

# Faster, Better, Cheaper: The Iron Triangle of Higher Education Assessment<sup>1</sup>

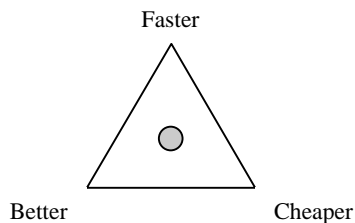
Marc Chun, PhD, Research Scientist, The Council for Aid to Education<sup>2</sup>

We live in a “choose two” culture. Although we typically seek to accomplish everything faster, better and cheaper, unfortunately we most often find that we can never have all three at once. Ask a chef to whip up a meal quickly and cheaply, and more likely than not it won't be very good; ask an architect to design a building of higher quality but in less time, and that won't come cheap; or ask NASA to create a new space probe that is better but to do so with limited funds, and it will inevitably take more time.

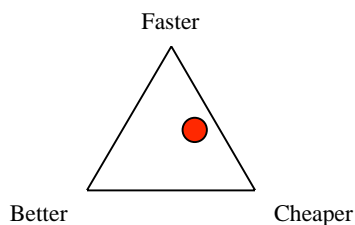
In this choose two culture, we instead find that we must trade off these triple-constraints against one another. We will have to decrease one dimension in order to improve the other two (or often to even keep the other two steady): for example, we can stay under budget and complete the work before the deadline by cutting corners, but this comes at the expense of quality. We can produce something of higher quality less expensively, but that will take longer.

In the same vein, if one dimension is unbounded, we have tremendous flexibility. If we have an unlimited budget, we could do whatever we want as fast as possible; it will just cost a bundle.

One way to illustrate this dynamic is the "iron triangle" below, with each dimension at one of the triangle's vertices. Think of the default mode as the gray circle in the center -- the project will be completed at moderate speed, of moderate quality and at moderate cost.



Any movement towards a side of the triangle means sacrificing at least one of the dimensions. As noted in the example below, the red circle indicates that by "choosing two" to complete the task *faster* and *cheaper*, one must be willing to accept that the process might not necessarily be *better*.



---

<sup>1</sup> The author gratefully acknowledges the helpful comments of Elizabeth McEneaney, Mary Rauner and Richard Hersh on an earlier draft of this essay.

<sup>2</sup> Please contact the author at 215 Lexington Avenue, Floor 21, New York, NY 10016-6023, 212.661.5800.

In many ways, the same can be said of higher education assessment, and more specifically, assessment of undergraduate student learning. Before proceeding, it is important to define terms as relevant to this case:

What's **faster**? Requiring less time to collect data and to complete analyses.

What's **cheaper**? Necessitating fewer resources (money, staff, technology, and other materials) to collect the data and complete analyses.

What's **better**? Having overall higher quality assessment (admittedly the least straightforward of these terms, but here it is defined as the accuracy and authenticity of the indicators, and the scope of the assessment).

### **Assessment at the Individual Course Level**

The individual faculty member goes through a version of this calculus when deciding how to assess student learning in the classroom. It is reasonable to assume that most faculty members would agree that a better way to assess student learning is going to be more holistic and comprehensive. A way to understand the degree to which a chemistry student has mastered the material might be to see if she can identify an unknown substance by meticulously doing titrations and observing interactions with reagents; an architecture student might build a model of a proposed home; a philosophy student might do an ongoing series of oral presentations that span the academic term. To assess general education skills such as critical thinking and written communication, students could be engaged one-on-one in a dialogue with the faculty member, could write theses, or could do some other performance-based project that require the demonstration of such competencies. Arguably, such forms of assessment of student learning and mastery of material are indeed better, but they take more time and greater expense of resources.

In a small course, however, this may be a fair trade-off. In an upper division seminar with half a dozen students, it is reasonable for a faculty member to make the investment of time and cost to do such assessment. It not only takes more class time and resources for the students to demonstrate their competencies, but also more time and resources for the faculty member to complete the evaluation and provide feedback. However, the quality of the feedback is also likely to be better. A faculty member can provide a more nuanced and specific set of critiques that may promote more learning on the part of the students.

Yet, once we scale this up to a full auditorium lecture hall with 500 students, this no longer seems feasible. Thus, the chemistry laboratory activity becomes a multiple-choice test; the individual model home project becomes a class group project; the series of presentations becomes a five-minute briefing. In order for the faculty member to maintain some reasonable semblance of faster and cheaper, it becomes inevitable that better must be compromised. And whereas a regular one-on-one discussion between the faculty and student may permit a faculty member to note specific areas of strength and development, the results of a multiple choice exam offer limited opportunities for such luxuries.

