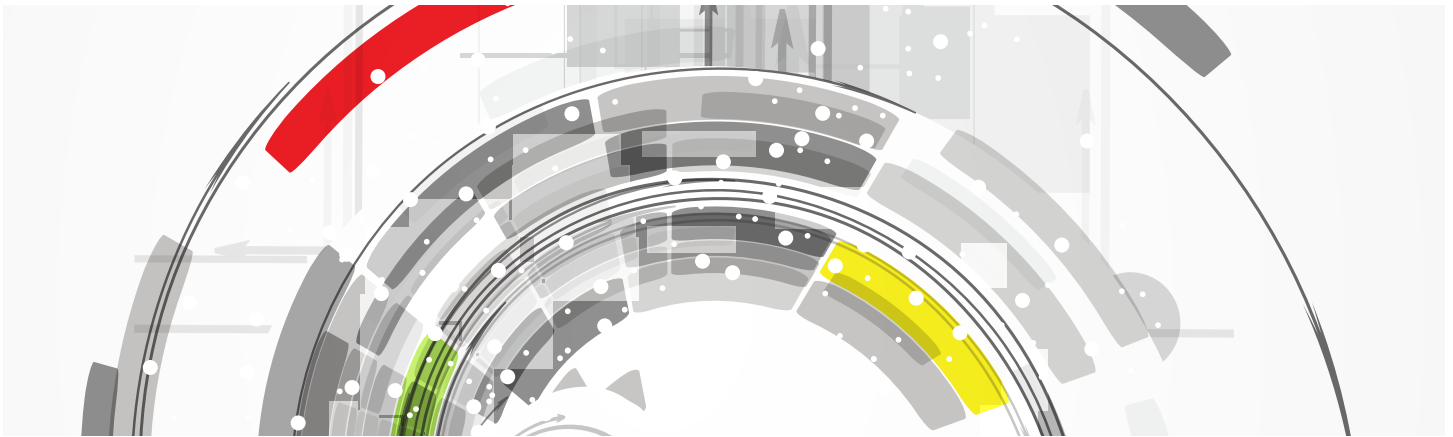


The Next Ten Years: The Future of Assessment Practice?



Forward

In these times of extraordinary change and hardship, there is perhaps no better time to consider how we as assessment professionals might reimagine our established practices. This is exactly what the RPA editors had in mind when we reached out to two visionary thinkers in the field with the following question: “If you are given unrestricted power to change assessment practice for the better over the next 10 years, what does assessment look like? What changes would you make over the next 10 years and why are those changes needed?” Both David Eubanks and Keston Fulcher responded to this prompt with thoughtful and telling insights into how we might improve assessment. While the two may disagree on some points, they both agree that we can do better. We invite you to explore their visions of the future of assessment along with an integrative response from RPA associate editor, Megan Good. We hope that these thoughts spur you to consider your own vision of assessment.

We also wish to note that this is purposefully an incomplete conversation. *Research & Practice in Assessment* is partnering with the IUPUI Assessment Institute to produce a podcast where David, Keston, and Megan will continue this engaging conversation about the future of assessment.



To listen to the podcast visit: <https://assessmentinstitute.iupui.edu/overview/podcast-episodes.html>

Forward by RPA Editor-in-Chief, Nicholas A. Curtis, Ph.D.

Authors

David Eubanks, Ph.D.
Furman University

David Eubanks holds a doctorate in applied mathematics and has served at four colleges and universities over a 29-year career. Since 2015 he has served at Furman University as Assistant Vice President for Institutional Assessment and Effectiveness. His 2016 article “A Guide for the Perplexed” challenged the practical usefulness of standard assessment practices for accreditation reporting and led to an ongoing conversation within the higher education community.

Keston Fulcher, Ph.D.
James Madison University

Dr. Keston Fulcher is the Executive Director of the Center for Assessment and Research Studies and Professor of Graduate Psychology at James Madison University. Keston’s research focuses on structuring higher education to foster learning improvement at scale. Keston’s larger work endeavors to help higher education transform from a “culture of assessment” to a “culture of learning improvement.”

