The Contribution of the Collegiate Learning Assessment To Teaching and Learning[1]

Roger Benjamin

Council for Aid to Education (CAE)

May 4, 2008

Introduction and the Steps in the Argument

The goal of the Collegiate Learning Assessment (CLA) is to improve teaching and learning in undergraduate education. The strategy used is to introduce performance assessment to higher education, a different testing paradigm that gives promise of being better aligned with an emerging reform agenda devoted to improving higher order skills. The CLA approach emphasizes the importance of using the institution itself as the initial unit of analysis, permitting faculty and administrators at an individual institution to understand where their institution stands in improving these skills compared to similarly situation institutions. There are up to three standard deviation differences between similarly situated institutions testing with the CLA (CAE, 2007), thus there is a large canvas of differences in best practices to be investigated that produce better student learning results. Of course the CLA is an assessment instrument but it is also much more than that. The performance assessments are designed to be direct contributions to teaching and learning in the classroom. The CLA protocol also includes a diagnostic logic template that suggests how faculty might drill down to the classroom to put in place changes that improve teaching and learning with respect to higher order skills.

The paper begins by indicating why performance assessment is important to higher education. Second, the rationale for comparative-based assessments of institutions as the initial unit of analysis is presented including a discussion of the rationale for the value added approach to this assessment. Third, the CLA template for continuous improvement of learning is presented followed by examples. This paper focuses on the CLA strategy. A recent publication (Klein et al., 2007) presents the technical arguments

[1] I thank Richard Hersh, James Hundley, Chris Jackson, Heather Kugelmass, and Richard Shavelson for comments on earlier versions of this manuscript.

underlying the CLA protocol, including the value added approach. That paper also demonstrates that the CLA performance assessments are highly reliable and valid in response to criticisms (Banta and Pike, 2006 and Kuh, 2006).[2]

The Current Assessment Regime Is No Longer Sufficient

Multiple-choice and short-answer tests remain the dominant testing regime, not only for facts, but also for higher-order skills. For example, multiple-choice tests may present examples of correlations and causation and then ask students to identify whether or not each is correctly or incorrectly applied. However, responding to such choices passively is very different from asking students, in performance assessments, to actively critique a case study presenting an argument about data in which correlation or causation are misused. In the latter approach, the student must not only recognize the mistake but also understand where and how the concepts are confused and explain why the argument fails.

Another critical shortcoming of the current assessment regime in education is that it pays little attention to how much a school or college contributes to the development of competencies students will need after graduation. For instance, the outcomes that are typically looked at by higher education accreditation teams, such as the college's retention and graduation rates and the percentage of its faculty in tenured positions, say nothing about how well the school develops its students' analytic reasoning, communication, and problem-solving skills. This situation is unfortunate in that the ways in which institutions are assessed affects what content and skills are taught, how they are taught, and what students learn.

Educational institutions are being challenged to do a better job educating tomorrow's workforce. A college education has never been more necessary for productive participation in society. Employers seek individuals able to think and communicate to meet the requirements of the new Knowledge Economy. This means that the skills taught in higher education are changing. There is less emphasis now on acquiring content knowledge and more on

[2] The Klein et al. (2007) article remains relevant to Banta's (2008) repetition of a number of technical criticisms of the CLA protocol Banta and Pike made in their 2006 article. This paper deals specifically with the argument that the institutional comparison strategy is not warranted, that ceiling effects prevent its applicability to selective institutions, and that the CLA is designed for accountability not assessment.

searching for information, using reasoning to solve problems, and using communication skills to convey this information.

Assessment has to catch up with this change. Performance assessment tools are necessary not only to evaluate whether students are learning the skills expected of them in today's workforce, but also to spur educational advances in teaching such skills. Performance assessments are realistic work samples, case studies, or real world scenarios presented to students for analysis. They differ from standard multiple choice or short answer tests in that the student is provided all the information they need---documents, tables, graphs, figures---to answer a set of questions that requires a reasoned response. There are no "right" or "wrong" answers; rather, the goal is to stimulate students to apply what they know to a problem and come up with a solid response.

As the U.S economy has evolved from the Industrial Era to the Knowledge Economy, it has become increasingly dependent on a workforce that can generate knowledge as a foundation for economic prosperity. Knowledge generation requires stronger skills in critical thinking, analytical reasoning, problem solving, and writing – referred to as higher-order skills. This development portends significant changes in teaching and learning as reflected in the educational reform movement now underway.