The high-quality assessments could be maintained if one of the other two constraints is unbounded. If the faculty member had the luxury of a large assemblage of willing and capable teaching assistants, such assessment again becomes possible; if the faculty member had no other commitments and could devote all time to the course, this also begins to seem viable. Thus, a faculty member can really have only two of the three.

### **Assessment at the Institution Level**

When we shift from the classroom-level to the institution-level of assessment presents a similar challenge as moving from a small to a large course. Although we may have great success with a smaller-scale assessment effort, bringing it to scale creates similar challenges. Here, we again see this same three-way tug-of-war. At the institution-wide level, ask a campus administrator to assess student learning in areas such as writing and thinking, and again although the desire may be to have all three, it's really only fair to expect her to choose two, and she must often settle for just one. Put another way, a campus administrator can only really control two dimensions, and can only really guarantee those two.

As a heuristic, the traditional approaches this administrator has at her disposal can be organized into four basic families or groupings: (1) actuarial data; (2) ratings of institutional quality; (3) student surveys; and (4) direct measures of student learning. Each will be considered along the dimensions of the degree to which they are faster, better and/or cheaper.

#### ***Method 1: Actuarial Data***

What are often seen as the most "objective" measures of higher education are the analyses based on "actuarial" data. These data include graduation rates, levels of endowment, student/faculty ratios, highest degrees earned by faculty members, selectivity ratios, admissions test scores of entering students, and levels of external research funding. Although not intrinsic to the data themselves, the way in which the analyses are conducted typically rely upon two central assumptions: that a "*higher quality*" institution has more resources (funding, faculty [which is defined as a higher percentage of any given cadre holding Ph.D.s], and students [those with high entrance examination scores]), and that *students learn more at such "higher quality" institutions*. As one might suspect, these assumptions are not universally accepted.

Researchers argue that the primary advantages of using actuarial data are that these data are relatively straightforward to collect, and the resulting statistics can be easily compared across institutions and over time. Indeed, actuarial data are arguably *faster* and *cheaper* to collect; given the highly systematized and standardized procedures, such projects capitalize on tremendous efficiencies of scale. However, although actuarial data have prima facie validity in objectively assessing higher education quality, it is not clear if such approaches are *better*, in that the tools often cannot tacitly measure student learning.

#### ***Method 2: Ratings of Institutional Quality***

A second approach of higher education assessment is based on analyses of ratings and rankings of institutions. This has typically taken the form of surveying either or both college faculty and administrators and asking these "experts" to rate the quality of different institutions and their programs on a series of dimensions, again with the assumption that students learn more at such

"higher quality" institutions. Further, the implicit logic here is that informed "experts" can best assess institutional quality. Some rankings are based in part on actuarial data (such as selectivity, faculty resources, and financial resources), but are also based on surveys of faculty and administrators that ask for their perceptions and opinions about academic quality and reputation.

Again, although such rankings are relatively *faster* and *cheaper*, concerns about ranking methodologies have raised doubts that this information is indeed *better*. Rating systems are susceptible to methodological concerns. For the rating systems published in popular periodicals, the weighting approaches are often editorial in nature and hard to defend on theoretical bases (e.g., they may weight student-faculty ratio as 20% of the total score and research productivity as 10%, without providing any empirical justification that student learning is affected by each variable in such proportions), and ratings can be highly sensitive to even minor weighting changes. Additionally, some variables may lack face validity: alumni giving is claimed to serve as a proxy for student satisfaction, when it can arguably be instead a function of effectiveness of the development office or the relative wealth of the students. Further, reputations change slowly, advantaging those that are slowly declining, and disadvantaging those that are rapidly improving.

### ***Method 3: Student Surveys***

A third approach used to assess institutions is based on self-reported student information. In contrast to the proxy data used in the actuarial approach and ranking data based on surveying faculty and administrators, these data are collected by asking students directly about their collegiate experiences, satisfaction, academic abilities, and educational and employment plans. Typically, individual institutions collect such data to gather feedback about their institution while national researchers collect data from a number of institutions in order to generate research on the effects of higher education in general.

In contrast to actuarial data and ratings systems, the use of student surveys demonstrates some of the trade-offs involved. Although such data aren't cheaper and faster than the first two methods discussed (it takes more resources and time to collect surveys from individual students), if one is seeking to understand student learning, these data are somewhat *better* in that such rough proxies are not employed. Further because many of the outcomes of interest cannot be empirically measured (e.g., attitudes and values), the use of student self-reports is commonplace in higher education research because they are seen as better. However, a key issue in student surveys, as in all surveys, is that of the reliability of the self-reported data (particularly given the "desirability bias").

