



AUTHORS

Charles Blaich, Ph.D.
*Center of Inquiry
at Wabash College*

Kathleen Wise, MB.A.
*Center of Inquiry
at Wabash College*

The More Things Change, the More They Stay the Same: New Challenges to Using Evidence to Improve Student Learning

“**W**hy so much collection—but so little utilization—of data?” (Blaich & Wise, 2011). This was the question we grappled with five years into the Wabash National Study. We created the Wabash National Study in the early 2000s as a response to increasing pressure for institutions to get serious about assessment. The Wabash National Study was a multimillion dollar, 40-plus-institution, longitudinal research and assessment project that measured liberal learning outcomes and the good practices that promoted the development of those outcomes.

The study assessed students at three points: when they entered college, at the end of their first year, and at the end of their fourth year. In addition to sending detailed summary reports and analyses to institutions after each assessment, we offered to combine, at no cost, data from the study with additional student data from institutions so that researchers at the institutions could learn more about what factors promoted learning for their students. We also held workshops and visited participating institutions, at no cost, with a focus on using data from the study for assessment.

We thought of the Wabash National Study as an “assessment test kitchen.” We thought that people resisted assessment because they didn’t understand what they could learn about their students from high-quality data. We hoped that providing institutions with such data would demonstrate the potential of assessment for improving student learning. By research standards the Wabash National Study was a success. The study led to hundreds of presentations and journal articles, and many theses and dissertations. But as a model of high-quality assessment the Wabash National Study was a bust.

CORRESPONDENCE

Email

blaich.cila.heds@icloud.com

We learned early on that few people at participating institutions were downloading and reading our carefully-crafted reports. And when we visited institutions, bringing what we thought were useful insights from our analyses, we often found that similar insights had already been identified in a report from an earlier project—a report that had, in many cases, come and gone without note. From a research standpoint this was good news. We were confirming findings of earlier institutional research. But from an assessment standpoint it was bad news. Most of the institutions were not using data from the study, or their own data, to drive improvements in student learning. And thus, our question, “Why so much collection—but so little utilization—of data?”

When does data make a difference?

A few institutions in the study countered this trend. What differentiated these institutions from the institutions that were only “assessment curious”? The lessons we learned from these more active institutions are simple and remain true. Assessment evidence on its own doesn’t lead to improvement. For assessment evidence to make a difference it has to address compelling questions that people have about student learning. Institutions that leveraged the Wabash National Study to improve student learning were institutions where data from the study addressed issues about student learning that faculty, staff, and academic leaders cared about. These issues, whether they were about academic challenge or critical thinking, were “in the air” at these institutions. Moreover, influential people were willing to commit their time, energy, and political capital to use data from the study to advance work on these issues.

Those instances in which we discovered data that people on campus already knew, but had ignored, were also instructive. These ignored findings were usually the result of routine data collection that was not driven by any compelling question. For example, one Wabash Study campus we worked with routinely collected data on the proportion of students who were employed on and off campus and another administered the same first-year student survey every year for over 30 years. In each case, someone would write a memo summarizing the results, distribute the memo, and that was the end of the story. The act of writing and distributing the data summary wasn’t connected to relevant conversations on campus. We often see the same thing with assessment today. Sometimes such data are collected routinely for compliance reasons, other times they are collected out of habit. We work with institutions that administer four different student surveys on a four-year cycle. The results are written up, distributed, and filed. When we ask people why they do this they reply that even though these kinds of routinely collected data aren’t connected with any urgent questions they can serve as a “dashboard warning light” to ensure that nothing is amiss. In other words, they help keep an eye on things.

Most of the institutions were not using data from the study, or their own data, to drive improvements in student learning. And thus, our question, “Why so much collection—but so little utilization—of data?”

What is the goal of assessment?

Consider the following departmental assessment plan. To ensure that its majors have acquired sufficient knowledge in the discipline, a biology department requires a sample of its graduating seniors to take a nationally administered, standardized exam (the ETS Major Field Test) on general biology. The department’s goal is that, on average, students in the sample will score at the 75th percentile on the test; as long as they do, the department takes no action. If students’ average score dips below the 75th percentile, the department will consider next steps.

This assessment plan is designed to keep an eye on things and make sure they are okay. However, it is designed in a way that can easily lead to improvements in student learning. Dips in Major Field Test scores might prompt inquiry into what’s behind those declines, and that inquiry might lead to changes that benefit student learning. On the other hand, dips in exam scores might also lead to conversations about whether a different test would be a better measure, whether this year’s cohort was a “bad class,” whether it would be prudent to test a few more cohorts to make sure the trend holds, or whether it’s time to rethink the 75th percentile criterion. Such responses to falling short of the standard might even be called “closing the loop” in an assessment report.

Sometimes “keeping an eye on it” assessment is done so that a program can focus on other ways of improving student learning. But sometimes it’s all that programs do. At some institutions, the need to fill up accreditation reports can exceed the patience and support necessary to ensure that programs engage in assessment that’s designed to improve student learning. For example, at a recent conference we heard an assessment director say that she appeals to programs that ignored requests for assessment reports by saying, “All I want from you is something I can aggregate and roll up into our overall reports.” While this may be a useful strategy for gathering information for accreditation, it removes student learning from the process, thereby making assessment less relevant for faculty and staff who care about what and how their students are learning.

