



# A New Field of Dreams

## The Collegiate Learning Assessment Project

By **Roger Benjamin**, president, and **Marc Chun**, research scientist, both of RAND Corporation's Council for Aid to Education<sup>1</sup>

In the film *Field of Dreams*, an Iowa farmer hears a spectral voice that instructs him, "If you build it, they will come." Despite formidable challenges and doubt from all those around him, he embarks on a quest to turn his Midwest farmland into a baseball field. Although he maintains his faith and conviction, he is plagued by the question of whether they will, in fact, come.

The RAND Corporation's Council for Aid to Education (CAE) recently undertook what some might consider to be just as foolhardy an endeavor: to build a new assessment approach for higher education. This approach, which assesses the "value added" of the institution, has now evolved into the Collegiate Learning Assessment (CLA) project.

There are numerous technical, political, and pedagogical factors that would make one hesitate before attempting such a project. However, our initial foray into this arena—a feasibility study in 2002 with more than 1,300 students at fourteen colleges and universities across the country—found that the approach was both viable and useful. The discussion that follows describes the key features of this feasibility study, organized around questions we asked when building the CLA project.

### Is There Really an Assessment Alternative?

The CLA project differs from most other approaches to student assessment in four ways. First, it uses direct measures of student learning rather than proxies for it; typical proxies include input or actuarial data (e.g., entrance examination scores or faculty salaries), student self-assessments of growth, or college faculty and administrator opinion surveys (e.g., the *US News & World Report* rankings). As we have reported elsewhere,<sup>2</sup> there are methodological concerns in interpreting such indirect measures. Although the CLA project does not dismiss input<sup>3</sup> or actuarial measures (which provide valuable information about a college or university), it recognizes that these measures do not focus explicitly on skills and abilities colleges and universities are committed to developing. Therefore, performance measures of actual learning are an important addition to existing approaches to assessment.

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

Third, the project uses a "matrix-sampling" approach to assessment. The traditional approach, which would be to administer an entire battery of instruments

NOTE  
This article provides an update on the Value Added Assessment Initiative, which was the focus of a special double issue of *Peer Review* (4:2/3 winter/spring 2002).

1. The authors—not RAND or CAE—are solely responsible for the views expressed herein. The authors wish to thank RAND's Dina Levy for helpful comments on an earlier version of this paper.

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

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



to all students, would be too time-consuming to be practical. Instead, the sampling design involves administering separate components of the full set to different (but randomly selected) sub-samples of students, thereby minimizing the time required per student yet still allowing complete coverage of the range of instruments and content areas. This matrix-sampling design provides comprehensive and reliable information about how well a school's students are doing as a group rather than about the proficiency levels of any individual student.

Fourth, the project was designed to assess value added, or the institutional contribution to student learning. We do this in two ways: (1) we measure how well an institution's students perform relative to "similarly situated" students (defined in terms of their SAT or ACT scores),<sup>4</sup> and (2) we measure how much students' skills improve during their tenure at the institution through a pre-test/post-test model. As the research continues, we will also consider establishing baseline benchmarks against which institutions can evaluate basic skill development.

### Why Focus on Assessing General Education Skills?

There are three related rationales behind the focus on assessing general education skills. First, most colleges and universities highlight general education as part of their undergraduate curricula. These are seen as the knowledge, skills, behaviors, and attitudes characteristic of an "educated per-

son." These general education skills—such as critical thinking, analytic reasoning, and written communication<sup>5</sup>—cut across academic disciplines and departments.

Although any given college or university may adopt different pedagogical approaches to develop such skills, they nonetheless all share an overall commitment to these dimensions of learning and assessment. However, there are limited tools available to permit systematic evaluations of how institutions are doing in reaching their general education goals. The CLA project, therefore, seeks to contribute to the overall assessment efforts by contributing new instruments and a method that reflect the value placed on general education.

