

# Toward an Improvement Paradigm for Academic Quality

By: Douglas D. Roscoe

The assessment movement that has emerged on American college campuses over the last twenty years emphasizes the need to carefully articulate the particular outcomes we seek for our students, and it demands that faculty and administrators provide evidence of their students' success with respect to these outcomes. It also requires that this evidence be used to improve the educational experience of students in order to better meet those outcomes.

The transformation of the original assessment movement into *paradigmatic* status has largely been driven by the accreditation process. The regional accrediting bodies, perhaps reasonably concerned about heading off government-mandated standards and testing, now expect institutions to engage in student learning assessment at all program levels, from campus-wide general education down to majors and minors.

This paradigm, like all paradigms, has closed off questioning of key assumptions and has facilitated the ossification of certain practices that may no longer serve us well. My view is that we should consider breaking apart this paradigm by moving away from a focus on assessment and toward an emphasis on improvement. This shift will retain what was valuable in the assessment movement while paring away some of the dysfunctions that have arisen as it has become paradigmatic.

## **Costs and benefits of the assessment paradigm**

I have been deeply involved in this assessment paradigm from the beginning of my career over fifteen years ago. As a new assistant professor with skills in survey research and social science methodology, I was recruited to be my department's director of assessment, even before I fully understood the significance of the assessment movement (or the faculty's reluctance to embrace it). Since then I've directed departmental assessment every year, across two universities; sat on a college-level committee that reviews program assessment; directed the assessment of my university's general education program since 2008; and even worked on general education assessment as a Fulbright Scholar in Hong Kong in 2011. I have implemented a variety of assessments, including off-the-shelf standardized tests, homegrown standardized tests, rubric-scored student writing, curricular mapping, focus groups, and more. During all this

activity, I have come to the conclusion that the assessment paradigm provides a distinctive benefit to our students.

It also generates a number of costs.

Unfortunately, the assessment paradigm tends to assume the benefits and ignore the costs. This has created a number of dysfunctions in the ways we attempt to improve higher education. And let's be clear: improvement is the point. Assessment should be about changing what happens in the classroom—what students actually experience as they progress through their courses—so that learning is deeper and more consequential.

The dysfunctionality of assessment today starts with the primacy of evidence and data. One of the key premises of the assessment paradigm is that the faculty's conventional wisdom about what students can and cannot do well is unreliable. We therefore must collect direct evidence of students' abilities to master the outcomes that we define to be part of their educational process. As someone whose own political science research is primarily quantitative, I am entirely sympathetic with this premise. But good data must be available to answer our questions about student learning successfully.

The problem is that assessment data can be either cheap or good, but are rarely both. The fact is that good evidence about student learning is costly. For example, valid and reliable standardized tests are not easy to produce, and without a large staff to create one in-house, institutions have to buy one created by an educational corporation. Undergraduate Major Field Tests from the Educational Testing Service cost about \$25 per student tested. The Collegiate Learning Assessment by the Council for Aid to Education runs about \$35. Multiply that by the number of students necessary to get a reasonably representative sample, add the costs of paying staff to procure and administer the test, and we're talking about a fairly significant cost.

Standardized tests have their limitations in terms of what they can tell you about student learning, and most don't match up very well to the particular learning outcomes we have defined for our students. As anyone who has administered standardized assessments will tell you, it is a battle to get students to take them (unless required in a course) and then another challenge to get students to take them seriously. This tends to introduce serious sampling biases and validity problems with the resulting measures. And even if you do get a representative sample of students' best work from a high-quality test, it often gives you great

answers to the wrong questions. Institutions define their own learning outcomes, and standardized tests rarely match up perfectly.