David Eubanks, Ph.D.
Furman University

This essay is a summary of my conclusions from working in assessment, accreditation, and institutional research capacities for two decades, as a practitioner, researcher, and peer reviewer. It reflects disillusionment with the rigid, almost dogmatic, restrictions inherent in accreditation-style reports that are intended to demonstrate the quality of academic programs. Criticisms of these reports have been widely reported, most recently starting with my (2017) article about methods, and followed by public statements and articles from luminaries (Lederman, 2019).

Measuring student achievements and using that information to improve academic programs is a perfectly fine idea; it's not that this goal is unreasonable or impossible, merely that it cannot be turned into a checkbox bureaucracy. In attempting to find a middle road between improvement and accountability, compliance standards for institutional accreditors have accomplished neither.

Using the SACSCOC standard 8.2 as an example, assessment reports are explicitly data projects: "The institution identifies expected outcomes, assesses the extent to which it achieves these outcomes, and provides evidence of seeking improvement based on analysis of the results." Accordingly, assessment offices churn out dozens or hundreds of reports a year, most with small samples of poor or untested data that are subjected to rudimentary analysis, so that—at best—the only conclusions that can be drawn are from an average or from a proportion being "too low."

In attempting to find a middle road between improvement and accountability, compliance standards for institutional accreditors have accomplished neither.

Regardless of whether a finding is possibly random, action is required, often resulting in anemic changes like "we added more critical thinking content to the syllabus." Accreditors complain about this "checkbox" reporting (CHEA, 2019), but seem unaware that the standards in place practically ensure that efforts won't pay off. This fact is probably due to a cadre of consultants and peer reviewers who continually reinforce the rules. As if the *system* is perfect and our local problems are due to our own lack of perfection. That's certainly what I thought for years: that if I just did exactly what the consultants and accreditors were describing, a flood of insights about student learning would follow. It's embarrassing to admit that it took years to realize that the same principles I was teaching in Statistics 101 applied to assessment reports too.

The way forward is to combine all the information we have and use the best methods available. In particular, it means overcoming the prejudice against grades. Accreditor's rules vary, but for most institutions, course grades are

considered invalid as primary measures of student learning, which leads to the need for a whole second set of books: rubric ratings of papers and so on. It's wasteful and ineffective to have two disconnected systems—accreditation reports and course grades—that have the same goal of assessing student achievement. We need a single integrated system with the goals of (1) improving success for all types of students, and (2) ensuring that transcripts are meaningful, both for individual courses and for degrees. For examples of the first of these, see the 2020 webinar put on by the United States Department of Education, "Predictive Analytics to Improve Student Outcomes."

A unified system uses measurement methods, meaning large samples of data gathered under similar conditions and tested for reliability and validity, but only when suitable. For everything else, we should trust faculty, who are the experts on their classes and students. Course grades lie in the intersection of those two sets: they represent a summative faculty judgment after seeing student work over a period of weeks, and there is usually a grade for every class a student takes, connected via student IDs to hundreds of other data points. As such, it is straightforward to evaluate the characteristics of grades, including reliability overall and within programs (Beatty et al., 2015), and instructor or program "leniency" (Millet, 2018). These assess the fairness of grading, which can be improved through feedback (Millet, 2010). One can look for courses that block students from curricular pathways or predict drop-outs, and include demographics or other factors as explanatory variables to identify systemic biases.

Grade validity is more difficult to assess than reliability (as in intra-class correlation), but taking the question seriously breathes new life into the assessment project. Because of the richness and completeness of the data, there are numerous strategies to try. For example, can a factor analysis of grades associated with course prefixes (e.g., BIOL or ENGL) extract dimensions that plausibly associate with domain knowledge? At my institution, the answer seems to be yes—we can distinguish "humanities" skills from "math" skills using grades.

