

MONOGRAPH

RETURNING TO LEARNING IN AN AGE OF ASSESSMENT:

A SYNOPSIS OF THE ARGUMENT

Introducing the Rationale of the Collegiate Learning Assessment

ROGER BENJAMIN
MARC CHUN

June 1, 2009



collegiatelearningassessment
COUNCIL FOR AID TO EDUCATION

215 Lexington Avenue, Floor 21 New York NY 10016-6023
p | 212.217.0700 f | 212.661.9766 e | cla@cae.org w | www.cae.org/cla

INTRODUCTION

Roger Benjamin

The principal goal of the CLA is to assist faculty in improving teaching and learning. Because arguments for and against assessment instruments often revolve around psychometric issues of reliability and validity, or larger policy debates about accountability, the question of the direct utility of assessment instruments for faculty in the classroom has remained marginal at best. Certainly, those associated with the CLA spend considerable time thinking and writing about the methodology and technical aspects of the CLA¹. Here, we place the question of the CLA's relevance front and center to indicate how it is being (or can be) used to improve teaching and learning. We argue that the CLA strategy discussed here is necessary, if not sufficient, for meaningful progress to occur in assessment in higher education.

Summative vs. Formative Assessment

Summative assessment—where assessment is separate from the learning process—usually plays the role of determining what aggregated knowledge or skills students attain at the individual, program or institutional level. In comparison, formative assessment is typically viewed as part of the teaching and learning process. The CLA as a performance-based assessment is designed to do both: to fulfill the traditional role of summative assessment for the institution, but more importantly, to also be used for formative purposes by faculty.

It is important to develop an integrated strategy that combines formative and summative dimensions in order to make realistic progress toward a more systematic approach to improving teaching and learning. This approach recognizes that faculty are the ultimate stakeholder of assessment. Unless there is formative value, faculty will not take assessment seriously; and in turn, they will not use the data to change practice to improve learning. Appropriate summative assessment is actually necessary in order to give faculty and administrators information they need to help frame a well-grounded formative assessment program.

Assessment approaches must do three things: first, satisfy the highest standards of reliability and validity so results can be meaningfully compared; second, be engaging to the student, constituting a significant learning as well as an assessment opportunity; and third, be recognized as authentic by faculty. Summative assessments typically meet one or two of these criteria, but rarely meet all three of them. The goal of the CLA is to satisfy all three criteria. Only in that way will these criteria be perceived as being of critical importance to faculty in higher education.

Evaluation Criteria For Formative Assessment

To satisfy the needs of the formative component of an assessment program, the answers to three key questions become central.²

¹ See reference section at the end of this paper. An asterisk is placed next to publications by CAE affiliated staff that address these questions.

² The creators of the CLA, like any group attempting to contribute to the field of learning outcomes in higher education, went through a number of decisions that make up a set of responses to the need for any assessment protocol to be reliable and

First, what components of student learning outcomes should be focused on? Higher order skills (critical thinking, analytic reasoning, problem solving and written communication) are a promising focus for a number of reasons: these skills are emphasized in liberal arts or what is called common learning, are thought to be critical in the knowledge economy, and are reflected as core to virtually all mission statements of colleges and universities.

Second, how should the selected outcomes be measured? Three plausible choices fall short:

- College GPAs are subject to grade inflation, which means there is too little discrimination among students. Additionally, grades are not calibrated across institutions (Klein et al., 2005).
- Tests given by academic departments may seem attractive; however, attempts to develop measures of the core outcomes of majors have proven problematic. Too few majors have agreed-upon definitions about what is central. And there are so many majors that the question of which ones are most central becomes controversial.
- Portfolios have strong face validity; faculty believe they can represent what students should know. However, selecting the “best” performances to go in a portfolio may not be representative of the abilities of the candidate. Moreover, the evaluation of portfolios is not seen as meeting minimum standards of reliability and validity. While portfolios are of understandable illustrative interest, it appears to be quite challenging to institutionalize their use for wide numbers of faculty for more than demonstration purposes. It is difficult, then, to consider making decisions about curriculum and pedagogy based on information largely based on portfolios.

Third, should comparative-based (summative) assessment at the institution-level be the initial focus? The case for comparison is itself strong when there is little or no consensus about whether existing theories provide minimum guidance for research and development in a field.³ If we had consensus about how to improve teaching and learning, we would not need such comparisons. That is not where we find ourselves in the field of teaching and learning. However, some standardized comparative-based assessment instruments should be used to answer the question, “Is what the institution doing good enough?”⁴ In other words, the right kind of comparative assessments (the ones that meet all three criteria for evaluation of assessments) present important signaling tools and checks for faculty and administrators to use as guidance at their own institutions.

