

Meaning More Than Measure: From Compliance to Commitment

Daryl G. Smith, Claremont Graduate University

The purpose of this conference, “Learning Without Measure: Assessing the Liberal Arts,” is important because it raises a number of salient questions, among them the ways in which assessment might be different in a liberal arts context, how to think about assessment, and the relationship between accountability, assessment, and accreditation. I would like to use my time to reflect on some of these issues and perhaps provoke some discussion about them.

Although I have been working in the area of evaluation and assessment for some time, I didn't come to it easily. Indeed, I think I can say I came to it with some resistance. Perhaps it was a reaction to so much of the early work that seemed to privilege tests. Perhaps it was the evidence of the limitations of standardized testing, especially high stakes testing, for the assessment of learning or anything. Perhaps it was my sense that talking about assessment and evaluation to groups of academics and academic leaders was a little like talking about a root canal. At the same time, years of accreditation work, thinking about my own need for institutional research to provide good information to inform decisions while I was an administrator, and my background as an empirical social psychologist meant that I could not simply dismiss the issue of assessment lightly and especially its connection to educational effectiveness and student learning.

What finally moved me from a supporter and user of assessment to my belief that this is essential to higher education was the research on diversity in which I have been engaged for years. About 10 years ago, I was involved in an effort that brought together evaluators from five national projects funded by Ford, Irvine, Kellogg, among others, to ask what it was that we were learning about the impact of these diversity efforts. As we met over a several month period, one of our key findings was that few of the campuses involved in our work had good information, let alone data, that could inform in any systematic way the efforts of the outside evaluators or the campus. We were thus left with information from long lists of project descriptions and interviews during campus visits. On one campus, which had invested a great deal of effort in getting faculty engagement with diversity, I asked what I thought would be a simple evaluation question: How many of the faculty had participated? Whereas I had been given names of seminars and activities, the campus itself did not know how many had participated. Because the campus was small, I asked if they might count. It turned out that over three fourths of the faculty had been deeply engaged in the seminar efforts begun during this initiative. This was an important and startling figure that was not even known to the campus. More problematic for all the evaluators was how little data was available to us or known to the campus about the graduation and persistence rates of students disaggregated by race and gender and whether there had been any change over time. For all of us, the absence of systematic information, and even pretty basic data, was a factor in the inability of campuses, or anyone else, to really know if they were making progress on diversity and in what areas.

This realization has led many who are concerned about diversity to begin to address the question of evaluation and assessment concerning diversity and educational effectiveness. For whom is the campus successful? What is being learned? Are students here thriving and which ones? For me and for many of us, the issue of assessment and evaluation is a core question about information and evidence (to use the Western Association of Schools and College's [WASC] term). How *do* you or *might* you know about the effectiveness of the institution in fulfilling your educational mission? The goal here is to describe, document, and understand what is happening with an eye toward improving educational practice. At its core these issues should frame the imperative of liberal arts institutions that, more than most any other sector, care about student learning and have a scale that makes deeper understanding and analysis more feasible.

The Challenges of Assessment

Part of the problem in bringing deep engagement to the issue of evaluation and assessment to higher education is one brought by my own field of psychology which in its effort to become a science has scientized and professionalized the fields of testing, assessment, measurement, and evaluation. Indeed, one of the barriers to effective work on our campuses has been the assumption that only experts in the science of evaluation and assessment can do evaluation and assessment. Responses on campus typically take the form of "I don't know anything about this area." People are intimidated by the scientific jargon of evaluation and assessment in the form of issues such as reliability, validity, and the language of statistics. Assessment, in particular, has been scientized to the detriment of a focus on institutional learning.

Another very important factor in the resistance to assessment, if there is resistance, is the current tendency (evidenced by the Spelling Commission and No Child Left Behind) to equate assessment with norm referenced and high stakes standardized tests, with cross institutional comparisons, and with efforts that are easily administered and scored. Whereas this may prove to be a challenge in any academic setting, it is especially challenging in the liberal arts context where reductionistic, simplistic, and homogenous approaches may seem anathema to the core values of the liberal arts.