Although this reform is present in K-12 education, particularly in the college readiness and accountability debates, it is most advanced in higher education. In higher education, the reform movement can be characterized along three dimensions:

- Shift from the long-standing lecture format to a student-centered approach that emphasizes students' active class participation and analytic-based writing;
- Change in the balance of curricular and textbook focus from its current emphasis on content to case- and problem-based materials that require students to apply what they know to new situations;
- Change in assessment instruments from multiple-choice tests that are best used for benchmarking the level of content absorbed by students to open-ended essay tests that are aligned with the other goals of the reform process.

Although significant advances have been made in the first two dimensions of higher education reform, assessment has lagged behind: Most standardized testing is still based on multiple-choice and short-answer formats. As a result, in higher education the current testing regime is not assessing the most critical skills required of students in the workplace and—just as important—it is not supporting the other two dimensions of reform. This is where performance assessment comes in.

Why comparison is essential

What is the justification for comparison? There clearly have been benefits to comparison on the research side. Higher education adheres to a transparent system for the allocation and evaluation of research based on the principles of peer review.

Higher education leaders do develop comparative benchmarks for undergraduate education. They make a variety of comparisons about student learning, as well as other indicators, over time within their institution. They may also develop comparative yardsticks to understand how well they are doing on a variety of outcome measures compared to other institutions. The question is whether the current mix of indicators, locally developed within institutions, is good enough to provide effective assistance to faculty and administrators in their effort to improve teaching and learning at their institutions.

While there is no question they are critical for assessments within institutions, two often cited candidates for comparison—within major comparisons and portfolios—face formidable hurdles.

Majors form the basic building block for all educational content produced in the university. However, majors are not well suited for comparative assessment as there are too many majors and each institution exhibits its own combination of them. Faculty have difficulty agreeing on what is essential for a student to learn in a major. Cultural anthropologists and physical anthropologists, for example, focus on distinctly different areas of knowledge. Portfolios, the other candidate (see Banta, 2008), suffer from serious reliability and validity problems that prevent their use for comparative assessment (Koretz, et al. 1993).

Still, some form of inter-institutional comparison is particularly critical now because we remain at the formative stage in the development of useful theories of pedagogy and learning. If we had powerful theories of pedagogy and learning, we would not need to engage as much in searching for best practices, experimentation or comparative assessment itself. However, we do not have agreed upon methods of teaching, methods of assessment or even standards to which students should be held accountable. Under these conditions, comparativists argue it is critical to expand the universe of cases that might be compared with each other (Ragin 1989; Munk and Snyder, 2007). From this larger universe of cases the researcher can separate out outlier cases that warrant greater examination. Alternatively, the researcher can study in greater detail the characteristics of the most "normal" cases, the ones closest to the mean of all cases being analyzed.

Comparative-based tests should not and will not replace locally developed institutional assessments developed. In fact, no one test can measure everything considered important. Comparative-based assessments should be augmented with local assessments designed by the faculty or campus administrators. In other words, comparative assessments should be viewed as one of multiple instruments an institution uses to assess its progress in student learning. That said, only comparative measures can produce important signals to the faculty and administration about where they stand in comparison to other institutions.

However, in addition to providing informative about performance relative to other institutions, comparative-based measures should also be designed to assist faculty in the improvement of teaching and learning. Only when faculty are comfortable with assessment instruments will they entertain discussions about the use of results from such instruments for accountability purposes.

Finally, the results of comparative-based assessments should not be used to rank or rate colleges and universities publicly. Single dimension rankings mask the great diversity of outcomes in higher education and are therefore inappropriate and disingenuous. Moreover, the field of assessment in higher education is far from reaching maturity. Time is needed to experiment with measurement instruments and ways to implement results.

Why higher order skills should be a principal focus and the institution be the initial unit of analysis

Majors were clearly the most important venues for students during the industrial era when the transfer of content knowledge was thought to be the primary function of undergraduate education. However, while it remains necessary for students to master content, it is equally if not more important for students to master the higher order skills they will need to access, structure, and use information in today's knowledge economy (Pellegrino, et al., 2001). Employers have emphasized for some time that they are not as concerned about what students major in; rather they want the college graduates they hire to be proficient in higher order skills (Immerwahr, 2000). Majors account for less than one-third of the courses students take in a typical four-year program. Under our present department-based model, faculty have no way of knowing whether their courses---delivered within departments and isolated one from the other---in fact develop these greatly valued higher order skills.