In comparison to these alternatives, performance assessments have proved to meet all three essential criteria for evaluating assessments that can be effective in higher education. This does not mean we have completed the journey of adapting the performance assessment paradigm to undergraduate education. Performance assessment should be extended to majors and to additional skills such as perspective taking, among many other possibilities.

The Relevance of the Peer Review Analogy for Reporting Scores

Higher education produces at least three principal public goods: research, undergraduate education and service. For research, the higher education community long ago adopted the principle

valid, fair to all takers, respectful of time and resource limits for testing, and understand that decisions about what to test and how to test influences curriculum and instruction (Klein, 2002).

³ Ragin, 1989.

⁴ See Graff and Birkenstein (2008) for a defense of standardization.

that research, research grants and potential publications should be governed by peer review (Bush, 1945). There are strict rules governing this process meant to maintain as much objectivity in the review process as possible. Both the submitting authors and reviewers need to know that their arguments, ideas, or critiques will be protected from public scrutiny. Otherwise, there is no incentive for a researcher to submit drafts for publication because they might be embarrassed by the content of the critique. Similarly, the reviewers do not want their identity to be made public. The peer review process, at its best, thus generates honest appraisal of research that benefits the authors even if their paper is not accepted for publication. Peer review makes research a process of continuous improvement. Further, research standing is documented by the process of peer review; no professor, department or institution can simply assert that they are the best researchers in the region, state or nation.

Comparative-based standardized assessments, if they meet the previously noted criteria of evaluation of assessments), can be seen as the equivalent of peer review. Both peer review and comparative-based assessment attempt to provide an answer to the question, “Is it good enough?” As in the case of peer review of research, comparative-based assessments should be designed to be objective characterizations of an institution, offering evidence that is judged to be reliable and valid, and in particular, to have strong face validity (thought authentic by the faculty). Assessments that meet the measurement science minimum standards of reliability and validity offer powerful reality checks for institution-based formative assessment efforts. The fundamental point, by analogy, is that the organization that does the comparative-based testing need not make the results public. The assessment organization should report results for the institutions it tests back to them in confidence. Otherwise, why would an institution, department or program want to permit comparative-based testing which, we argue, plays a critical role in making formative assessment more systematic?⁵

Let us be as clear as possible, then, about the rules that govern reporting of CLA scores. They are and will only be reported back to the institution that contracted to use the CLA and the students who took the CLA tests. These rules have been promulgated by CAE’s Board of Trustees which is largely made up of distinguished current and past presidents of colleges and universities.⁶ To do anything else would cripple the usefulness of comparative-based tests and would result in distorted results being made public.

Implications for Assessment and Accountability in Higher Education

The kind of assessment evidence discussed here may be used in accountability systems agreed to by the institution. But the individual colleges and universities should play an important role in deciding whether and what to make public about its assessments, both comparative and local based.⁷ National and state organizations have recently launched pilot experiments such as the Voluntary System of Accountability (VSA) and the University of Texas system of accountability. The Council of Independent Colleges has created a consortium of colleges that is making progress in figuring out how to use the CLA results to good effect to improve teaching and learning. Other national organizations such as the Association of American Colleges and Universities are focused on promoting a broad set of liberal arts goals and measuring them with portfolios. The Council for Higher Education Accreditation and most of the regional accrediting organizations have or are moving to requiring

⁵ See Benjamin and Klein (2007) for a more detailed discussion of the relationship between assessment and accountability.

⁶ See CAE Board Statement on assessment presented to the U.S. Department of Education’s Spellings Commission (2006) at www.cae.org.

⁷ See Benjamin and Klein, 2007.

evidence of student learning improvement. And Robert Connor, president of the Teagle Foundation, with other leading higher education leaders, has created a Leadership Alliance for Student Learning and Accountability to focus attention on the need to systematically improve teaching and learning, with assessment as an essential ingredient.