Grades generally only show student development qualitatively, through the courses taken, since per-student grade averages tend not to change significantly over time. That makes assessment measures on a developmental scale attractive as a complement to grades. For example, an institutional study of student writing that can measure growth over time is a project that can benefit all programs. I co-authored a validity study of our work along those lines, where we found evidence of differential growth related to grade averages (Eubanks & Vanovac, 2020). The findings suggest that educational opportunities are not equally accessible to all students. Without the combination of grades and assessment data, these insights would be lost.

The framework I've described here, combining the official records of student achievement (grades) with

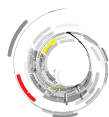
complementary high-quality research, eliminates the need for assessment reports that attempt to improve learning outcomes one by one. That business of writing down learning outcomes, finding a plausible data element to match, and so on, has a giant plot hole: there are a lot of learning outcomes in a college degree: far too many to treat that way.

By my count, using section headings from a textbook, a first calculus course has 30-40 substantial learning outcomes, just in that one course. Focusing on individual outcomes is the wrong way to go about it; it makes much more sense to increase faculty teaching ability in general through faculty development, including pedagogy and assessment. A faculty member who notices a problem with a learning outcome while a class is going on and fixes it right away raises the level of learning generally, and there's no multi-year lag between noticing the problem and fixing it.

You've probably noticed that my description of fixing assessment means essentially doing the opposite of what consultants and accreditors have been advocating or requiring for decades. Indeed, that's my general rule of thumb by now: if it's considered a "best practice" in accreditation reporting, it's probably the reverse of what you should do to get results. I recommend that you test this for yourself by asking for evidence to back up claims.

The main purpose of ensuring program quality—the reason for the accreditation requirements—will never be achieved with checkbox-style reporting. Our students and institutions deserve better, and the need is urgent.

None of the foregoing takes away from the work that assessment staff does to support academic programs by helping with curriculum and course designs, thinking through what students are expected to learn, designing tests and marking processes, and so on. The good that has come from the assessment movement is undoubtedly driven in part by the accreditation mandate. However, the main purpose of ensuring program quality—the reason for the accreditation requirements—will never be achieved with checkbox-style reporting. Our students and institutions deserve better, and the need is urgent.



Keston Fulcher, Ph.D. James Madison University

Twenty years ago, as a graduate student, I believed that higher education was a tightly conceptualized, rigorously executed enterprise. From a seat in the classroom, I felt the vast majority of professors were passionate about their subject areas and cared deeply about student success. If asked how I would make higher education better, I would have scratched my head. More parking, perhaps?

Ten years ago, as an early-career assessment professional, I was in a better position to think about the efficacy of higher education. My gaze narrowed in on academe's ability to foster student learning. Shouldn't learning be the most prized outcome of higher education? Don't we want students to have the knowledge, attitudes, and skills that prepare them for a successful career and a meaningful life? My positive view of higher education persisted. Nevertheless, the post-secondary sheen did not appear as bright. I began realizing that programs were not perfect. They could be more effective if tweaked. And, of course, the mechanism for tweaking could be nothing else than solid assessment practice (said from the myopic lens of an assessment professional).

I realized that the requirements for assessment, through accreditors and internal college policies, promoted a checkbox mentality. Assessment often was treated as a bureaucratic chore to accomplish rather than a mechanism for real change. If I were asked a decade ago how to make higher education better, I would have suggested more attention to assessment and more rigorous methodology. My assumption was that if higher education professionals had access to better assessment data, then they would use it to improve the enterprise!