The institution as a whole is the appropriate unit to begin evaluating the extent of success in improving higher order skills in its graduating students because these skills are accrued in a cumulative fashion as students progress through all their courses and other experiences at their institution. Moreover, virtually all institutions commit to improving these higher order skills in their mission statements. Higher order skills have the quality of public goods in the sense that they are produced by a number of suppliers, i.e., departments and programs throughout the institution. One may analyze the relative contribution of, say, specific majors to higher order skills. But the first goal is to determine the degree of improvement of higher order skills students reach in the first place. This is why the institution rather than the major is the initial unit of analysis.

However, this does not mean the department or the major are not important. Majors are an important consideration in any analysis in order to understand their contribution to the institution's value added score, and they remain critical for establishing the degree of content achieved by students. Moreover, it is important to drill down to the individual classroom to determine how best practices in teaching and learning can effect and improve value-added results at the institutional level (see below pp. 11-20). However, undergraduate education should be viewed as greater than the sum of its parts. In other words, it is arguably as if not more important to focus on the cumulative growth of higher order skills in students at an institution than only the transmittal of the content of knowledge.

Why the value-added approach is important

If the argument for comparison is accepted, the issue then is what and how to test. An important goal in education should be to learn how much our teaching leads students to improve? This is the value-added question, prominent in today's assessment and accountability debate. Let us consider two definitions of value-added. One focuses on improvement within an institution over time (e. g., seniors score higher than entering freshmen with similar SAT scores). The other focuses on the improvement between classes within an institution compared with that of other institutions.

If we are to use the value-added approach as a way to judge the contribution of the higher education institution to student learning, there are at least three arguments against this position that must be considered (Lombardi, 2006; Banta and Pike, 2006; Banta, 2008).

The first is that students who enter college well-prepared exit well-prepared, thus the contribution of college to this growth may be assumed to be minimal. If this were the case, controlling for the resources these students bring to the institution is not worthwhile. However, we have clearly seen that some institutions improve higher order skills as measured by the CLA scores of their graduating seniors compared to graduating seniors in similarly situated institutions (Klein et al. 2007).

A second argument is that there are ceiling effects on student learning growth in elite institutions, therefore there cannot be as much student learning growth in institutions that admit students with high ability as measured, for example, by the SAT or ACT. This suggests the value added approach penalizes these institutions when compared to institutions that admit students of lower ability.[3] In fact, the CLA has been designed to permit maximum variation in the quality of responses. Based on the performance of students at highly selective institutions participating in the CLA, the higher order skills of students at our nation's most selective

[3] Parenthetically, the reverse is the case now. Because colleges and universities do not widely use the value-added approach and do not measure student learning outcomes in a comparative manner, only institutions that admit students with high SAT/ACT scores are labeled to be of high quality.

institutions can use as much improvement as students who begin their academic careers at comparatively less selective colleges and universities.[4]

This, however, raises the third argument—that the value-added approach privileges institutions that demonstrate more value added growth over highly selective institutions that exhibit high standards as asserted in different ways. This criticism of would be valid if institutions could not set standards for minimum proficiency and low to high ranges in performance. One may, for example, establish ranges of scores on comparative assessments that provide institutions with examples of sample answers for a variety of scores along the performance spectrum.[5] Institutions can then set minimum cut off scores that are appropriate to their context.

To summarize, the main focus of comparisons should be made not between absolute levels of achievement but between the amount of value that colleges and universities add. However, this leads to two different and somewhat contradictory counter-arguments by those who object to the use of value-added comparisons at elite institutions. The first is that elite institutions should only be compared to their peer institutions that face the same challenges with similar students. The second argument is that elite institutions might want to compare the absolute level of student attainment they achieve to that of their peers since this is the scale they excel on. Both of these arguments can be accommodated in a voluntary system of assessment where institutions can choose what to focus on--- value-added or absolute levels of performance.