A Key Design Feature: Respect for Faculty Autonomy

In higher education, faculty are provided broad discretion to decide what to teach, how to teach, and how to assess. The CLA recognizes the importance of faculty autonomy in their role as educators and does not proscribe any particular teaching and learning approach for faculty to follow. The CLA is built on the assumption that a successful assessment program in higher education needs to be designed to not only recognize the importance of faculty autonomy in teaching and learning but create strategies that encourage faculty to develop problem- and case-based assessments appropriate to their fields of expertise. In our view, there is a virtually unlimited range of faculty driven ideas to give content to the shift underway in undergraduate education to student centered teaching, represented by problem- and case-based curricula and text materials, and open-ended assessments which the CLA performance measures are designed to provide models for.

From the point of view of the CLA, assessment can be a form of experimentation that places the focus on improving teaching and learning. However, experimentation should be judged by how well the assessment strategy stacks up against the criteria for evaluation of assessments presented here. As I recently wrote:

We should apply to any proposed strategies for assessment and accountability the same logic by which Vanevar Bush developed the peer review research policy that respects the diversity of institutions and American higher education's institutional design, with its decentralized governance structure and respect for faculty autonomy. But we need to engage the kind of layered comparison strategy described here to improving the quality of teaching and learning in higher education.⁸ If we did not have the ability to do the kind of comparative assessment described here, the status quo might be acceptable. But we now can do sophisticated, meaningful, appropriate comparisons – and, since we can, there is no legitimate argument against doing so (Benjamin, 2008, p. 55).

The argument that some critics make that any comparisons between institutions are inappropriate is not credible. In fact, without appropriate comparisons that serve to link individual campus efforts, current formative assessment strategies are doomed to remain isolated silo activities at colleges and universities that, in the end, cannot be built upon to create the more systematic approach to student learning improvement many in the academy seek to achieve.

This is a time for serious innovation in assessment and accountability, and we should encourage fresh efforts to do so but while also subjecting them to rigorous scrutiny which, above all, means

⁸ By which I mean developing and using assessment instruments such as performance assessments that can be used for both summative and formative purposes.

employing the standards of reliability and validity developed by assessment scientists as well as the critiques of faculty who ultimately decide whether and how to assess.

We next turn to a discussion of how we apply the logic of the CLA.

INTRODUCING CLA EDUCATION: BRIDGING THE GAP BETWEEN ASSESSMENT AND TEACHING/LEARNING ⁹

Marc Chun

Background

In 1990 the U.S. Department of Education stated as a goal that “the proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially.” Despite such an important charge, a generation later this effort is still ongoing. As children born in 1990 were preparing to go off to college in 2008, then-presidential candidate Barack Obama stated that “We’ll teach our students not only math and science, but teamwork and critical thinking and communication skills, because that’s how we’ll make sure they’re prepared for today’s workplace.”

Consistent with Obama’s claim that critical thinking and problem skills are needed in the workplace, Halpern (1993) found that “virtually every business or industry position that involves responsibility and action in the face of uncertainty would benefit if the people filling that position obtained a high level of the ability to think critically.”

National educational associations have also recognized the importance of critical thinking skills as well. The National Assessment of Educational Progress (commonly referred to as NAEP) results suggest that “although basic skills have their place in pedagogy, critical thinking skills are essential” (Ballantine and Spade, 2007). And a report from the American Association of Universities (“Standards for Success”¹⁰) indicated that habits of mind (critical thinking, analytic thinking and problem solving) were essential for college success.

Although calls to action from the federal government, employers and national associations remind us of the importance of these skills, it is college and university faculty who should have the most important voice in this conversation. In fact, according to then-president of Harvard University Derek Bok (2006), national studies have found that more than 90 percent of faculty members in the United States consider critical thinking the most important goal of an undergraduate education.

Given this unanimity of commitment to ensuring that students have critical thinking skills, one might reasonably expect that college graduates would develop these core abilities. But do they?

A 2006 American Institutes for Research (AIR) study found that 75% of two-year college students and 50% of four-year college students did not perform at proficient levels of literacy (here, meaning being unable to compare credit card offers with different interest rates or summarize the arguments of newspaper editorials). More troublesome concerns arise when we look to college graduates. A 2005 National Center for Education Statistics study found that only 31% of college graduates could read a complex book and extrapolate from it. Further, the aforementioned AIR study found that 20% of college graduates had only basic quantitative skills (operationalized as being unable to calculate the total cost of ordering office supplies, compare ticket prices, or calculate the total price of a salad and sandwich on a menu).

So, given that the federal government, employers and, most importantly, faculty are behind the charge to equip students with critical thinking and other higher order skills, how is it the case that students are not developing them by the time they graduate?