However, it is my view that the current national and statewide movement for accountability has come in part because of the resistance at all levels of education to look at student success and learning even at the most basic of levels. Education has resisted the efforts to document how students are doing and what they are learning, or it has been inclined to simply blame the students, the family, the culture, and the lower schools for any lack of performance. For decades and decades, schools let themselves off the hook in terms of the performance of students from poor or underrepresented minority backgrounds. In those contexts, schools have also either not known or have been willing to excuse the blatant achievement gaps in ethnicity and gender. What is clear here is that the concern of most of the accountability movements was and is not really on learning outcomes at their highest but rather a focus on core indicators of institutional failure—

school dropouts and graduation rates—and the perceived unwillingness of institutions to do anything about them.

When Counting Makes Sense

Similarly, in higher education today, even the most basic questions where measurement and counting might be appropriate gets avoided to the peril of students, to the peril of the integrity of higher education, and to society. So if you will, let me take on the term measurement reflected in the title of this meeting and suggest that too many campuses don't measure what can be measured and don't use what they have measured or could measure to inform decisions. For me, as I go to campus after campus, most of them earnestly and urgently working on issues of diversity, and ask for things like disaggregated graduation and persistence data by race and gender, I find statements like, "We can't get that" or "The computer system doesn't do that," or I find tables and tables of data that require me to take my own calculator and ask whether this campus has an achievement gap at the most basic level of graduation, persistence, or grades. How many of you know where your campus stands on this issue? I have gone to presidents who didn't know. However, I can go to a wonderfully easy to use program at Ed Trust that has the data from one cohort of students of six-year graduation rates disaggregated by race for almost every campus in the country. Paradoxically, I can get these data because they are provided, through the Integrated Postsecondary Education Data System (IPEDS), by campuses. I did this for this meeting. Some of you appear to have substantial achievement gaps that are contributing to a national concern about accountability and performance. Some of you don't. For me the question is whether any of the campus self studies for accreditation, for example, are looking at these data, deciding if there is an issue, and deciding what to do about it. These data are at the most superficial of levels. What about our success or lack of success in the Science Technology Engineering Math (STEM) areas? What about time to degree for low-income students? What about levels of engagement on campus and for whom?

Yes, there is criticism of six-year persistence rates for places like state universities where people are older, attend part-time, and are non residential. If not six years, what would be better? Yes, the IPEDS only includes first-time, full-time students. But what stops us from knowing how our part-time and transfer students are doing? Yes, students may transfer and graduate elsewhere, but in terms of a given institution's effectiveness, dropping out is important for the institution as well as for the student.

It is my belief that simply criticizing all efforts to assess student learning and success and going on with business as usual is what inevitably forced government agencies to institute policy level efforts such as No Child Left Behind (NCLB). However, what NCLB did do was force schools to disaggregate and not hide the success or lack of students from all racial and ethnic backgrounds, and it laid as a foundation that it is not okay to simply ignore whole groups of students. Some liberal arts colleges in this country have graduation rates for African Americans that are 20 percent. Others have graduation rates for African Americans that are as good or better than for whites. Latinos on some campuses are succeeding very well, and some Asian students are not. Do you know?

Should we know? In reality, the accountability movement is primarily focusing on basic data—things like graduation rates. For example, a recent Association of Independent California Colleges (AICCU) report wrote about the movement toward accountability in California. The urgency of the movement was framed almost entirely on degree holders in the state, time to degree, and graduation rates disaggregated by race and ethnicity. This can be measured, but do you know?

What seems clear is that the accountability movement, accreditation, and effective assessment practices are linked but are not the same. Whereas the accountability movement is clearly being driven by public policy concerns for the inability of education at all levels to increase education for all, assessment is about improving educational effectiveness and the ability to provide feedback to students and/or institutions. In between, accreditation is the mechanism by which institutional quality and educational effectiveness is to be monitored by a peer process of institutional members who voluntarily undertake the improvement of education and the validation of standards. However, when our institutions do not hear and take seriously public concerns, it is through accreditation, or even more directly, that pressure will be placed on our campuses to do better, to do it quickly, and to do it publicly. If we are not careful, No Child Left Behind will be coming to our campuses, and there will be no opportunity to distinguish that which can be measured from those things which should not be measured but which should be assessed. Moreover, it seems clear that the compliance attitude (we have to do it and what do we have to do) helps get things going but interferes with really engaging in a way that benefits our campuses and students.

Assessing Outcomes that Matter

Having said that, I know assessment is not always easy or obvious, so I do want to spend a little time addressing the question of ineffability so powerfully raised by Peter Ewell (1991) some years back. I teach a graduate seminar that looks at issues of the assessment of learning outcomes at all levels of education. In that seminar, I ask students to take a learning outcome about which they care. That is, they should choose something that matters to them in terms of learning. Moreover, whereas graduation certainly is an indicator of student success, I encourage them to choose something for which data is not so easily available. What they and I learn as a result of that process provides some caveats about assessment:

1. The tools to assess the outcome of interest are not always available, although I would say that more and more tools are becoming available. For example, I just had the occasion to look at some of the work being done on the assessment of spiritual development.
2. One cannot be comprehensive in assessment. It would be impossible to comprehensively assess everything all the time given the level of intentionality and engagement now required of assessment efforts.
3. In addition, assessment will not be perfect, definitive, or highly scientized. Even if tools or instruments look scientific, they are still limited.
4. Be cautious about the language of value added, as its ability to be used is still limited. (Banta and Pike, 2007)

While some important outcomes can be measured, let us look at some others that suggest ways to look at assessment that are systematic and can be shared with others, but which are more holistic and nonreductionistic. I want to start with one of my favorite examples that was done in a large elite public research university in an effort to understand the learning outcomes for students who were being exposed to new pedagogies involving problem-based learning in chemistry. The motivation was in part defensive (defending the new approach) and in part the result of the concern of some of the faculty that time spent on content was being lost to process. As a result, they agreed that an assessment was needed. The chemists had spent a lot of time arguing and trying to agree on the traditional assessment question of what every student should know and should be able to do. They finally gave up because of lack of agreement and concern that coming to an agreement would stifle curriculum creativity. Instead, one of the chemists had what is to me a brilliant and elegant approach. The students had been taking introductory chemistry using the traditional methods and the problem-based methods pretty randomly. The chemists agreed that one of their goals was that their colleagues who taught upper level courses in chemistry or in fields that required introductory chemistry should feel good that these students were “prepared.” They randomly chose students who had taken the traditional or new chemistry and asked the faculty who received students to interview the students following the completion of the introductory course. They could ask any questions relevant to their concerns for their courses (no agreed upon content knowledge and no written test). They were then to rate the students holistically on a scale concerning preparation for their course. The analysis showed that students who had taken the problem-based course dramatically outperformed those in the traditional classes. The approach was transparent, relevant, and quite convincing to the faculty.

Another example is in the area of writing. Some of the concerns about accountability and student learning emerge out of a concern for some basic skills such as writing. For me, the question, especially in a liberal arts context, would be about levels of writing and the degree to which students (and which students) are writing at a complex and sophisticated level appropriate to the context in which they are writing. I would want to know what percentage of graduates is writing at what level and what strengths and weaknesses are emerging. How might we know about this? After I left Scripps, I thought often about the senior thesis exercise required of all students and thought it would not have been hard to develop or even to use existing rubrics that describe effective writing. For example, what percentage of the graduates were proficient writers? Where were their strengths and weaknesses as writers? Were students in certain fields more proficient than those from other fields? What percentage of the graduates had only minimal proficiency? Faculty colleagues could take a sample of the students’ theses, within fields or between fields, and evaluate the writing so as to answer the question about graduates. One could, if writing was an important issue, also look at students over time. It would not be hard to consider taking a subsample of students who enter and invite them to develop a portfolio of work that could be evaluated (for the student as well as for the institution) in terms of writing, critical thinking, and a number of other outcomes of interest. I know of one campus that did this to see whether students had developed in their capacity to engage the