Consequently, many institutions have turned to rubric-scoring of authentic student work. That is, student work produced for credit in classes is collected, and readers determine how well each student performs on a number of criteria connected to the defined learning outcomes. This has the advantage of capturing students trying to do their best and mapping their work directly onto the key outcomes. Unfortunately, creating good data from a rubric-scoring process is very difficult—and the availability of substantial resources makes it only slightly less so. The main problem is that scoring is a necessarily subjective process that requires all kinds of judgments about what key terms mean, how to distinguish between performance categories, and how to sort students' work into those categories. Calibration sessions that give readers training can be helpful, at least in promoting greater reliability, but may not help in establishing validity (e.g., everyone agrees what a "proficient" response looks like for scoring purposes, but is a student who wrote it really and truly *proficient*?). Moreover, it is often difficult to determine whether a student is unable to show mastery of an outcome or whether an assignment just didn't do enough to prompt the student to show that mastery. All these issues are complicated in general education curricula, where student work is coming from a variety of disciplines using an assortment of assignments that may not be commensurate.

Of course, content analysis methodology is used in social science research all the time, and there are ways to improve, if perhaps not perfect, these measurements. But these are costly. For example, double- or triple-coding helps, but now two or three times as many readers have to sacrifice the time (presumably at the expense of their other teaching, scholarship, and service obligations) or, for institutions that can afford it, stipends or course releases now go to that many more faculty. And as these projects get bigger and more complicated, the assessment paradigm has justified the creation and expansion of assessment administration offices. To do these kinds of assessments properly and produce high-quality data, it is probably necessary to hire staff with expertise in the methodology who can design and implement the process. Campus-wide administrators hired mainly to be "assessment directors" were rare when I started my career; they no longer are.

As enrollments stagnate, state appropriations dwindle, and campus budgets get squeezed, the opportunity costs of this kind of high-cost, high-quality assessment are sharply felt. Many faculty are reasonably upset when another assessment administrator is hired while a vacant

line in their department remains unfilled. Given accreditors' demands for assessment and the need to do more with fewer resources, campus leaders often have to choose between good assessment data or cheap assessment data. For a select few institutions, good data can be easily afforded. For most, it is a difficult cost to bear.

To make matters worse, even the best, most valid and reliable assessment data about student learning are often not really all that useful. At the same time, useful information that could improve practices is easily available without having to score or test our students separately.

To explain what I mean, consider a stylized example. Imagine a particular program has four defined learning outcomes. Student work is collected from a large random sample of students in the program, and readers score the work on each outcome. The resulting data need to be quantitatively summarized, given the number of cases. So averages might be calculated for each outcome, or the percentage scoring at or above a certain score is calculated. What do these numbers tell us? If the average on outcome one is 3.2 on a five-point scale, what does that mean? If 76 percent score at or above a 3, is that good or bad? Standardized tests create comparator benchmarks, at least, but rubric-scored work can't be judged objectively like that. More importantly, while the numbers might be used to show external audiences what our students achieve, what does it suggest we do to improve student outcomes? What is the actionable information from these data?

We might identify the outcome with the worst scores and focus on improving it. But likely all outcomes could stand improvement, and even without any of the assessment data it would be possible to guess about the worst one and save the expense and hassle. And if we guessed wrong, the worst-case scenario is that we improve one outcome instead of another that needed it more. But we would still focus on improving an outcome, so it is not exactly a disaster.

Even if the assessment data can be used to narrow our focus for improvement, they don't really tell us how to improve. To do that, we need other data that tell us about the educational experiences of the students. We need to know what classes students took, in what order, and what they did in their classes; we need to know what the assignments they completed were like and what the instructors did to support their learning; we need to know what kind of cocurricular experiences they had; and we need to know something about the individual students themselves, like their work habits, natural intelligence, attitudes about their

education, and mental health. These data, correlated with student outcomes data, would show us what works and what should be more broadly implemented.

Fortunately, these kinds of correlational studies already exist and show us what improves student learning. If we want to improve our students' ability to write or think critically, there is an enormous literature in the scholarship of teaching and learning to which we can turn. These are areas of scholarly expertise in their own right, and people devote their careers to careful measurement and to the testing of pedagogical research questions using sophisticated multivariate methodologies. Institutions do not need to create new data sets as part of an administratively imposed assessment process in order to get answers to our questions about how to teach our students better. Our faculty are already doing this work as part of their scholarship.

The reality is that the improvement piece of the assessment paradigm often takes a back seat to the collection of data. This manifests in the need for continual reminders to “close the loop”—that is, to use the data to change the educational experience of students.