We think using rubrics to evaluate student work shows great promise as a form of assessment that can improve student learning. But we’ve also seen rubrics used for assessment in ways that barely make the keeping an eye on it standard. At a recent conference we heard a presentation about a rubric-based general education assessment process that asked faculty in general education courses to 1.) choose, on their own, a general education outcome they thought their course contributed to; 2.) select at least 2–5 pieces of student work from their course where students demonstrated that outcome; 3.) evaluate those artifacts on a four-point, one-dimension scale for the outcome (“does not meet” to “exceeds standards”); and 4.) submit those scores to the assessment director. The assessment director then summarized scores by outcome and posted them on the institution’s website.

Why this approach? Assessment leaders at this institution wanted to use rubrics and student work because they thought it would engage faculty. They also had no resources for faculty training, norming sessions, or other work that might improve the value of this process. Finally, they were gearing up for accreditation and needed to implement a general education assessment process as soon as possible. Admittedly, anything that gets faculty to think about how students engage their assignments is a good thing. But without more developed rubrics, norming, a better process for looking at the alignment between assignments and outcomes, or systematic plans for engaging faculty in sustained conversations and responses to the findings, this approach has little chance of systematically improving what students are learning in their general education courses.

Of course, a better designed and resourced use of rubrics can also result in keeping an eye on it assessment. A recent post on an assessment listserv described a project to assess the impact of experiential learning programs for an institution’s upcoming accreditation. Students in these programs were required to complete pre- and post-program reflective essays. A stratified random sample of these essays was collected and scored by a team using one of the AAC&U VALUE rubrics.

This assessment was not designed to help program leaders improve the impact of their programs. It was designed to answer the question, “Overall, is experiential learning making a difference?” This is the kind of keeping an eye on it question that accreditors want answered. But improvements in student learning come from changes in what students encounter in specific courses, experiences, or programs—not from courses, experiences, and programs in general. In addition, this experiential learning assessment process doesn’t provide information that people can readily use to either assess or improve student learning in their particular programs.

Interestingly, at the end of this post, the author reported that they had also implemented a fidelity survey to ask students what they experienced and learned in their specific programs. The author stated, “We figured that this indirect measure would be both helpful to us to see whether what we think we’re doing is actually taking place, and *it might also provide useful information to the faculty and staff who are involved in offering high-quality ELOs.*” (emphasis added) So, despite the care and effort that went into the rubric work, a survey was the measure that was seen as providing useful information to faculty and staff.

Sometimes such data are collected routinely for compliance reasons, other times they are collected out of habit.

At some institutions, the need to fill up accreditation reports can exceed the patience and support necessary to ensure that programs engage in assessment that’s designed to improve student learning.

Conclusion

If many assessment programs are aimed at monitoring rather than improving student learning, the lack of broad evidence indicating that assessment improves learning should not be surprising.

We have no quarrel with using assessment to make sure that things are okay. But it's worth considering how often people use the term assessment to refer to processes of collecting, making sense of, and acting on data related to student learning, and then testing to learn whether those actions had the intended effect, or whether they are referring to keeping an eye on it data collection.

Fulcher, Good, Coleman, and Smith (2014), and Brown and Knight (1994), have correctly pointed out that learning doesn't get better just because you measure it; and assessment that's designed to keep an eye on it is assessment that focuses on measuring things. So even though there's more assessment happening now than ever before (Jankowski et al., 2018), our work with institutions today sometimes feels like déjà vu all over again, with so much data collection, but so little of it done in a way that's structured to readily improve student learning.

Perhaps this is what's behind the frustration about assessment that has bubbled up recently in *The New York Times*, *The Chronicle of Higher Education*, and *Inside Higher Education*. As Molly Worthen (2018) put it in her editorial, "All this assessing requires a lot of labor, time and cash. Yet even its proponents have struggled to produce much evidence—beyond occasional anecdotes—that it improves student learning."

If many assessment programs are aimed at monitoring rather than improving student learning, the lack of broad evidence indicating that assessment improves learning should not be surprising. Nor should it be surprising that faculty and staff, many of whom believe their classes, departments, and programs are doing well, might find assessment aimed at keeping an eye on it to be pointless. We're not arguing that all courses, departments, and programs are accomplishing their goals for students. But if we're selling assessment on its potential for improving student learning, it's probably time to consider the extent to which we're overpromising and underperforming, and what we can do to diminish that gap.

References

- Blaich, C. F., & Wise, K. S. (2011, January). *From gathering to using assessment results: Lessons from the Wabash National Study*. (Occasional Paper No. 8). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Brown S., & Knight, P. (1994). *Assessing Learners in Higher Education*. London: Kogan Page.
- Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014, December). *A simple model for learning improvement: Weigh pig, feed pig, weigh pig*. (Occasional Paper No. 23). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Jankowski, N. A., Timmer, J. D., Kinzie, J., & Kuh, G. D. (2018, January). *Assessment that matters: Trending toward practices that document authentic student learning*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Worthen, M. (2018, February 23). The misguided drive to measure 'learning outcomes.' *The New York Times*. Retrieved from <https://www.nytimes.com/2018/02/23/opinion/sunday/colleges-measure-learning-outcomes.html>