complex issues of race and gender over their time at the institution. While sampling is an important way to avoid overloading the process, on some campuses that are smaller, and for some outcomes that are central or are known to be issues, one might even consider looking at everyone. This is the obvious advantage of smaller institutions.

A last example deals with the educational benefits of diversity. A few years ago, pushed by the Michigan case at the Supreme Court and the need for evidence, groups of academics came together with lawyers to develop the research which would answer the questions about the educational benefits of diversity and the learning outcomes that result from having a diverse student body. There was a great deal of attention to what outcomes would be relevant not only to the courts but also to Michigan's educational mission. I was a bit angry at first because it seemed to me obvious that there were benefits. Moreover, I didn't see anyone pushing for evidence about the benefits of technology in a way that would justify educational expenses for technology. At the same time, I knew that it had to be done. The courts wanted evidence, not just anecdotes. One of the outcomes assessed by Patricia Gurin and her colleagues (2002) at Michigan was cognitive complexity. As a result of thinking about learning outcomes from diverse experiences, a robust body of evidence now suggests that engaging with others from diverse racial and ethnic backgrounds improves a student's cognitive complexity—a quality quite consistent with a core value of the liberal arts. This work on cognitive complexity along with improvement in values and post-graduate behaviors has shed an important light on how diverse experiences influence learning by interrupting habituated thinking and suggests important findings for designing classroom and co-curricular experiences. The tools they used had already been developed and were quite useful and transparent.

The tools associated with assessment are growing. In addition, the use of rubrics is a wonderful way to develop meaningful categories that effectively take holistic and complex information and develop it into something that can be communicated to others and which can make meaning out of the information. Portfolio development combined with the use of rubrics to assess the portfolio is another wonderful way to see how students have progressed. These approaches are very interesting, realistic, manageable, and holistic.

There are also new developments in norm referenced or peer comparative tests like the National Survey of Student Engagement (NSSE) and the new Collegiate Learning Assessment (CLA) that can, if a campus wishes, be used to compare your campus to others. Their use places somewhat less burden on the campus, and they have qualities that merit their consideration. In the case of NSSE, data can also be disaggregated to look at the levels of engagement of different subgroups. I think there are always questions about whether the comparison to others versus the comparison to criteria established within the campus are more or less useful. I will say that the CLA, which is a writing and thinking exam, might be useful even if the peer comparisons are only heuristic. When I look at NSSE items, there is a sense that the questions are getting at things that matter like research and writing experiences.

Organizational Learning

It is clear that it is possible to ask and to develop a way to answer the question: How do we know that we are succeeding in terms of learning and student success? To what degree are the institutional practices and approaches to education well aligned with what we learn from our assessment activities? To what degree is the institution assuming some responsibility for this alignment of effort and results? The key phrase here is “how do we know this” and how engaged are we in linking good evidence to institutional improvement—in other words to the ways in which organizational learning can take place?

The paradox for higher education is this: All too often we resist these questions. We resist that they can tell us anything meaningful. We resist a culture of evidence about the core activities of student learning and educational effectiveness. At the same time, we are institutions of higher learning. We value (almost too much) the approach of science. We ask students in every academic writing exercise to document findings, sources, and claims with evidence. We ask for this but don't often enough do it ourselves.

For me it seems clear that the answer to these questions can and will emerge if we frame assessment not as a reductionistic effort or one that simply focuses on those things that are easily measured or as something that requires compliance. It turns out that whether in the assessment movement or in the efforts to monitor progress on diversity, taking an organizational learning approach can be quite a useful framework that focuses on systematic efforts to document and learn about the institution and to use those efforts to improve educational effectiveness. It is also clear to me that organizational learning is an approach that is well aligned at least in theory with the culture of academe. What is not so clear to me is whether enough of our institutions of higher learning can or will become learning organizations. With its focus on learning, with its pattern of faculty involvement, and with its scale and mission, liberal arts colleges can and I hope will become the model for organizational learning concerning learning.

In developing such an approach to assessment, there are a number of key elements to consider. First, embed the effort in core institutional processes. Program reviews, institutional self studies, capstone courses, and the use of existing institutional information are all ways to build assessment into ongoing processes and potentially to get greater meaning from these processes. Further, accreditation, as it is currently defined, provides an incredible opportunity to make something that could be an exercise of compliance into an exercise of commitment. Second, keep assessment manageable, strategic, and focused. One cannot assess everything all the time. Building on what is available and making use of it is key. However, one can make good use of what is already available and build from there. Third, it is clear that too many of our campus efforts these days run on parallel tracks. Why have a planning process for diversity efforts that is not deeply embedded in one's accreditation efforts concerning educational effectiveness? So often accreditation efforts, strategic planning efforts, and diversity plans do not engage one another. Fourth, the questions about assessment are central to the educational core of the institution and must engage, excite, and empower faculty.

Finally, there are many models and tools being developed on other campuses. Assessment does not and should not require reinventing the wheel. It is extremely important to benefit from many of the excellent ideas being developed throughout the country. Indeed, building on and referencing the work of others is a hallmark of good scholarship, and it certainly should be a hallmark of good assessment. The Carnegie Foundation's Scholarship of Teaching with an emphasis on peer review, knowledge of the scholarship of others, and sharing information does seem to apply. Parenthetically, as I review some of these tools I am not certain that liberal arts colleges would necessarily find them irrelevant simply because they were developed at another kind of institution. I do think, however, that liberal arts colleges have opportunities to do some things that other institutions cannot, including focusing on distinctive aspects of the mission, addressing outcomes more holistically, and perhaps studying outcomes in ways that link the academic and nonacademic aspects of the campus such as leadership development.

In a recent *Change* article (2007), Lee Shulman suggests that rather than being a victim to the intrusive and potentially damaging calls for accountability, rather than resist the currents of the time, rather than just go with the flow as minimally as compliance will allow, higher education should do what white water rapids folks do—they should paddle faster than the current and get ahead of it so that they are not forced into practices that don't fit. I think, had the K-12 sector gotten ahead of the challenge of the achievement gap, other ways of addressing the issue might have been developed that would have had greater alignment with educational values. Indeed, I think that rather than reacting to accountability issues, campuses should seize the educational challenge of systematically assessing student learning. Our challenge today in higher education is whether our dear institutions of higher learning, especially liberal arts colleges so invested in student learning and higher order values, can be learning organizations—organizations that develop a culture of evidence, make meaning out of the information, and tie improvement to that information. This is not a model of proof, causality, and reductionism, and it is not perfect, but it will help us know how we as institutions are doing and how our students are doing. This cannot and should not be about compliance. It should be about commitment to educational effectiveness and learning and *meaning more than measure*.

Works Cited

- Banta, T.W., and G.R. Pike. (2007). "Revisiting the Blind Alley of Value Added." *Assessment Update* 19(1): 1ff.
- Ewell, P. (1991). "To Capture the Ineffable: New Forms of Assessment in Higher Education." *Review of Research in Education* 17: 75-125.
- Gurin, P., E. Dey, S. Hurtado, and G. Gurin. (2002). "Diversity and Higher Education : Theory and Impact on Educational Outcomes." *Harvard Educational Review* 72: 330-366.
- Shulman, L.S. (2007). "Counting and Recounting." *Change* 39(1): 20-25.