In my experience, most of the value from assessment comes from closing the loop, but the data are useful mainly in that they provide the opportunity for faculty to have conversations about improvement. It is the requirement to discuss the data that provides an opening for fruitful dialogue about what is happening in our classes, what our students struggle with, what we are doing that works, and how we might change programs and courses to better steer our students toward the outcomes we are aiming them toward. It is remarkable how much improvement happens when faculty can carve out some time in their busy schedules just to talk about student success. The imperatives of scholarship and creative activity, teaching and grading, and service work leave little space on the plate for collective conversations about curricula and instruction. The assessment paradigm has been successful in demanding that these conversations take place on a regular basis.

To me, this has been the most important achievement of the assessment movement. The institutionalized collection of student outcomes data is really only a side note, only a part of the broader process—and a part that is far less important than it appears within the strictures of the paradigm. For this reason, I propose that we eliminate the assessment paradigm. In its place, we should embrace an improvement paradigm.



## **An improvement paradigm**

What would an improvement paradigm look like? How would it differ from the current emphasis on assessment? What would be its tenets?

First, an improvement paradigm would place at the forefront collective conversations about curricula and instruction. Faculty often want to discuss teaching and curricula, but this simply gets pushed off the plate. An improvement paradigm would create spaces where these discussions are required; it would wall off an area on faculty plates crowded with scholarship, teaching, and service demands. Put differently, this is about institutionalizing regular, serious faculty conversations about curricula and instruction. The assessment process, when it functions well, can and does create these conversations—in fact, that is what's most valuable about assessment today. But it would be better to put front and center the part of the process that is most worthwhile.

In practice, this would mean shifting what the campus administration and regional accreditors demand from faculty. Rather than requiring each department to identify an assessment guru who collects and analyzes data for the department, it would be far better to require regular department discussions about how to improve student learning. Deans might require a report of minutes from these meetings, rather than a report on what the assessment data showed. Indeed, the content of these conversations can be seen as the central product of the improvement paradigm (whereas in the assessment paradigm, student learning data are the product, which then must be used by faculty to close the loop). Data can be important in showing faculty how our perceptions about student learning can be incorrect, but the data must be viewed as only one component of the faculty conversation about improvement.

Second, an improvement paradigm would emphasize front-end intentionality over back-end assessment. Improvement requires changing what we do in ways we expect will help. The focus of faculty efforts to boost academic quality should be on what we think we need to change in order to make improvements. This is not to say that student assessment data are irrelevant. We can learn important things about how to improve by looking at our students' performance. But, as I noted earlier, intentional improvements can be driven just as successfully by professional research about teaching and learning. An improvement paradigm would ask faculty to rely on this research just as much as student learning data.

For example, the important work of George Kuh, using the National Survey of Student Engagement data, documents the powerful effect of the so-called high-impact practices on student outcomes.<sup>1</sup> In an improvement paradigm, faculty might begin with this research, which already establishes the efficacy of these practices without any new campus assessment data, and ask how they might implement service learning or learning communities or any of the other high-impact practices. Front-end intentionality might also be captured in a curricular mapping process, which involves an analysis of which courses in a program are intended to contribute to particular learning outcomes. If a curricular map shows only a couple of courses that, say, work to improve information literacy skills, then improvement is likely to occur if instruction focused on that outcome is pushed into other courses.

Ultimately, there are two key questions. What do we want our curriculum to do? What can we do to aim more squarely and effectively at those goals? Though this is a forward-looking perspective, it is also a reflective process. We can and should draw from what we have learned about students and their ability to master key learning outcomes—and there is a great diversity of information to draw upon.

This leads to the third key tenet of an improvement paradigm: a broadened view of what should be used to inform the improvement process. Collected data on student learning outcomes are fine, but they are far from the only useful input into a forward-looking improvement cycle and often have a low benefit-to-cost ratio. There are various alternatives that can be used: published research about instructional practices; curricular maps; surveys of instructional techniques and assignments; course and assignment grades (particularly if pieces of rubrics can be utilized); surveys of faculty opinions about student outcomes; focus groups; and, of course, faculty discussions about what our students struggle with and where they need help.