Still, although you may have better data, this does not necessarily lead to better analyses. It may be problematic to determine the actual impact of any process variables. Moreover, the traditional positivistic approach often employed in such analyses assumes that individual aspects of the college experience can be studied atomistically; it can be seen as denying the holistic nature of student learning.

### ***Method 4: Direct Assessments of Student Learning***

A fourth approach to assess institutional quality is to measure student learning directly. Direct assessments of student learning are perhaps the least systematically used of the four methods

discussed here, but have the greatest face validity (to assess what students have learned, assess, well, what students have learned). Direct assessment may involve analyzing course grades; administering standardized tests, performance tasks, or open-ended questions to assess general academic skills or subject matter knowledge; and obtaining data from evaluations of student projects or portfolios of student work.

Referring back to the earlier discussion of individual classroom level assessment, such forms of assessment reflect the ideal that many faculty embrace. To assess students' abilities in ballet, for example, one could count the number of toe shoes they've gone through, one could ask an external expert how strong the program is, and one could have the students complete a survey to capture what they think about their skills and how much they've grown; or, alternatively, one could have them actually *dance*.

Researchers tend to agree that this approach is a valid measure of students' abilities, but the use of one performance indicator may not be reliable. For example, a student may write an excellent term paper on one topic, but not on another, due to varying levels of motivation or interest in the topic. The key here is to ensure that enough information is collected, with an intentionality to the range of settings and contexts. While such approaches arguably provide *better* data, they are generally neither faster nor cheaper.

### **The Four Approaches Compared**

When it comes to understanding what students have actually learned in college the literature suggests that we are faced with a conundrum. There is general agreement that student learning is important and valued, but there is little (if any) agreement on how to assess said learning. Collection of actuarial data is commonly used because of the ease of data collection and the patina of scientific objectivity, but this approach equates quality with discrete, available (and perhaps most significantly, easily measurable) indicators of quality. Institutional rankings rely on a formula using actuarial data and ratings by informed experts, but these rankings are limited (and questionable) because they provide only an indirect measure of quality and because they tend to conflate quality and reputation. Student surveys use student perceptions of their learning, but research has shown that such measures may be problematic because they depend upon student self-evaluation; still, this research has been an important step in connecting student learning with educational quality. Finally, direct measures of student learning arguably have the greatest face validity with regard to assessing undergraduate education, but the literature indicates that there are numerous issues that complicate their implementation.

### **Another Methodological Note**

An important note is that the temporal design of any study will also shape the faster/better/cheaper analysis. A cross-sectional snapshot using any of these data sources has the advantage of being faster, but may limit the ability to make claims about changes over time in student learning in any value-added way. To study college impact using a pre- and post-test model will arguably produce better analyses (in that having two time points has clear advantages over a solely retrospective survey design), but this will neither be cheaper nor faster. Tracking specific members of a particular cohort of students can be even more expensive, but can provide richer data about student growth.

### So Which to Choose?

To return to our campus administrator, she most likely faces a particular version of this tug-of-war. She can meet two of the demands, but it's impossible to accomplish all three. Typically, she won't have unlimited time and she won't have unrestricted resources; to prepare for the looming accreditation visit or faculty committee meeting, she may have to sacrifice a better assessment in order to stay on schedule and under budget; this is perfectly understandable, and might make sense given the very real exigencies she faces. Further complicating the matter is, as noted before, there isn't universal consensus about what "better" means when it comes to assessing writing and thinking. Given this, the range of assessment approaches may be considered of equivalent quality, and thus essentially interchangeable; thus, the ones that are fastest and cheapest will win out.

The point to keep in mind, however, is that it's important for all on campus to recognize these trade-offs. A campus that focuses exclusively on cost and schedule must be prepared for the fact that some compromises in quality will be made, in just the same way that a campus that focuses on having a high quality assessment in short order must be willing to put its money where its mouth is. The key is to identify what is most important to the campus: doing work faster, better or cheaper.

I argue here that in the best interest of our students and improving academic programs, however, campuses should place a high priority on better. Quality should never be negotiable. If one completes analyses cheaper and faster, but there are no meaningful conclusions that can be drawn, the entire exercise is rather worthless. Having better data, conducting better analyses, and coming to better conclusions should be the key. Although it may cost more and take more time, an unspoken fourth dimension I would add here is the return on investment; this will assuredly be higher with better assessment. Better assessments have the potential to ensure that better programmatic choices will be made in the long run, that our understanding of student learning will be deeper and more accurate over the long haul, and that our ability to conduct more sophisticated and nuanced analyses will occur over the long range. Until the day comes when we can choose three, perhaps we should instead start by choosing one:

