private Benefit

This moment of tolerance, however, is to be short-lived. Higher education is increasingly at risk and the faith in higher education has begun to erode. Where once this country was unified in its belief that higher education is a public good (hence the extensive network of land-grant institutions, tax exemptions for all of higher education, the federally supported GI Bill), we have increasingly come to believe that individual beneficiaries of higher education should shoulder most of the cost and thus we have been able to rationalize decreased state tax support, sharply rising tuition, and federal and state grant aid increasingly being replaced by student loans, and individual student debt.

In addition to the increasing intrusion of a marketplace mentality as part of the American public's ethos, the transformation of American thinking about its higher education systems as less a public good and primarily a private benefit has occurred because of the way higher education has presented itself in terms of explicitly promising that a college's ranking and a degree's cost is worth the ultimate, individual financial gain. But that too is now being questioned. I am reminded of the study done by the Princeton economist Alan Kreuger a few years ago, in which he tracked the earning record of students who were admitted to Ivy League universities but chose to attend non-ivies versus the earning power of Ivy graduates. There was no difference.

But the increasing loss of public faith may also be a result of the lack of a transparent accountability system that could continuously renew people's faith in the public value of higher education. The confusion that a college education is nothing more than another marketplace commodity to be purchased as an individual choice is not helpful. And, finally, our academic language has hurt. While higher education is rightly touted as the fuel for the development of society's "human capital," such language and implied focus solely on market utility may also help to alienate the public from its colleges and universities in terms of its understanding the larger public good served by "higher" education. Compounding the issue, the increased calls from state and federal voices for academic accountability have put the academy on the defensive. We in higher education have little empirical evidence to offer in support of undergraduate quality and our consequent defensiveness has only emboldened our critics.

Reconciling Student Assessment and the Legitimate Need for Accountability

I believe what is needed now is a reconciliation of assessment and accountability. I suggest that the academy has the opportunity—I would say responsibility—to take the lead in reasserting the merit of the public good argument. And the best way to do that is to implement a learning assessment system that first and foremost helps to improve student-learning and institutional efficacy and second, provides appropriate student learning evidence in response to calls for accountability. The key issue is to develop an assessment of learning metric that meets the following criteria: that it measures important outcomes; that it is transparent and comprehensive; that it also can be used with the 40% of students who attend more than one college; and that it is initiated and controlled by the professoriate rather than imposed by agencies outside of the academy. Such a metric, if properly communicated, would go a long way in reconciling the various calls for

improvement, the measurement of “value,” and the increasing demands for accountability.

To begin it might be useful if higher education acknowledged its historic passivity on this subject. Until now the academy has successfully and resisted any attempt at externally imposed accountability claiming it a violation of academic freedom, claiming that the professoriate was already doing a satisfactory job via its own course tests and grades, and claiming that many of the most important things that one would want to teach are not capable of being measured. Yet, significant grade inflation has given the lie to the notion that grades represent quality of learning. A recent story in the NEW YORK TIMES tells of corporate recruiters asking new college graduates what their SAT scores are because such scores are perceived as better measures of talent than inflated grades!

Assessment of student learning must be acknowledged by the academy as inextricably linked to the teaching-learning process and as part of the feedback loop that can enhance institutional efficacy. Indeed, few question the legitimacy of many existing assessments, such as the Medical Board exams, Bar exams, and certifying exams for engineers, architects, and nurses. But these are all summative examinations and research on learning tells us we would be far better off educationally if we formatively assessed student learning earlier and more often. Appropriate assessment and feedback to students in a timely manner is a powerful aide to both student learning and institutional improvement.

The reconciliation of assessment and accountability requires a much more assertive academy in this matter lest its many constituents decide to take the matter exclusively into their own hands. The Collegiate Learning Assessment project (CLA), that I will describe in a few minutes, provides a powerful example of how the academy can indeed provide sophisticated performance measures of student learning in ways that might well suit both those outside the academy calling for accountability and those inside the academy who rightly care about academic freedom and being able to measure important learning. In this sense, the CLA project may help that reconciliation.

The Nature of Accountability

Accountability is multi-faceted. Students, faculty, administrators, Boards of Trustees, and legislatures all have a role to play but certainly all would agree that each campus has an obligation to be held accountable. That is, each institution has an educational, fiscal, legal, and moral obligation to be transparent—to demonstrate publicly to themselves and to their constituencies that the promises made to students, parents, funding agencies, and the larger body politic are being kept.

The primary function of accountability should be to optimize educational efficacy—that assessment and the reporting of the results serve the educational purposes of higher education. Further, we need to acknowledge that educational efficacy, as measured primarily but not only by student learning, is the single most salient standard of accountability. This standard, while necessary, is not sufficient, because it is quite legitimate to ask about the costs for such value that goes to the heart of fiscal and

political accountability. But the point I wish to emphasize here is that it is impossible to properly assess the educational value of an institution unless one has something concrete and causal with which to examine its cumulative impact. “Inputs” such as entering SAT scores of students, selectivity of admissions, endowment per student, and number of books in the library, while important, are not sufficient measures. Nor does the graduation rate tell us very much about what is actually being learned. Student learning should be the central component of any effort to measure the quality of an institution or program.

Value Added: The Most Valid Assessment Measure

Virtually everyone who has thought carefully about the question of assessing quality in higher education agrees that “value added” is the most valid approach. Excellence and quality should be determined by the degree to which an institution develops the abilities of its students. By “value added” is meant the value that is added to students’ capabilities and knowledge as a consequence of their education at a particular college or university. Measuring such value requires assessing what students know and can do as they begin college and assessing them again during and after (including years beyond graduation) they have had the full benefit of their college education. Value added is thus the *difference* between the measures of students’ attainments as they enter college and measures of their attainments when they complete college. Value added is the difference a college makes in their education.

Value added assessment is appropriate for the variety of higher education institutional missions. Community colleges, which account for close to 40% of all undergraduate enrollment, as well as colleges and universities, clearly need to determine the effectiveness of their programs. We know, too, that more and more students begin in one institution and finish in another and that the problems of appropriate program placement and transfer credits within and/or across institutions can be ameliorated with better and more transparent assessment data. Moreover, with increased institutional efficacy we might also see fewer students feeling the necessity to transfer. Within an institution, value added assessment to provide diagnostic feedback to both students and faculty within single courses, programs, and majors, catalyzes improvement efforts. Finally, with the advent of distance education and its myriad educational promises, value added assessment may have its greatest use since the demonstration of competence, rather than the simple accretion of credit hours, will become the new academic currency.

Benefits of Value Added Quality Assessment

Value added learning assessment enables the continuous improvement of student learning, promotes institutional efficacy, and can richly inform state policy. Most importantly, timely and appropriate assessment provides feedback to students to improve their learning in much the same way that doctors’ and coaches’ assessments help patients and athletes improve. In this sense, assessment should be an inextricable part of the teaching/learning process.

An assessment system can serve as a catalyst for continuous institutional improvement. Some students do not learn because they have not been responsible and assessment will have obvious consequences for failed student effort. On the other hand, if assessment shows large numbers of students not doing as well as expected, there is a faculty and institutional responsibility to make changes in courses, programs, and teaching. Moreover, the development of effective measures of the value added to student performance would create a new metric for the performance of departments and instructors that could be added to or replace research productivity as the criterion most often used to evaluate faculty performance. It is worth noting here that only 100 institutions out of more than 3000 account for 90% of federal research dollars!

From a state perspective, value added assessment data can provide the basis for continuous improvement of state policy intended to enhance higher education. Policies to provide smaller classes, technology for teaching, or more effective advising, for example, could be judged by their direct impact on students

In sum, value added assessment of student learning can have enormous educational value in that it can help:

- 1) faculty and their students make better sense of the teaching and learning in which they are mutually engaged;
- 2) institutions of higher education measure the cumulative impact of their curricular programming;
- 3) higher education generally by providing benchmark data for comparisons by sector (e.g. community colleges, liberal arts colleges, research universities, on-line instruction), as well as creating the potential basis for an incentive system focused on student learning;
- 4) public policy decision-makers concerned with issues of access, quality, cost, accountability, and equity;
- 5) students make better decisions regarding selection of appropriate colleges and universities, rather than relying on current incomplete ranking systems.

As we hear the chant for “learning organizations” in the 21st century, as we elevate the need to educate for an information-driven society, and as we are asked to make wiser use of scarce resources, value added quality assessment has many benefits.