⁹ Marc Chun is the director of CLA Education.

¹⁰ Visit www.s4s.org for more information.

What We Know from the Literature

One possibility is that we don't have a standardized, coherent theory of practice about the best way to develop these higher order skills. In Pascarella and Terenzini's (2005) comprehensive, meta-analysis of the literature on higher education student development, they summarize the evidence mounted to date about how to explain within-college effects on the acquisition of cognitive skills (with specific discussions of developing critical thinking skills). With respect to critical thinking, what we know from the literature boils down to just three points.

First, major field of study had little consistent connection to critical thinking, although Pascarella and Terenzini found that different fields of study led to the development of different reasoning skills. The main conclusion here could be that critical thinking skills are not the province of any one academic domain and that it is a fair expectation that all students should be able to develop such skills, regardless of field of study.¹¹

Second, there are particular student behaviors (outside of the classroom) that are associated with the development of critical thinking skills (e.g., the use of computers, hours studied, and number of non-assigned books read). Also, student-faculty non-classroom interactions were also found to have a positive effect. National Survey of Student Engagement (NSSE) research affirms the theoretical claim that student engagement is positively associated with student development.

Third, opportunities to interact with those who are different are positively associated with critical thinking skill growth; involvement in diversity experiences and service learning do as well.

Taken together, these three points indicate that critical thinking skills do develop across majors, when students are engaged, and when students come face-to-face with difference. But can they be taught?

There is good news and bad news. As we might hope and expect, Pascarella and Terenzini's review of the literature suggests that critical thinking can indeed be taught. Unfortunately, the observed effects have been quite modest. They indicate that this may stem in large part due to differences in definitions of what constitutes teaching critical thinking. As noted, with no shared conception of what pedagogical and curricular practices constitute such instruction, it is difficult to measure this. Further, since most of this research used multiple-choice or student self-reports to measure critical thinking, the research was hampered in its ability to measure the key dependent variable.

So what does the research literature tell us? The preponderance of evidence about how to improve critical thinking skills falls may largely be beyond the scope of what happens in the classroom, and thus, beyond the scope of what faculty can do. The key though is that evidence also suggests that it *is* possible to teach these skills, but we must be more clear about what that means and how to do it. CLA Education seeks to be a practical response.

AUTHENTIC ASSESSMENT AND PERFORMANCE TASKS

Both CLA Assessment Services and CLA Education share a common starting point: a demonstration of critical thinking and other higher order skills. As noted previously (Chun, 2002, 2006), we often are left finding simple and convenient proxies for these skills that faculty agree are most important, and all too often better is sacrificed for that which is faster and cheaper. The result

¹¹ That is, it is not only students who earn degrees in or take courses in philosophy who should develop critical thinking skills.

often has been assessment that serves as a compliance exercise rather than a means to improve teaching and learning.

Authentic assessment and performance tasks provide a means of bridging the chasm between institutional assessment and classroom practice; between the summative and the formative. In fact, the authentic assessment movement informs the work of the CLA. Grant Wiggins (1990), considered to be a guru of authentic assessment, noted:

Assessment is authentic when we directly examine student performance on worthy intellectual tasks. Traditional assessment, by contrast, relies on indirect or proxy ‘items’ -- efficient, simplistic substitutes from which we think valid inferences can be made about the student’s performance at those valued challenges.

Wiggins argues that authentic assessment presents students with an array of tasks that reflect best instructional activities: writing, doing research, engaging in oral analysis, collaborating with others.¹² Further, authentic assessment requires students to effectively perform using acquired knowledge, often by completing tasks that are intentionally ambiguous and not clearly structured. This presents challenges and roles that help students rehearse for the situations they will face later in life: Wiggins notes that authentic assessment achieves validity when the task simulates larger world “tests” of ability. Authentic assessment challenges students to create complete and justifiable performance, answers or products in response to a meaningful prompt. Finally, authentic assessment gains validity and reliability through appropriate scoring criteria for the product.