Second, whereas it is common to assess outcomes of individual courses, we believe that general education is not so neatly compartmentalized. It is, rather, the sum total of the combination of courses a student takes, plus the learning that occurs "between" courses, that contributes to overall skill development. As a result (and as will be discussed below), the focus on the institution as the unit of analysis is motivated by an interest in understanding the overall impact of the college or university as a whole. This, we argue, is a more holistic way to understand general education.

Third, whereas discipline-specific measures focus on content, and some instruments might assess the ability to recall facts or formulas, the CLA project measures students' demonstrated ability to use information.

Focusing on general education skills makes

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

### Can These Skills Be Assessed?

Two different sets of performance measures were administered during the feasibility study. One set consisted of six performance tasks. The tasks measure a student's ability, for example, to read a table of data, make sense of a literature review, analyze an interview transcript, and review a newspaper report, and then to weigh the relative value of each document, synthesize the material, and prepare a cogent response to a question. These tasks, which take ninety minutes each to complete, are set in various contexts such as science, social science, and arts and

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

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

humanities. We used four of the “Tasks in Critical Thinking” (developed by the New Jersey Department of Education) and two CLA performance measures specifically developed for the project.

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

Student responses can be graded by a trained reader or by a computer.<sup>6</sup> There was a 0.50 correlation between a student’s college GPA and scores on the CLA measures. This correlation was substantially higher (0.65) when corrected for the less than perfect reliability of the measures. The corrected coefficient (which uses the institution as the unit of analysis) provides a more relevant indicator (than would student-level measures) of the degree to which the CLA measures tap skills that schools value (as reflected by the students’ grades).

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

### Can the Institution Be the Unit of Analysis?

The CLA performance measures we used were not designed to assess the same construct or provide scores that would be reported for individual students. Instead, a combination of measures was used from different clusters of academic disciplines. We would not expect that a measure set in a science context would necessarily correlate especially highly with one in the arts or humanities,<sup>7</sup> but the combination of measures across disciplines would provide a more robust measure of the institution’s contribution to overall student learning.

### How Can Value Added Be Assessed?

We explored the “value added” of the college experience by analyzing both within- and between-school effects. The within-school

effects analysis found that, after controlling on the students’ SAT scores, upperclass students (juniors and seniors) tended to earn higher scores on our measures than did underclass students. This suggests that the measures capture institutional effects (recognizing that learning occurs both in and out of the classroom).<sup>8</sup> The correlation between years in school and test scores was statistically significant. A school’s average score on the CLA measures also correlated highly with the school’s average SAT score ( $r = 0.90$ ), yet we found statistically significant institutional effects after controlling on SAT.<sup>9</sup>

The between-school effects analysis examined whether the students at some schools were, on average, scoring higher or lower than would be expected<sup>10</sup> on the basis of their mean SAT scores. Thus, the amount of education a student receives is related to the kinds of skills we assessed, and these relationships transcend the abilities tested by college entrance exams. We use this approach as a means to quantify “value added.”

### Can Such an Assessment Be Done Economically?

The assessment can be done in a cost-effective manner and within a relatively short

---

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

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

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

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

10. Operationalized as more than two standard errors relative to the campus’ spread of scores.

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

### Will Schools Teach to the Test?

There is nothing wrong with teaching to the test if test performance demonstrates skills or abilities that are valued. This is analogous to intentionally teaching student pilots how to land an airplane in a cross-wind because the final pilot's exam involves performing that task; there is inherent value in teaching to such a test. Hence, we would encourage schools to teach to the test if that activity involved working with students to develop their analytic reasoning and writing skills and developing skills that students will need to demonstrate but still have value outside of the testing situation.<sup>11</sup> In fact, we recognize that if an assessment approach does not reflect educational goals that faculty support, it inevitably will fail. Thus, the measures have been designed

specifically to address some of the common elements that cut across higher education sector and academic field and that we believe faculty will endorse.

### Will Students Participate?

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

### Will Institutions Participate?

From the inception of the project, we knew that the question of institutional participation would be one of the greatest challenges. However, given the realization that the measures are ready to be used and the subsequent interest, CAE has created a nonprofit service that will allow institutions to pay a nominal fee to use the measures.

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

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

### Lessons Learned

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

---

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