Institutional and Cultural Barriers to Value Added Assessment of Quality

The primary initiatives for comprehensive, coherent, and ongoing assessment of education quality have come mostly from outside colleges and universities by state and local boards of education, corporations, state legislatures, governors, and market-oriented online educators. States have garnered the most headlines in this regard with their K-12 school reform priorities, explicit state-wide standards, and so called “high-stakes testing.” Such assessment, however, is difficult. It requires political and educational consensus about what is worth learning, developing valid and reliable assessment measures, constructing efficacious curricula, improving teacher education, providing appropriate

reward and incentive systems, and offering the financial resources and time for the development and sustaining of a comprehensive, systemic assessment program. These are equally salient issues for higher education whose history and culture make it especially resistant.

Higher education leaders offer a number of reasons to deflect a more rigorous assessment of teaching effectiveness and student learning. First, the academy has observed the problems States are having with assessment of K-12: the tendency to reduce testing to what is easily measured; inappropriate coaching or even cheating on the part of teachers and schools; narrowing the curriculum to just what is tested; and confusing assessment designed for diagnostic purposes with the politics and economics of holding individual schools accountable. These are serious issues and have reinforced the usual questioning by higher education of the value of the entire assessment enterprise.

Second, the culture of higher education is based on a number of assumptions conflicting with the notion of assessment. Teaching is understood to be a very human, person-to-person enterprise. “We teach people, not produce things,” is the refrain. Education is by its very nature, academics say, an inefficient enterprise, not amenable to the measures employed by, for example, the corporate sector. Indeed, the academic culture is one in which *systemic* assessment of undergraduate learning is neither practiced nor valued. Professors are trained to be scholars rather than teachers, with their scholarly disciplines and professional communities given a higher priority.

Third, it is believed that teaching cannot be properly evaluated, certainly not by the ubiquitous use of student evaluations of courses and professors that primarily measure affective – not cognitive – dimensions of learning. Moreover, it is argued, that what is truly worth learning cannot be measured at all or certainly not during the college years, since the real effects of a college or university education are evident far beyond the undergraduate years.

Fourth, a prior history of intermittent but inappropriate federal and state administrative intrusion into curriculum raises a legitimate concern for invasion of “academic freedom.” For example, the attempted federal directive on accountability that was to require the creation of State Post-Secondary Review Entities (SPRE) in all 50 states was strongly rejected by the states. And recently, the New York Board of Regents mandated a core curriculum for the State University of New York. Assessment beyond individual course grading, say professors, is just the first slide down the slippery hill of external intrusion. In short, both the realistic demands of efficacious assessment and the culture of higher education pose formidable barriers to a value added approach.

Implications of Value Added Assessment for Private Institutions and State Policy and Practice

We are virtually the only country with both private and public higher education systems and the issue of learning assessment and accountability are relevant to both albeit in different ways. In spite of the barriers and legitimate concerns noted, assessing quality in a serious, comprehensive, ongoing, and systematic way can have profound effects on educational practice within and across institutions as well as on state policies and practices.

First and foremost, value added assessment should have as its goal continuous improvement of curricula, pedagogy, admissions and retention. For both private and public institutions this is equally true. And certainly the higher direct costs borne by parents and students at private institutions demands the best educational value for the money. So too the same pressure for public institutions but the governance system in which the State has legitimate authority to make policy affecting its institutions adds another interesting dimension. Here I argue that State policies regarding funding of higher education, incentive structures for innovation, differentiation of mission among the various levels of higher education, and assessing the comparative advantages among state educational institutions, for example, should when possible be measured against whether or not they promote or inhibit teaching and learning.

The problem is that each educational policy issue, such as access, retention, true costs of instruction, and quality – whether debated inside or outside the academy – is too often treated in isolation. Few are, in the parlance of the day, connecting the dots. Policies to increase inexpensive access to higher education, for example, cannot be divorced from the issue financial aid and what students learn once access and affordability are achieved. Access is a hollow promise, indeed, if poor educational programming and teaching quality results in low retention. But only by assessing student learning will we have appropriate data to answer these vexing concerns. This also raises the important question of what constitutes appropriate readiness for college work but I save that issue for another day. On the other hand, assessment of student learning requires clear expectations, something higher education can improve upon that in turn would help make for a better connection between K-12 and colleges.