All CLA programs embrace these principles of authentic assessment, and this big idea of looking for the demonstration of key skills takes the form of performance tasks (which are a specific form of authentic assessment). Michael Hibbard¹³ (2000) described central features of performance task assignments or projects that are:

- Engaging (they grab students’ attention)
- Activating (more student work leads to being drawn into the task)
- Authentic processes (steps taken mirror those of similar performance for similar audience)
- Authentic products (products or performance similar to that of “larger world”)
- Essential (connected to important—and not trivial—standards)
- Integrative (knowledge, thinking skills, problem-solving skills, and writing are utilized together)
- Embedded (used with—and not added-on-to—instruction)
- Appropriately structured (sufficient explanation such that students understand the task)
- Feasible (tasks are possible to complete given time and other resources)
- Equitable (Fair to students based on background knowledge)

By design, all of these features are included in the CLA Performance Tasks.¹⁴ The authentic nature, focus on key higher order skills, and means of delivery are all hallmarks of the CLA.

¹² It is surprising, perhaps, that academics have their students complete such engaging classroom work, and yet accept standardized, multiple-choice tests to assess student learning.

¹³ Hibbard writes about middle school science, but arguably the same principles are applicable to other domains.

¹⁴ See the explication of the goals and underlying structure of the CLA performance tasks in Chapter 5 of the complete monograph.

Additional features that Hibbard noted include:

- Feedback and revision loop (a chance to make revisions)
- Group work and individual work (a balance between the two approaches)
- Promotion of deeper understanding (expects students to go beyond surface-level knowledge)

These last three points are central for using performance tasks as a classroom activity to promote learning. Although these features of authentic assessment are beyond the scope of CLA Assessment Services, they are built into performance tasks created by faculty as part of CLA Education.

What may seem obvious upon reviewing this list is that performance tasks can bridge the gap between assessment and classroom practice. What CLA has done is capitalize on this. The same tools that can be used as an effective means to assess learning can also be used to improve learning. It is by connecting these two features that systematic improvement of teaching and learning can and will occur. There is a simple poetry to this stance: if we want students to develop critical thinking skills, we should have them practice critical thinking skills. If we have students learn and develop critical thinking skills in an authentic, practical manner using performance tasks, we can assess the growth in their skill development using performance tasks. Here, assessment and teaching and learning become one-and-the-same.

CLA EDUCATION

CLA Education focuses on curriculum and pedagogy and explicitly recognizes the central role of faculty in the process of developing higher order thinking skills like critical thinking. This is accomplished through a number of products and services that are described in detail below¹⁵.

Performance Task Academy

The first Performance Task Academy was offered in March 2008. This basic two-day Academy includes a mixture of mini-lectures, small group activities, large group discussion, and independent work. The Academy is designed to be an introduction for faculty who have little to no experience creating performance tasks, and a chance for developing tasks for faculty who already have some practice.

The first day of the Academy focuses on authentic assessment practices, performance tasks, rubrics, and providing diagnostic feedback. The second day begins with a series of activities to help faculty learn how to put these concepts into practice. Faculty then spend the balance of the day creating an actual performance task that they might use in their own classrooms. They receive feedback from each other, as well as from the facilitators.

The number of participants is capped to keep activities highly interactive. Institutions are encouraged to send teams of at least two faculty members to the Performance Task Academy. To date, approximately 700 faculty members have participated in one of the 26 Academies offered in the inaugural year. These Academies have been offered across the country, as well as in Korea and Japan.

Feedback about the Performance Task Academy has been overwhelmingly positive, and perhaps reveal what professors found to be lacking in other faculty development activities. One associate professor of psychology noted:

¹⁵ See Appendix I for a full list of CLA Assessment and CLA Education products and services.

Generally if I come back from a conference/workshop with one small nugget of good info, I am satisfied. Here, however, I was amazed that the entire two days were fabulous. I never felt bored or wished things would move along more quickly.

A professor of English noted:

"This stuff gives us something concrete to work with, to analyze and evaluate. It's much better than just waving our hands and saying, 'well, you know, it's about 'critical thinking'."

Other participants' responses indicate that they appreciate the active learning aspects built into the workshops. One faculty member wrote:

The facilitators were the best I've ever had at a workshop. I was particularly impressed by the connection between teaching/learning and assessment and that there was a focus on authentic assessment and changing the culture of teaching and learning at our institution.

Another commented:

The workshop was developed to ensure maximum learning. The two-day workshop consisted almost entirely of active learning. It is refreshing when presenters practice what they preach.¹⁶

Participants also value the opportunity to work with other faculty. One participant wrote:

I've made important contacts that I will maintain as I continue to develop my performance tasks.

New workshops will be launched in the 2009-10 academic year. These will include the Advanced Task Development Academy—a special four-week program, in which initial in-person training will be provided on creating performance tasks. Faculty will then work for four weeks (with web-based workshops to support task development), followed by another in-person workshop that will be held to share and attain feedback on the finished performance tasks. Additionally, a Course Development Academy will instruct faculty on how to create an entire course centered on the use of performance tasks. Discipline-based Performance Task Academies are also being planned

Finally, there has been interest among faculty about being “certified” in developing the skills to create such performance tasks. Some faculty have indicated that they would be interested in including this in their tenure files. Accordingly, we are exploring how to provide such certification and to create a community of faculty members committed to using the performance task pedagogy and curricular approach to promote development of these higher order skills.

Performance Task Library

Faculty from across the country who have participated in one of the Performance Task Academies have created course projects and assignments that utilize authentic assessment approaches

¹⁶ Moving forward, the plan will be to have an independent, outside evaluator observe and document the experiences at an Academy.

introduced during the workshop. These performance tasks include a range of topics based in different disciplinary orientations, and often cross disciplinary boundaries. While some tasks that come out of the Performance Task Academies are appropriate for general education courses, one hallmark of the Academy is that it encourages faculty members to use the architecture of the CLA performance tasks to embed their own course content. These content-based performance tasks can be used by faculty members as class assignments, group projects or even (as one participant has done) the structure of an entire semester's curriculum. These faculty-created performance tasks reinforce the key principle of teaching and learning improvement; namely, that the way faculty assess their students' skills should be in line with the way that they are teaching these skills.

The Performance Task Library is designed to encourage and incentivize the sharing of performance tasks between faculty and among institutions. On the CLA in the Classroom website (www.claintheclassroom.org) many of the tasks that have been created by Performance Task Academy participants are available for downloading and sharing. For instance, currently available tasks discuss such topics as: whether or not to subsidize corn ethanol, if and how federal stimulus money should be used to support weight loss programs, and determining where it makes most sense to site a new firehouse in a city.

The caveat to downloading previously created tasks from the Performance Task Library is that in order to access an available task, a faculty member has to have submitted a performance task of her or his own; these task are then reviewed by the CLA Education staff. This is designed to reinforce the communal and collaborative aspects of the CLA and in particular the ideals of the CLA Education programs. As suggested by the saying “a rising tide floats all boats,” we trust that in encouraging the sharing of performance tasks across institutions we promote the idea that all institutions can benefit from a collaborative focus on improving teaching and learning.

Student Diagnostic Report and Institutional Diagnostic Report

In order to facilitate further the improvement of student learning, CLA Education will be launching in the Student Diagnostic Report (SDR) and the Institutional Diagnostic Report (IDR). These programs allow individual faculty members to assess the students in their own classes, using a CLA Performance Task no longer in rotation as part of CLA Assessment Services. Student responses are scored by a trained reader, but rather than controlling for initial ability and providing analyses of overall student growth (as is part of CLA Assessment Services), responses are scored relative to the basic demonstration of the key higher order skills in an absolute sense. Using a scoring rubric (that may eventually anchor planned diagnostic aspects of CLA Assessment Services), faculty receive sub-scores of student performance on analytic reasoning and evaluation, problem solving, persuasiveness, and mechanics. The SDR reports back individual student results with careful caveats to note that the results should not be used for high-stakes decision making, but instead as one piece of evidence that could be used as part of a diagnostic conversation with students about their skills. This should be part of a larger, ongoing process focusing on student development. The IDR aggregates these results the whole institution.

The SDR and IDR programs permit faculty to have on-demand assessment (that can be conducted off-cycle from the formal CLA), since it need not be scored with and reported back relative to the rest of the institutions participating in the CLA.

SUMMARY AND CONCLUSION

In the complete monograph, Benjamin, Chun and Jackson (chapter 2) note that the CLA is strategically placed in the context of assessment and accountability, and that it is a multifaceted approach for engaging in assessment, education, and research. CAE policy ensures that these components will inform one another. Benjamin, Chun and Shavelson (chapter 3) outline the role of diagnostic logic employed by the CLA, which provides a justification for the work of CLA Education, and Benjamin (chapter 4) focuses on faculty and how the CLA is a test worth teaching to. Specific CLA performance measures are outlined by Hardison, Hong, Chun, Kugelmass and Nemeth (chapter 5).