Evidence of the importance of institutional comparisons

One attribute usually and correctly cited in arguments asserting the superiority of American higher education is its diversity. But diversity is also used to argue that colleges and universities cannot be compared at all. Certainly, there is great diversity in the missions of four-year educational institutions (see Carnegie classification), ranging from liberal arts orientation to science and technology, to business, music and performing arts, to art and architecture. Moreover, institutions are public or private and non-profit or for-profit. These distinct missions clearly result in differing emphases, e.g., population served, configurations of majors and minors, types of core curriculum, and expectations of student performance. The implication is

[4] Even students with SAT scores at the 99 percentile score no more than 50 percent of what is possible on CLA performance tasks.
[5] As presented in Hardison and Vilamovska (2008).

that colleges and universities vary so much that it is not possible to compare them and that any such effort would be fatally reductionist, factoring out the rich diversity which makes American higher education so unique. To support this position some argue that *within* institutional variation on student learning is greater than *between* institutions (Pascarella and Terenzini, 2005; cf Kuh, 2007) and that therefore one should focus on internal comparisons within institutions and, in particular, concentrate on measuring important factors associated with student learning such as student engagement.

Pascarella and Terenzini's argument—that colleges and universities are more alike than their internal majors and programs are different—undercuts the diversity claim. It is thus important to examine the Pascarella and Terenzini findings more closely. Pascarella and Terenzini find within institutional changes on a number of key indicators, including critical thinking measures, to be greater than changes between institutions, thus directly contradicting the diversity argument. The critical thinking and student learning content measures they review range up to 1.2 standard deviation of growth *within* institutions (Pascarella and Terenzini, 2005, pp. 573-574). By comparison, none of the measures they review suggest statistically significant differences in change in student learning or other key indicators *between* institutions.

Pascarella and Terenzini's review of the literature ends at the conclusion of the 1990s. They did not have available evidence from the CLA. The CLA corroborates the finding that there is substantial growth within institutions. There is well over 1.0 standard deviation value-added growth in CLA scores within institutions (estimated cross-sectionally). Recall, that the CLA measures higher order skills that almost all institutions commit to improving. Therefore, with respect to these skills, the diversity argument does not come into play.

Let us turn then to comparative, between institutional findings. When the CLA compares the value-added growth of institutions while controlling for the SAT scores of the students sampled, there is up to 3 standard deviation differences in CLA results between institutions with similar SAT scores. This is a very large effect, much greater than within institution growth. This demonstrates that, in fact, it does matter where one goes to college. There are significant differences between colleges at least with respect to their ability to improve higher order skills. The variance therefore suggests that there are interesting best practices to be studied in colleges doing better than

expected, best practices that can be adapted for use in all colleges focused on improvement of teaching and learning.

Finally, the subject of whether and how information about higher education institutions should be reported publically needs a great deal of deliberation (Fung, et al., 2007). We may be entering a period when we have comparative measurement instruments of sufficient quality to use for assessment and accountability purposes but we are only in the beginning stages of serious discussion of their public policy implications. The National Association of State Universities and Land Grant Colleges (NASULGC) and the American Association of State Colleges and Universities (AASCU) have called for greater transparency and accountability and a common template that institutions will use to collect information to be reported publically. But raw scores on assessment measures may vary from one year to the next (Klein et al.), low graduation rates may signal that an institution is doing its job in recruiting at risk students, etc. How should these and related issues be sorted out? Moreover, arguments can be made for current national efforts comparing higher education institutions across states can be made but the process is not as straightforward as one might think (Benjamin and Klein, 2006).

Even if the equivalent of the Bush report (Bush, 1945) for assessment and accountability in higher education were written today it would not be adopted. Clarity about the nature of the problem is not equivalent to the need for a research policy that Vanevar Bush dealt with. Nor are the solutions as clear-cut. Bush and his colleagues had only to continue to support the peer review system already in place in American higher education. But as institutions, systems of higher education, and states devise templates of assessment and accountability measures that produce results that they are comfortable with, they will publicize them. And once a number of successful examples are made public, many other examples will follow. This is the historical pattern in the American society and economy.

Implications

Today, comparative-based assessment is a key ingredient to gaining knowledge about best practices in teaching and learning available for institutions to adopt. We need to compare institutions to establish the extent to which they are diverse and similar and to then share best practices for improvement in teaching and learning. Moreover, we should engage in the kind of layered comparison strategy described here to improve the quality of

teaching and learning in higher education. If we did not have the ability to do the kind of comparative-based assessment presented here, the status quo might be acceptable. But we now can do sophisticated, meaningful comparisons and, since we can, there is no legitimate argument against doing so.

A basic question, then, is whether higher education institutions, public or private and across the Carnegie Classification categories, should participate in testing with comparative-based assessments before all the rules of engagement about interpretation and use of the results are worked out. These rules of engagement will be worked out through trial and error led by national associations, systems of higher education, states, and individual institutions over the next several years. There is no other practical way to accomplish this. It is therefore important that higher education institutions and their leaders all be involved.

Colleges and universities exhibit considerable diversity in producing the higher order skills they pursue in common. This presents a rich common ground for faculty and administrators to identify, examine, and compare best practices in teaching and learning. This real world of educational experimentation, evaluation, and improvement, facilitated by institutional comparisons, gives promise of providing educators information critical to the improvement of theories of pedagogy and learning which, in the end, are the ultimate goals all share in higher education.

A Template: How the CLA Can Contribute to the Improvement of Teaching and Learning in Higher Education

Of course the CLA can be viewed as an assessment instrument. Equally, however, it can be viewed as an instrument for the reform of teaching and learning in higher education. It is important to present the diagnostic logic chain that links and institution's CLA scores to the class room, providing examples about what this means since the word assessment places the CLA in a box occupied by many other assessment tools, including multiple-choice tests. When examined for its contributions to teaching and learning, the CLA is in a league of its own.

Here is a template that indicates how an institution might respond to the initial institutional level CLA scores followed by illustrations of how administrators and faculty are benefiting from using the CLA along with

other measures related to student learning. These illustrations are offered as early examples of productive uses of the CLA.

From the Institution to the Classroom: The CLA Comparison Strategy

1. The CLA's global institutional score is based on the average performance of the sample of freshmen and senior students taking the CLA. In reports to the institution, its score is presented in comparison to those of all other participating institutions. To account for variation in competencies the students bring to college, the CLA institutional scores are adjusted for the SAT scores of the participating students. The CLA score, then, reflects the amount of value-added improvement in performance between the freshman and the senior year graduating students, while controlling for incoming academic ability. When the scores of all institutions taking the CLA are placed in a regression equation, the institutions cluster along a straight line. More specifically, a college can be compared against the performance of colleges with similar average SAT scores. Any particular institution's value added may fall anywhere from the lowest ten percent to the highest ten percent on the plotted line at a particular level of SAT intake.

In a basic sense it does not matter where the institution falls the first time the institution tests. The comparison gives faculty and administrators a benchmark, a signal about where their institution stands. The question then is what should the faculty and administrators of institutions do to improve the degree of their value added. That leads to the following subsequent steps.

2. Correlate inputs, processes, and outputs. A logical next step is for the college's institutional research office to correlate the inputs and processes (or such proxies as class size, expenditures per pupil, incoming freshmen SAT scores, per student endowment expenditures, etc.) with outputs of undergraduate education (such as retention and graduation rates and, of course, CLA performance and other measures of learning). The goal here is to develop an efficient description of the factors that correlate with positive CLA results.

3. Conduct in-depth analysis. While the institutional score signals where an institution is placed when compared to all other institutions administering the CLA, administrators and faculty members will want to know more about the relative contributions to that score by colleges (if the institution is a university) or by certain departments or programs (if the institution is a

college). Which departments or programs, for example, are particularly strong or weak contributors to their CLA results?

4. Conduct audit of existing assessments. There is a saying in the assessment world that a curriculum is determined by what faculty test for. Thus it will be useful to understand the extent to which faculty are using multiple choice or essay tests in their classrooms. Are the tests given truly measuring what this paper has argued are important skills--such as critical thinking, problem solving or analytical reasoning—driven by the Knowledge Economy and expectations of employers? How well are the students doing on current tests?

5. Examine best practices found to produce good CLA results. Many colleges participating in the CLA are working together in consortia of similar institutions. They are highlighting and sharing best practices that are correlated with noteworthy CLA scores. For example, it appears that schools that require more analytic- based writing do better on the CLA than those that do not.