Policies intended to ensure quality, productivity, and accountability must each be informed by a common metric— student learning best serves that purpose. This requires value added assessment. Anything short of this systemic approach will leave academic and or political leaders without a basis for determining the effectiveness of their own policies, costs, or benefits. At the same time, let me emphasize that the culture of higher education is unique. It is just not sufficient to import from industry the rhetoric of assessment and efficiency. The nature of teaching, learning, and scholarship, in the context of college and university cultures, requires an assessment system designed specifically for those environments. Moreover, public and private Boards of Trustees must understand that there are real investments required to provide the time, energy, and resources necessary for such an endeavor.

If governors and other state policymakers attempt to use the gross, surrogate measures now mandated by a number of state legislatures and higher education governing boards – such as graduation rates, time to degree, or simple percentages of 18- to 34-year-olds in post secondary education – they will gloss over the fundamental causes of student and institutional success or failure. The key questions are clear. What institutional and system attributes support student success? What costs, even if they are higher than one would prefer, are justified because they enable professors to be better teachers and more students to succeed at higher levels of learning? Which state incentives and rewards best result in the kinds of institutional priority setting, cost containment, and innovations that result in increased student success? Assessment of student learning over time is the key variable in answering these questions.

Cumulative Learning and Institutional Accountability

While assessment of student learning has as its primary purpose helping individual students learn, such assessment can be of unique importance for educational and public policy makers if student-learning data are aggregated at the institution level. How well does a particular college or university fare with regard to the *cumulative learning* of its students over a typical four or five year undergraduate program? For public accountability it is important to find out how well the institution is doing in terms of learning gains students make as well as a measure of their absolute levels of competence. Moreover, such data not only allow for comparisons within the institution over time regarding changes in admissions, curricula, etc. but also for comparisons with similar institutions who choose to use the same value-added measures.

While assessment in individual courses by faculty is very important, it is not enough because this additive approach does not fully account for cumulative effects of instruction and work outside the classroom. The holistic effect of an entire institution's culture, ethos, curricula, and pedagogy can be measured—the whole is indeed greater than the sum of its parts. In fact, institutions do matter and they vary in the degree to which they add value to student growth. It is important to understand how they vary and then explain why this is the case. Such answers will aid the efforts by faculty and administrators to improve institutional efficacy.

Comparisons across similar institutions in large enough samples will allow for the setting of quality benchmarks to be used by public and private institutions themselves, parents, accrediting agencies, and state policy makers. Prospective students and their parents will monitor and evaluate the degree to which individual institutions are exceeding benchmarks. Their conclusions will result in fewer or greater numbers of students attending the institution. Policy makers at the state level will be able to determine the impact of greater or lesser funding.

Individual examples of powerful assessments of learning exist within and across higher education institutions in America. But only a few have constructed an institution-wide value added approach and there is only one initiative, the Collegiate Learning Assessment project that is developing a nationwide approach to value added assessment.

The CLA project attempts to incorporate each of the functions and assessment principles I have just described.

A Model for Higher Education Value-Assessment of Quality: The Collegiate Learning Assessment Project (taken from “A New Field of Dreams,” PEER REVIEW, Summer, 2003

Consider the chart I have here on the overhead. (see attachment). In our pilot-study, now duplicated by other researchers in the past few months, we tested students in 14 different colleges and universities to see if the institutional effects were greater or lesser than predicted in terms of our learning outcome measures assessing critical thinking, problem solving, and writing. Note that some institutions are on the regression line but that others did far better or worse given the nature of whom the students were when they entered. Clearly you can see it does make a difference where one attends school. Some campuses do add more value than others! How did we arrive at this conclusion?

In the fall of 2000, the RAND Corporation's Council for Aid to Education (CAE) began a national Value Added Assessment Initiative, a long-term project to assess the quality of undergraduate education in America. With beginning funding by a consortium of major foundations, the overall goal of the Initiative was to develop an assessment system that measures the impact of undergraduate education. The Initiative involves the continuum of higher education from community colleges to doctoral-degree-granting private and state colleges and universities. The objective was to serve as a model and incentive for continuous improvement of higher education and create measures of quality that all the major stakeholders – university administrators, faculty, students, parents, employers and policymakers – could use to evaluate the quality of academic programs nationwide. The Initiative has now been retitled, Collegiate Learning Assessment (CLA).

The CLA project differs from most other approaches to student assessment in four ways. First, it uses direct measures of student learning rather than proxies for it; typical proxies include input or actuarial data (e.g., entrance examination scores or faculty salaries), student self-assessments of growth, or college faculty and administrator opinion surveys (e.g., the *US News & World Report* rankings). As we have reported elsewhere,¹ there are methodological concerns in interpreting such indirect measures. Although the CLA project does not dismiss input² or actuarial measures, which provide valuable information about a college or university, it recognizes that these measures do not focus explicitly on skills and abilities colleges and universities are committed to developing. Therefore,

¹ See Chun, Marc. 2002. Looking where the light is better. *Peer Review* 4:2/3, 16-25.

² For example, SAT-I scores of entering freshmen purportedly provide information about the general intellectual ability of these students. SAT-II and ACT scores reflect a combination of achievement (i.e., what they learned in high school) and general intellectual ability.

performance measures of actual learning are an important addition to existing approaches to assessment.

Second, the CLA project focuses not on discipline-specific content; instead it focuses on general education skills—critical thinking, analytic reasoning, and written communication. The measures are all open-ended rather than multiple-choice.

Third, the project uses a “matrix-sampling” approach to assessment. The traditional approach, which would be to administer an entire battery of instruments to all students, is too time-consuming to be practical. Instead, the sampling design involves administering separate components of the full set to different (but randomly selected) sub-samples of students, therefore minimizing the time required per student, yet still allowing complete coverage of the range of instruments and content areas. This matrix-sampling design provides comprehensive and reliable information about how well a *school's* students are doing as a group rather than about the proficiency levels of any individual student.

Fourth, the project was designed to assess value added, or the **institutional contribution** to student learning. We do this in two ways: (1) we measure how well an institution's students perform relative to “similarly situated” students (defined in terms of their SAT or ACT scores)³ at other institutions and (2) we measure how much students' skills improve during their tenure at the institution through a pre-test/post-test model. As the research continues, we will also consider establishing baseline benchmarks against which institutions can evaluate basic skill development.

Why Focus on Assessing General Education Skills?

There are three related rationales behind the focus on assessing general education skills. First, most colleges and universities highlight general education as part of their undergraduate curricula. These are seen as the knowledge, skills, behaviors, and attitudes characteristic of an “educated person.” These general education skills—such as critical thinking, analytic reasoning, and written communication⁴—cut across academic disciplines and departments. Although any given college or university may adopt different pedagogical approaches to develop such skills, they nonetheless all share an overall commitment to these dimensions of learning and assessment. However, there are limited tools available to permit systematic evaluations of how institutions are doing in reaching their general education goals. The CLA project, therefore, seeks to contribute to the overall assessment efforts by contributing new instruments and a method that reflects the value placed on general education.

Second, whereas it is common to assess outcomes of individual courses, we believe that general education is not so neatly compartmentalized. It is rather the sum total of the combination of courses a student takes, plus the learning that occurs “between” and

³The feasibility study results for each institution were reported back to that institution only.

⁴ It is important to note that this list is not exhaustive; there are other dimensions to general education. See Shavelson, Richard J. and Leta Huang. 2003. Responding responsibly to the frenzy to assess learning in higher education. *Change* 35:1, 11-18.

outside courses, that contributes to overall skill development. As a result, the focus on the institution as the unit of analysis is motivated by an interest in understanding the overall impact of the college or university as a whole. This, we argue, is a more holistic way to understand general education.

Third, whereas discipline-specific measures focus on content, and some instruments might assess the ability to recall facts or formulas, the CLA project measures students' demonstrated ability to **use** information. Focusing on general education skills makes possible institutional comparisons, both within sectors (e.g., Carnegie Classification) as well as across the system of higher education as a whole. Again, because nearly all institutions work to develop general education skills, the CLA project makes possible benchmarks and analyses across type, such as between research universities and liberal arts colleges, or between historically black colleges and large public colleges. Even despite the differentiated missions characteristic of the higher education system, assessing the common elements helps us to avoid some of the pitfalls of comparing apples with oranges. Moreover, the CLA project does not prescribe any particular approach for developing such skills but, instead, makes possible research to allow institutions to make relative comparisons about how different programmatic or pedagogical designs work to promote student learning in general education areas.

Can These Skills Be Assessed?

Two different sets of performance measures were administered during the feasibility study. One set consisted of six performance tasks. The tasks measure a student's ability, for example, to read a table of data, make sense of a literature review, analyze an interview transcript, and review a newspaper report, and then to weigh the relative value of each document, synthesize the material, and prepare a cogent response to a question. These tasks, which take ninety minutes each to complete, are set in various contexts, such as science, social science, and arts and humanities. We used four of the "Tasks in Critical Thinking" (developed by the New Jersey Department of Education) and two CLA performance measures specifically developed for the project.

The second set of measures consisted of the two kinds of Analytical Writing Measures that are now part of the Graduate Record Examination (GRE). The forty-five-minute "Present Your Perspective on an Argument" type prompts students to state an opinion and provide supporting reasons and examples on a given topic; the thirty-minute "Analyze an Argument" prompts students to critique an argument by discussing how well-reasoned they find it.

Student responses can be graded by a trained reader or by a computer.⁵ There was a 0.50 correlation between a student's college GPA and scores on the CLA measures. This

⁵Analysis of the feasibility study data found that readers agreed highly with one another in assigning scores (median inter-rater correlation = 0.85). We also found that scores assigned by the computer to a student's answer to a pair of GRE essay prompts correlated highly with the scores assigned to those same answers by a human reader ($r = 0.78$).

correlation was substantially higher (0.65) when corrected for the less-than-perfect reliability of the measures. The corrected coefficient (which uses the institution as the unit of analysis) provides a more relevant indicator (than would student-level measures) of the degree to which the CLA measures tap skills that schools value (as reflected by the students' grades).

We also asked students to complete a task evaluation form. Their responses to the questionnaire indicated that they felt the time limits were generally more than adequate, that the tasks were engaging and authentic, and that the measures tapped skills that college students should be able to perform.

Can the Institution Be the Unit of Analysis?

The CLA performance measures we used were not designed to assess the same construct or provide scores that would be reported for individual students. Instead, a combination of measures was used from different clusters of academic disciplines. We would not expect that a measure set in a science context would necessarily correlate especially highly with one in the arts or humanities⁶, but the combination of measures across disciplines would provide a more robust measure of the institution's contribution to overall student learning.

How Can Value Added Be Assessed?

We explored "value added" of the college experience by analyzing both within- and between-school effects. The within-school effects analysis found that, after controlling on the students' SAT scores, upperclass students (seniors and juniors) tended to earn higher scores on our measures than did underclass students. This suggests that the measures capture institutional effects (recognizing that learning occurs both in and out of the classroom).⁷ The correlation between years in school and test scores was statistically significant. A school's average score on the CLA measures also correlated highly with the school's average SAT score ($r = 0.90$), yet we found statistically significant institutional effects after controlling on SAT.⁸ The between-school effects analysis examined whether the students at some schools were, on average, scoring higher or lower than would be expected⁹ on the basis of their mean SAT scores. Thus, the amount of education a student receives is related to the kinds of skills we assessed, and these

⁶ The mean internal consistency (coefficient alpha) for the CLA performance measures was 0.75, but the mean correlation between any two was 0.42.

⁷ This is notable because previous longitudinal and cross-sectional studies that utilized multiple-choice indicators have not found any such systematic differences. Still, an issue that faces all educational assessment is the difficulty in parsing out the direct educational contribution of a particular institution (as separate from general skill development and learning that theoretically might have happened irrespective of which college or university a student attends) or even learning that might have happened if the student instead hadn't attended college (also called maturation effects). Further complicating this matter is that 60 percent of students attend more than one institution while pursuing their undergraduate educations. We will refine our matrix sampling and methodological strategy to take these concerns into account.

⁸ With a sample size of 100 students per school, and with SAT scores explaining more than 80 percent of the variance, institutional effects were still detected.

⁹ Operationalized as more than two standard errors relative to the campus' spread of scores.

relationships transcend the abilities tested by college entrance exams. We use this approach as a means to quantify “value added.” Future research will also be longitudinal; rather than using SAT and ACT scores to predict value-added, pre and post-tests will be administered over the four or five years a student is in college.

Can Such an Assessment Be Done Economically?

The assessment can be done in a cost-effective manner and within a relatively short time frame. We found that a three-hour test battery consisting of one CLA performance measure (which takes ninety minutes) and two GRE measures (which together take seventy-five minutes) provides a sufficiently reliable and valid total score for assessing between-school effects. We also found that it is possible to calibrate the scores on different tasks to a common scale and, with the matrix sampling approach, to expand the range of measures used. In the future, we plan to administer the measures over the Internet, which will substantially reduce costs and increase the number of institutions that can participate in the assessment activities. We are also investigating ways to use machine scoring of performance tasks that will be as accurate as human scoring.

Will Schools Teach to the Test?

There is nothing wrong with teaching to the test, if test performance demonstrates skills or abilities that are valued. This is analogous to intentionally teaching student pilots how to land an airplane in a cross-wind because the final pilot’s exam involves performing that task; there is inherent value in teaching to this test. Hence, we would encourage schools to teach to the test, if that activity involved working with students to develop their analytic reasoning and writing skills¹⁰ and developing skills that students will need to demonstrate but still have value outside of the testing situation. In fact, we recognize that if an assessment approach does not reflect educational goals that faculty support, it inevitably will fail. Thus, the measures have been designed specifically to address some of the common elements that cut across higher education sector and academic field and that we believe faculty will endorse.

Will Students Participate?

As with all approaches to assessment, student motivation is a key issue. Because there are no high-stakes consequences at the individual student level, there must be another set of incentives to encourage students to participate and be motivated to do well on the measures. By participating, students will be able to receive an individual score (calculated as the mean score of the two GRE Analytical Writing Measures and one CLA performance measure). In addition, students can be provided with a CLA Certificate of Participation, which they can note on their resumes and which could be rewarded by their institution. Also possible are institutional incentives, such as framing participation as an element of school pride and responsibility and suggesting that students will want to do

¹⁰ Of course, teaching to the test should not include practice with the exact performance measures that will be used.

well so that their college or university will receive better information to improve curricular offerings.

Will Institutions Participate?

From the inception of the project we knew that the question of institutional participation would be one of the greatest challenges. However, given the realization that the measures are ready to be used and the subsequent interest, CAE has created a non-profit service that will allow institutions to pay a nominal fee to use the measures. **To date we have close to 75 institutions signed up to participate with us starting in September 2004.**

Many colleges across the country will soon use, or have expressed interest in using, our approach to higher education assessment. We have found that their reasons for doing so differ markedly. Some would like to use our measures as benchmarks for their own or other assessment measures. Some want to use them to monitor overall student progress within their institution over time, while others want to see how well their students are doing relative to those of comparable ability at other institutions.

We will continue with our research project by conducting a longitudinal study that will follow freshmen through to graduation at approximately fifty institutions. This will provide a rigorous basis to address important questions such as the relative merits of smaller, liberal arts colleges versus institutions with other instructional formats. Because this research also will include a cross-sectional component that involves testing at all class levels in the first year of the study, we should be able to learn a great deal by the end of the second year of the study.

Lessons Learned

So, what have we learned? Creating this assessment project has been quite a challenge. We sought to create an approach to assessment that is scientifically valid and reliable, that can be executed economically, that avoids the problems of teaching to specific test questions, that focuses on the value added of the institution, and that will be attractive for student and institutional participation. If you build it like that, they may come.

Implementation of the Value-Added Assessment Approach at the State Level

We now have three state systems of higher education that will be utilizing our measures this coming fall. Thus, we have also developed a model for states to consider as a process of phases in eventual implementation of value-added assessment as we have defined it. Much of this is equally applicable to private institutions as well.

As discussed above, it is crucial that colleges and universities take the leadership and control of a comprehensive student learning based assessment metric. We have found during the past several years in working with private and public campuses and state systems of higher education that the institutionalizing of CLA needs to take place in

phases if we are to overcome the usual academic barriers to assessment and accountability.

Phase I: Experimentation, Incentives, and Rewards

In this phase, value added measures need to be borrowed and/or developed and experimented with by faculty and administrators in a variety of colleges and universities within and across states. There is a wealth of material available but not widely known or shared. The goal in this phase is to establish an appreciation for a culture of evidence of learning within the institutions. Faculty and staff must gain confidence in the system of value added measures they develop that will provide them with the way to engage in continuous improvement of curriculum and pedagogy.

An example may be helpful here. One highly regarded private university recently convinced its faculty of the usefulness of assessing the effectiveness of its required freshman year core courses only to find that those students who took the courses in the fall fared no better than those who had yet to take them in the spring. Needless to say, such data surprised the faculty but also has led to significant questioning of the curriculum and pedagogy involved in that first year program.

Given the barriers described earlier it is crucial to any such endeavor, however, to provide appropriate incentives for the institution and the faculty, with significant rewards to encourage experimentation and develop trust.