Despite consensus about the need for students to develop higher order skills like critical thinking, research has shown that students are not developing them. Although we can argue that faculty are the key to improving the teaching and learning of critical thinking skills, the research has provided limited lessons supporting this. We have hints that faculty can teach these skills, but little evidence given the divergence of ways in which teaching critical thinking has historically been operationalized.

CLA Education takes a straightforward approach: that students will develop these skills if they have opportunities to practice them. Drawing on the literature on authentic assessment and performance tasks, CLA has found a key link between assessment and classroom practice. When substantive classroom work focuses on development of these skills, but assessment takes another tact (e.g. multiple-choices measures or student self-reports of learning), it is not surprising to find a disconnect between the two (and for faculty to be frustrated by the lack of validity in the measures used to assess their work). CLA connects the two by using performance tasks that serve both purposes. This coherent approach has served to galvanize faculty, since they can see the connection between assessment and pedagogy. A dean of Academic Affairs noted:

One of our faculty members was skeptical of the utility of the CLA given her own professional needs and concerns. However, after she participated in the workshop on designing a unique, course-based performance task to assess student learning, she was a convert. Immediately, she saw how a performance task—especially one that she could tailor to reflect the concerns of her science course—could energize student learning and serve as a platform on which students could synthesize and integrate knowledge from various different disciplines and courses.

The continuous system of improvement described herein (and more completely in the full monograph) makes possible the direct connection between assessment and classroom teaching and learning. Having a clear mandate to develop critical thinking skills sets the stage for this work to occur, and recognizing the role of faculty is crucial. But after reconfiguring the landscape of assessment and accountability, and by recognizing that authentic assessment and performance tasks may finally provide the tools not only to directly connect assessment and classroom practice, but more importantly to do it in a way that faculty endorse, it is only now that the real work can begin. It will not be easy work and there may be new challenges teaching faculty how to do this. But now that the stage is set, we can abandon all that distracted us and confused the process. We are ready to move forward.

We can now return to learning.

REFERENCES¹⁷

- Ballantine, Jeanne H. and Joan Z. Spade (Eds.) (2007). *Schools and Society: A Sociological Approach to Education*, Third Edition. Wadsworth Publishing.
- *Benjamin, Roger (2008). "The Case for Comparative Institutional Assessment of Higher-Order Thinking Skills" *Change*, Vol 40, No. 6, November/December: 51-55.
- *Benjamin, Roger and Stephen Klein (2007). "Assessment Versus Accountability In Higher Education: Notes On Reconciliation." United Nations Educational, Scientific and Cultural Organization (UNESCO) Commissioned Paper Series: 1-26.
- *Benjamin, Roger, Marc Chun, and Richard Shavelson (2007). *Holistic Tests In A Sub-Score World: The Diagnostic Logic of the Collegiate Learning Assessment*, New York, NY: CAE.
- *Benjamin, Roger (2007). "Recreating the Faculty Role in Governance in Research Universities," in Joseph C. Burke (ed.) *Fixing the Fragmented University*. Bolton, MA: Anker Publishing, 70-98.
- Bok, Derek (2006). *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*. Princeton University Press.
- Bush, Vanevar (1945). *Science: The Endless Frontier*. Washington, D.C.: United States Government Printing Office.
- *Chun, Marc (2002). Looking Where the Light is Better: A Review of the Literature on Assessing Higher Education Quality. *Peer Review*, 4 (Winter/Spring): 16-25.
- *Chun, Marc (2006). *Faster, Better, Cheaper: The Iron Triangle of Higher Education Assessment*. New York, NY: CAE
- *Council for Aid to Education (2006). Board Statement on Assessment. www.cae.org
- Graf, Gary and Carol Birkenstein (2008). A Progressive Case for Educational Standardization: How Not to Respond to the Spellings Report. *Academe On line*, May/June. Research. San Francisco, CA: Jossey-Bass. <http://www.aaup.org/AAUP/pubsres/academe/2008/MJ/Feat/graf.htm>
- Hibbard, Michael (2000). *Performance-Based Learning & Assessment in Middle School Science*. Eye on Education.
- *Klein, Stephen (2002). Direct Assessment of Cumulative Student Learning in *Peer Review*, Vol. 4, No. 2/3: 26-28
- Klein, Stephen, David Freedman, Richard Shavelson, Roger Bolus (2008). *Assessing School Effectiveness*. *Evaluation Review*, 32, 6: 510-525.
- *Klein, Stephen, Marc Chun, Laura Hamilton, George Kuh, and Richard Shavelson (2005). An Approach to Measuring Cognitive Outcomes Across Higher Education. *Research in Higher Education* 46.1: 251-276.
- *Klein, Stephen, Roger Benjamin, Roger Bolus, and Richard Shavelson (2007). "The Collegiate Learning Assessment: Facts and Fantasies." *Evaluation Review* 31.5: 415-439.
- *Koretz, Dan, Brian Stecher, and Stephen Klein, Dan McCaffrey, and Edward Deibert (1993). *Can Portfolios*

¹⁷ Asterisks indicate publications by CAE-affiliated authors who address key questions outlined by Benjamin in the first chapter.