### **Prospects for a paradigm shift**

Scientific paradigms usually arise for good reason, reflecting broad, useful, and accurate understandings of the world. Paradigm shifts occur because new information or theories promise even better ways of understanding. Similarly, the formation of the assessment paradigm was perfectly reasonable at the time. The 2006 report of the Secretary of Education's Commission on the Future of Education (also known as the Spellings report) raised concerns about academic quality in US higher education. The commission concluded that colleges and

universities “must become more transparent about . . . student success outcomes, and must willingly share this information with students and families.”<sup>2</sup> The report called for a focus on innovation, recommending that “America’s colleges and universities embrace a culture of continuous innovation and quality improvement” by developing “new pedagogies, curricula and technologies to improve learning.”<sup>3</sup>

Following just four years after passage of the No Child Left Behind Act, which mandated state standards for primary and secondary education and standardized testing to assess student success in meeting those standards, the Spellings report sent shock waves through higher education. Would colleges and universities soon be mandated by the federal government to follow legislatively determined curricular standards and then to engage in high-stakes standardized testing to determine how well their students meet them? The prospect was scary for those who embrace the value of liberal education. It would be better, the thinking went, to get out in front and begin demonstrating voluntarily how well higher education prepares graduates.<sup>4</sup> Student outcomes assessment, already a growing movement among those interested in outcomes-based curricula, would become the key to these efforts. The focus on collecting outcomes data even eclipsed the Spellings report’s emphasis on innovation and improvement.

If it is not entirely accurate to say, a decade after the Spellings report, that the wind has gone out of the sails of standardized testing, it is fair to say the gale has softened to a mild, if consistent, breeze. There has been a considerable backlash against standardized testing in US schools, and the initial bipartisan support for No Child Left Behind has fractured on both sides of the aisle. Fights over standards and the pros and cons of testing are now more, not less, evident. Consensus has devolved into dissensus. A recent poll found that only 19 percent of Americans support the Common Core standards, with 54 percent admitting they do not know enough to have an opinion.<sup>5</sup> Almost two-thirds of Americans (64 percent) say there is too much emphasis on standardized testing in public schools.<sup>6</sup>

At the same time, Americans’ concerns about higher education continue to focus much more on cost than quality. Fully 74 percent of Americans agree or strongly agree that traditional colleges and universities offer high-quality education; only 5 percent disagree or disagree strongly. In contrast, only about a quarter of Americans believe a postsecondary education is affordable for everyone who needs it.<sup>7</sup> College costs are the third most commonly mentioned



concern when people are asked about the most important financial problem facing their families, and the top concern for adults under the age of fifty.<sup>8</sup>

So the time is right for a reassessment of assessment. We need to deemphasize the use of student learning data for external accountability and focus more on improvement based on whatever information is most useful and cost-effective. It is time to reclaim and remake the assessment cycle as an improvement cycle in ways that will benefit our students the most. It is time to embrace a paradigm of improvement.

## Notes

1. George D. Kuh, *High-Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter* (Washington, DC: Association of American Colleges and Universities, 2008).
2. The Secretary of Education's Commission on the Future of Education, *A Test of Leadership: Charting the Future of Higher Education* (Washington, DC: US Department of Education, 2006), 4.
3. *Ibid.*, 5.
4. The Voluntary System of Accountability, now part of the College Portrait, was one important result of this thinking. See <http://www.collegeportraits.org>.
5. CBS News, "2016: A Wide Open Republican Field, While Clinton Leads the Pack for the Democrats," CBS News Poll, March 21–24, 2015, <https://cbsnewyork.files.wordpress.com/2015/03/260281625-cbs-news-poll-2016-presidential-campaign.pdf>.
6. PDK/Gallup, "The 47th Annual PDK/Gallup Poll of the Public's Attitudes toward the Public Schools," Kappan 97, no. 1 (2015), [http://pdkpoll2015.pdkintl.org/wp-content/uploads/2015/10/pdkpoll47\\_2015.pdf](http://pdkpoll2015.pdkintl.org/wp-content/uploads/2015/10/pdkpoll47_2015.pdf).
7. *Americans Value Postsecondary Education: The 2015 Gallup-Lumina Foundation Study of the American Public's Opinion on Higher Education* (Washington, DC: Gallup), 23.