Phase II: Development and Diffusion

Through a process of experimentation with alternative strategies, faculty, state policymakers and higher education leaders eventually reach a confidence level in the system of measures developed within institutions. These can then be widely shared among institutions – a diffusion of best practices phase. As the usefulness of the value added measures is established, some institutions will want to publicly record their progress and individual institutional benchmarks become articulated. This allows for a public conversation to begin about the standards and expectations appropriate to each institution's mission. Individual states may wish to establish commissions or committees under the aegis of their governors, legislative higher education committees or higher education coordinating boards, to lead these studies, establish benchmarks, and report progress. The key here is to develop a coalition of academic and state system partners.

Phase III: Comprehensive Assessment System Development and Implementation

In this phase, well-defined outcomes for general education and majors, performance measures of those outcomes, and a significant sample of students to be assessed are agreed to across a consortium of private institutions or the state system of higher

education. Data are collected and utilized by each institution for internal educational program improvement.

Phase IV: Value Added Data Used to Inform Institutional and State Policy

Once individual institutions are comfortable in the development and use of an assessment system for educational improvement purposes, it and the state where applicable, with value added data, can craft policies that better inform questions of quality, cost, accountability, and productivity. A number of productivity enhancements could be implemented, linked to an evaluation of the costs and benefits of the enhancements as they impact student learning. For example:

- 1) One could provide incentives for faculty and define and evaluate faculty productivity in terms of student learning as well as research performance. This is particularly relevant to non-research colleges and universities where research incentives are not primary.
- 2) One could begin to better evaluate mission differentiation among the state's colleges and universities on the basis of student learning as a variable in the reallocation of system resources.
- 3) One could compare the cost/effectiveness of distance learning compared to similar content taught on campus.

At the risk of repetition, we note a final caution. Assessment of value added requires a radical cultural shift within higher education, a great deal of time, effort, cooperation, risk-taking, and funding. It takes more time, more skill, more trust at all levels, and more safeguards than are currently extant. It is, however, an investment with potential large payoff because, for the first time, many proposed changes would be evaluated against their positive or negative impact on student learning.

Conclusion

I have tried to make the following argument. Assessment is crucial for teaching and learning, for institutional improvement, and for public accountability. Currently there are no common or widely used assessment measures other than student surveys or the ranking of "input" variables and retention. We would be best served if we in higher education would assess student outcomes and better yet if we developed a strategy to measure value added in terms of student learning during and over the entire undergraduate experience. This in turn would help us provide the best kind of evidence to those people legitimately concerned with accountability.

Finally, I have described the CLA as one example that all this is not just theory but is in fact possible. I hope many of your campuses would be willing to join our endeavor and I now welcome your comments and questions.

SOURCES THAT INFORMED THIS PRESENTATION

Benjamin, Roger and Chun, Marc. "A New Field of Dreams: The Collegiate Learning Assessment Project," in PEER REVIEW, Summer, 2003

Benjamin, Roger and Hersh, Richard H. "Measuring the Difference College Makes: The Rand/CAE Value Added Initiative," in PEER REVIEW, winter/spring, 2002

Chun, Marc. "Looking Where the Light is Better: A Review of the Literature on Assessing Higher Education Quality," in PEER REVIEW, winter/spring, 2002

Hersh, Richard H. and Benjamin, Roger. "Assessing Selected Liberal Education Outcomes: A New Approach," in PEER REVIEW, winter/spring, 2002

Hersh, Richard H. and Benjamin, Roger. "Assessing the Quality of Student Learning: An Imperative for State Policy and Practice," in HIGHER EXPECTATIONS: Essays on the Future of Postsecondary Education. National Governors Association, Washington, D.C. 2001.

Klein, Stephen, et.al "An Approach to Measuring Cognitive Outcomes Across Higher Education Institutions," in press, RESEARCH in HIGHER EDUCATION.

In addition, there are a number of other articles in the PEER REVIEW winter/spring, 2002 issue that are worth perusing including four "responses" to the essence of our value-added approach. See articles there by Carol Schneider, ("Can Value Added Assessment Raise the Level of Student Accomplishment?") Peter Ewell, ("Keeping the Value in "Value Added") Michael Strada, ("The VAAI and Higher Education Quality Control") and Michael McPherson, ("Eyes Wide Open: A Look at the Risks")