Assess Student Performance and influence Instruction? The 1991/92 Vermont Experience. 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.

Pascarella, Ernest and Patrick Terenzini (2005). *How College Affects Students: A Third Decade of Research*. San Francisco, CA: Josse-Bass.

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

*Shavelson, Richard (2007). *A Brief History of Student Learning: How We Got Where We Are and a Proposal for Where to Go Next*. Washington, DC: Association of American Colleges and Universities.

*Shavelson, Richard (forthcoming). "The Collegiate Learning Assessment," in *The Quest to Assess Learning and Hold Higher Education Accountable*. Stanford, CA: Stanford University Press.

*Shavelson, Richard, Amy Kurpulis and Matt Bundick with Richard Hersh, Daniel Silverman, Corey Keyes, and Lynn Swaner (unpublished paper, 2009). *On Assessing Learning Broadly and Responsibly*.

Simon, Herbert (1996). *The Sciences of the Artificial*. Boston, MA: M.I.T. Press.

The College Portrait: the VSA Template. www.voluntarysystem.org

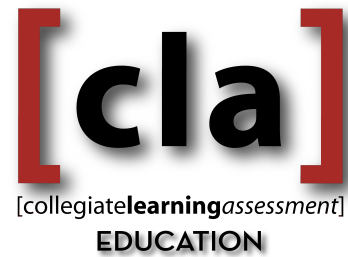
U. S. Department of Education (2006). *The Secretary of Education's Commission On The Future of Higher Education*. Washington D.C.: Government Printing Office.

Wiggins, Grant (1990). *The Case for Authentic Assessment*. *Practical Assessment, Research & Evaluation*, 2(2).

Appendix I: CLA Products and Services



- **Institutional Reporting**
Cross-sectional and longitudinal models permit value-added comparisons across participating institutions and over time.
- **Student Level Data File**
Allows institutions to run cross comparisons with locally collected variables, or to disaggregate data in sub-groups for institutions that choose to in-depth sample.
- **Sub-group Analyses through In-Depth Sampling**
Allow institutions to drill down and consider variation with their student samples (e.g. by course-taking patterns, student demographics, etc.)
- **CLA Coffee [cla]tch**
CLA staff facilitated networking opportunities at national conferences where the primary goal is institutional collaboration and the sharing of best practices.
- **CLA Spotlight**
A monthly web conferencing series that features campuses doing interesting work using the CLA, guest speakers from the larger work of assessment, or CAE staff members who provide practical advice.
- **Peer Group Analyses (*in development*)**
Designed to enable institutions to select the criteria (e.g., size, sector, geographic region, etc.) that reflect various institutional characteristics that are of interest to them, allowing them to control for other characteristics (beyond entering test scores) that narrow the group of similarity.



- **Performance Task Academy**
With a focus on curriculum and pedagogy, the Performance Task Academies allow faculty to work with retired CLA tasks to learn the basic “technology” behind the CLA. In so doing, faculty learn how to create their own content-embedded performance tasks, which can be used as class assignments, group projects, or semester long curricular activities.
- **Performance Task Library**
Academy attendees prepare complete or near-complete performance tasks of their own. They are encouraged to share these with faculty at other schools through the Performance Task Library, where complete tasks can be downloaded, adapted and implemented.
- **Faculty Certification (*in development*)**
Starting this fall, faculty members participating in a Performance Task Academy will have the opportunity to be certified for their ability to create their own tasks, which they can add to the professional development portfolio.
- **Student Diagnostic Report (*in development*)**
Students in one classroom take a CLA Performance Task that is scored on demand, enabling the reporting of absolute performance at the student level.
- **Institutional Diagnostic Report (*in development*)**
Using the same diagnostic rubrics at the Student Diagnostic Report, the Institutional Diagnostic Report provides information on overall cohort performance, and includes specific examples of student work.