6. The most important step: get published CLA performance tasks into the hands of the faculty so that they can:

a. Use them in their classroom where they have greater knowledge of the strengths and weaknesses of their students;
b. Develop their own performance tasks based on the scoring guide of the published tasks;
c. Choose case studies and problems for text material that is congruent with the documents in the CLA performance tasks rather than the content dominated textbooks extant;
d. Adopt a student-centered approach to teaching that calls for much more analytic-based writing and diagnostic feedback to the student about how they can improve their performance.

In sum, the above steps comprise an early version of what we, at CAE, intend to become a reinforcing system of continuous improvement of teaching and learning.[6] The institution's global score provides a critical

[6] This is precisely what higher education has in the research realm. Through peer review research has a public face that encourages and requires researchers to respond to criticism and evaluate the claims of other researchers: in short, engage in a never ending process of continuous improvement. If we followed the above steps for undergraduate assessment,

signal that triggers an internal focus on what correlates with the score. While an institution's performance on the initial test administration serves as an important benchmarking mechanism, the important and essential follow-up questions become related to (a) understanding what led to those results and (b) deciding what improvement goals might make sense for the future.

Examples of CLA Use

Kalamazoo College Report by Paul Southerland.[7]

Results from the CLA and NSSE can be enlightening, challenging, and affirming. Trying to understand our students' CLA performance has led us to examine features of our curriculum that might bring about changes we see in students between matriculation and graduation. Through these analyses we are finding that at least some of our students' experiences seem to have a "value-added" effect, and we are beginning to discern how this effect might be expanded to reach more students.

Performance of Kalamazoo College Students on the CLA

Through a grant from the Teagle Foundation, and as part of an assessment collaboration with Colorado College and Earlham College, we administered the CLA to first-year students and seniors during the 2005–6 academic year. First-years had a mean performance at the 80th percentile (at the lower end of the "at expected" range) of the CLA, even though their mean SAT scores were at the 92nd percentile compared with first-years who took the CLA in 2005–6. Seniors had a mean performance at the 99th percentile (at the upper end of the "above expected" range) of the CLA, whereas their mean SAT scores were at the 92nd percentile compared with other seniors who took the CLA. The "value-added" (mean senior CLA score minus mean first-year CLA score) of a Kalamazoo College education was "well above expected."

At Kalamazoo College, CLA performance seems to vary with the academic division in which students majored. Adjusted CLA scores differed significantly among divisions, even though actual CLA scores did not, with students in natural sciences having the lowest Adjusted CLA. This observation is corroborated by the distribution of students among the three

---

we could hope to eventually also create a continuous system of improvement of teaching and learning.

[7] Excerpts from Southerland (2007)

performance categories. The natural sciences showed a bimodal distribution with eight "below expected," three "at expected," and eleven "above expected" scores, whereas all other divisions showed uni-modal distributions, with the vast majority of scores in the "at expected" and "above expected" ranges. The bimodal distribution in natural sciences led to hypotheses about causes for the "below expected" performance of some science majors and prompted us to examine NSSE results more closely.

Insights from Student Interviews

Interviews of Kalamazoo seniors provide additional information about effects of various educational experiences. Students in a qualitative research methods course administered, transcribed, and analyzed interviews of thirty-one seniors who took the CLA. Examining the interview transcripts from students with high CLA scores and students with low CLA scores revealed intriguing intergroup differences that corroborate insights gained from examining disaggregated CLA and NSSE scores. The following "patterns" emerged: foreign language proficiency seemed to correlate positively with CLA scores; students who used phrases like "personal initiative" generally did better on the CLA; and some science majors seemed to get "lost" in their major, but those who did explore other disciplines tended to do well on the CLA.

Preliminary Inferences

Clearly, a college education enhances critical thinking, analytical reasoning, and effective writing, and the trajectories students take through that education seem to affect the degree to which those abilities develop. Although small sample sizes preclude our reaching definitive conclusions about factors affecting CLA performance, at this point in our explorations we surmise the following: a high "value-added" education emphasizes all skills measured by the CLA and creates opportunities for students to experience, reflect on, and learn from "high-intensity dissonance." Analytical reasoning and critical thinking are essential for performing well on the CLA, but without effective writing students cannot fully demonstrate those skills.

Insights from Interviews with Faculty

Because of the importance of getting the CLA performance assessment into the hands of faculty in the classroom, CAE has launched CLA in the Classroom which includes CLA Faculty Academies. These academies are designed to teach faculty how to develop their own CLA type tasks. Here are reactions from faculty involved in an academy workshop from Cabrini College following the short course in performance assessments they took. The reader will see reactions regarding the difference between performance assessments and multiple choice or short answer assessments, scoring, the goals of the assessments, what was learned about pedagogy and learning.

(1) Because I'm a psychologist, one of my issues was scoring their answers, because it's qualitative. One of my issues would be inter-rater reliability. We sat around the room – myself and the two researchers that had brought the materials to me – and we went through each student's responses one-by-one and we scored it on our own, and then we came together to arrive at a consensus, discuss answers. But I felt like perhaps there was little disagreement as far as scoring the answers and that's hard because you try to be unbiased but they're my students, so I know how they perform in class. Maybe I know what they are trying to say because I have them in class, and so that bias affects how I would score their ratings, or rate their performance, versus non-biased parties that come in and don't know who these students are.

(2) They learned a lot about themselves, and I learned a lot about them, and how to better tailor the information I present in class so it's more effective. But they gravitated to more of the visual sources, for the most part, gravitated to the visual sources but it wasn't because they learn visually, it's because -- and they admitted this to me -- that they're lazy and it's easier to just look at a graph and take information from a graph than actually have to read a full page of text.

(3) I do want to say I was pleasantly surprised. When I receive the e-mail, I'm thinking, "How am I going to fit this into my course, I have so much material to cover …" but I was pleasantly surprised with it. I think one of the joys was being able to discuss it with [CLA Staff] and [the director of the campus' teaching and learning center], that was one of the joys, and two was the ability to discuss it with the students and to really learn how the students learn."

(4) I know we were presented with data at the faculty meetings – so we learned, "We're value-added." But I don't think I really appreciated what that meant until I saw the test. So I was, "OK, it's another assessment piece." I didn't really think much about it, until I had the opportunity to really see what we were talking about. I thought it was going to be another satisfaction-style survey or another content knowledge type of survey, and clearly it's not.

(5) I think one of the things that would have to happen if we're going to have a full adoption, I think we're really going to have to do for everyone what that CLA workshop this summer did for me. It was eye opening. So once I saw it, the other component that [CLA Staff] took us through, was how can you apply this template for the performance tasks to your own course. So we had the opportunity to break out into small groups of two or three faculty members and literally fill in the blanks. The template was that simple. And we learned how we had to develop certain graphs, certain tables, certain literature references, certain personal statements that might imply bias into a certain thing, and I came out with something that with very little modification I could actually use in my own class. So I think that if we could share that with everyone on campus, I would say that most people would be excited. You aren't going to get a lot of people who are going to say, "I can't use that. There's no way I could do it." It was very simple to do from a science perspective, especially with all the data tables, it seemed very simple for me to develop, but I saw people who were in business developing it, economics would have a really simple time developing the template, there were some people in art, and I thought they could never do it, and it was very, very easy for them. I think everybody at the end was just really excited. It could be applied to almost every topic that was in that room, very, very easily. I think it's just a matter of sharing it at the level that we had this summer.

(6) I got excited about the one that [CLA Staff] presented to me. I'm like, "OK, I can't use this." Once I did my own, I was, "This is so cool!" It was really incredible to see that I could do that so simply. And it would add value for my students. It wasn't just an add-on kind of thing that I did at the end of the semester.

(7) The outcome is somewhat indicative of how they went through getting there. And by observing what they did, it gave you an indication as to how much thinking was involved. In other words, in other tests I may give, and they can be long, higher-level type questions, they're just writing,

writing, writing, writing. In this response, they stop writing, and they had to process and think, so that told me that this kind of test is asking for something other than just pure output. There's a process that's involved, and I can observe that by observing them.

(8) One of the first responses they gave me that really made me think, was, "We never get tested like this." Of course I wanted to defend it, and say, "Oh, yes you do! Don't you remember in that particular (laughs)!" But I was very calm at that point, and I said, "Could you expand on that?" They said, "We're never really tested to think about or to analyze information. We're really tested to kind of let you know what we know." And I think they're right. So therefore their responses told me that this kind of assessment, they've never been exposed to before, not only in higher education, but also in K-12 education.

(9) If a faculty member is going to use this kind of assessment, they are going to rethink the way they teach because this type of assessment is asking for students to pull together various forms of information and to analyze it at a higher level. And therefore when you're teaching content, you cannot teach it just at a lower level; you've got to teach it to the point where students have a basic understanding, but then you prompt them to analyze and evaluate the content at a much higher level.

Other examples of CLA Use

1. Evaluation,
    - Of investments to make some arts and science departments more student-centered---Duke
    - To compare humanities, social sciences, sciences colleges---Harvard
    - Of the effectiveness of core curriculum programs---Alaska Pacific College, Seton Hill University, Stonehill College

2. Use of CLA to enhance work in the assessment of writing---Westminister College among others

3. Use of the CLA to provide external validation of internal positive

assertions about the quality of an institution's undergraduate quality---many private liberal arts colleges, among them Lynchburg College

4. To assist faculty to develop their own CLA-type performance assessments---CLA Faculty Academies---one per month

5. To assist faculty to think more in terms of learning outcomes than teaching objectives---Allegheny College

6. Comparison of CLA results of transfer students from community college to upper division---California State University

7. Use of CLA results in combination with multiple measures in accountability regime---University of Texas system of Higher Education, West Virgina system of Higher Education

References

Banta, Trudy, and Gary Pike. "Revisiting the Blind Alley of Value Added."(2006) Assessment Update. January-February, 1-5.

Banta, Trudy (2008). Editor's Notes, Assessment Update. Vol. 20, No. 2 March-April, 3-4;15-16.

Benjamin, Roger. "Recreating the Faculty Role in Governance in Research Universities," in Joseph C. Burke (ed.) (2007) Fixing the Fragmented University. Bolton, MA: Anker Publishing, 70-98.

Benjamin, Roger. and Stephen Klein (2007). "Assessment Versus Accountability in Higher Education: Notes for Reconciliation," UNESCO Forum Commission Paper. Paris, France: UNESCO, pp. 27.

Bush, Vanevar (1945). Science The Endless Frontier. Washington, D. C.: United States Government Printing Office.

CAE (Council for Aid to Education) (2007). Institutional Report. New York: CAE.

Fung, A., M. Graham, and D. Weil. (2007). Full Disclosure The Perils and Promise of Transparency. New York: Cambridge University Press.

Hardison, Chaitra, and Anna-Marie Vilamovska (2008). Critical Thinking Performance: Setting and Applying Standards For College-Level Performance. Project Memorandum, PM-2487-CAE, Santa Monica, CA: RAND Corporation

Immerwahr, John (2000). Great expectations: how the public and parents White, African-American and Hispanic view higher education. San José, CA: National Center for Higher Education Public Policy, August.

Klein, Stephen, Richard Shavelson, Roger Benjamin, and Roger Bolus (2007). "The Collegiate Learning Assessment: Facts and Fantasies." Evaluation Review vol. 31, no. 5, October, 415-439.

Koretz, Dan., Stecher, Brian., Klein Stephen., McCaffrey, Dan., and Deibert, E. Can Portfolios Assess Student Performance and Influence Instruction? The 1991-92 Vermont Experience (1993. RAND Institute on Education and Training and UCLA: CSE Technical Report 371, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA: . CRESST/UCLA.

Kuh, G. (2006). Director's Message in: Engged Learning: Fostering Success for All Students. Bloomington, Indiana: National Survey of Student Engagement.

Kuh, George (2007). "Risky Business," Change, September/October, 30-35.

Lombardi, John, Inside Higher Education (2006). August 2.

Munck, Gerardo L. and Richard Snyder (2007). Passion, Craft, and Method in Comparative Politics. Baltimore, MD: Johns Hopkins University Press.

Pascarella, E. & Terenzini, P. (2005) How College Affects Students: A Third Decade of Research. vol.2. San Francisco CA: Jossey-Bass.

Pellegrino, James. W., Norm Chudowsky, and Robert. Glaser (eds) (2001). Knowing What Students Know: The Science and Design of Educational Assessment. Washington D.C.: The National Academy Press.

Ragin, Charles (1989). The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. Berkeley, CA: University of California Press

Southerland, Paul, Anne. Dueweke, Kiran. Cunningham, and Robert. Grossman, —all of Kalamazoo College (2007). PEER REVIEW, 22, Association of American Colleges and Universities (Spring)