Between then and now, several assessment insiders, including my team, have rejected this assumption. It turns out that assessment, even conducted with pristine methodology, rarely catalyzes improvement efforts. Blaich and Wise's (2011) excellent work on the Wabash Project shined a bright, expensive light on the misunderstanding. They, too, believed learning improvement would be propelled if institutions could access robust assessment methodology. Two million dollars later, the team had helped dozens of institutions gather trustworthy data but found little evidence of use, much less evidence of improved student learning. In other words, Blaich and Wise debunked the Copernican-like-view that assessment lies at the center of the learning improvement universe. Similarly, at James Madison University, we began examining our assessment reports across time. Over the years, we had provided assessment guidance and support to academic programs. And, the work appeared to pay off. Almost all areas of assessment were demonstrably better. The exception: the use of data for improvement. Ugggh. It was the Blaich and Wise finding at a smaller scale.

Today, I think about the bigger question. What would it take to improve student learning at scale? I also believe that higher education's treatment of student learning does not need a tweak. It needs an overhaul. My colleagues and I are working on a new model that centers on faculty, staff, and administrators. A model where higher education is structured to make evidence-informed changes to the learning environment.

It turns out that assessment, even conducted with pristine methodology, rarely catalyzes improvement efforts.

For the next 10 years, our vision is to improve higher education by inspiring and empowering faculty, staff, and administrators to make evidence-based decisions to enhance student learning and development. Enhancing (i.e., improving) student learning and development is explicit in the vision. If we don't see it, we haven't achieved it. Furthermore, the lever to achieve improvement is empowering faculty and staff to make evidence-based decisions. Any effort to improve student learning flows through these on-the-ground educators. It's this empowerment that we are putting our mouths and our money behind.

While not explicit in the vision statement, our strategy for empowering faculty is through professional development. In other words, we believe that a major obstacle to improved learning at scale is a lack of knowledge and skills related to evidence-based decision making. Professional development will help higher education push through this considerable obstacle. If educators can make better decisions, they can make informed changes to the learning environment that can foster better student learning.

We are not abandoning high-quality assessment designs, far from it. However, we argue that without the skills to make evidence-based decisions, faculty and staff will get little use out of quality data. It's like tossing car keys to an unlicensed driver. Educators need time, guidance, and practice to make evidence-based decisions to enhance student learning and development.

I think it is possible for every institution in the United States to achieve at least one example of program-level learning improvement by 2031. For example, a biology program could show that a future cohort of students could write better than a previous cohort because of targeted writing interventions. Operationally, what would it look like to "improve learning" in 10 years?

First, all of higher education, and particularly assessment professionals, must abandon the notion that the lack-of-improvement problem can be solved merely by developing better assessment methodology. Tweaking rubrics won't get us there. Better sampling designs won't get

us there. "Big Data" analyses won't get us there. These are all useful but insufficient tools to improve student learning.

Second, we must think of assessment as part of a larger learning system, a system where assessment is integrated with components that influence the learning environment (e.g., program theory, evidence-based interventions, implementation, and change management).

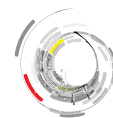
Third, if the end game is better student learning, we should create professional development around that notion. In addition to developing assessment practitioners/professionals and faculty developers, shouldn't we be developing learning system coordinators?

Fourth, and this is the toughest. We need to flesh out, in great detail, the needed knowledge and skills for higher educators to be evidence-based decision makers in a learning systems framework. In other words, how do we help faculty and student affairs professionals prepare to lead successful learning improvement efforts? I suspect that if leaders in assessment and faculty development put their heads together, we could make great strides in this area.

Finally, and this is not mentioned explicitly in the vision, we need to provide guidance to administrators. Deans, provosts, vice presidents, and presidents are the leaders that set priorities and create infrastructure by which initiatives happen. They, too, need guidance to create an environment by which professional development flourishes for on-the-ground faculty and student affairs professionals.

I believe that the future of better assessment in higher education ironically calls for its subordination; subordination under a larger learning systems framework.

In sum, I believe that the future of better assessment in higher education ironically calls for its subordination; subordination under a larger learning systems framework. The path forward includes articulating this new framework and providing professional development around it.



RESPONSE BY:
Megan Good, Ph.D.
RPA Associate Editor

“If you are working on something exciting that you really care about, you don’t have to be pushed. The vision pulls you.” – Steve Jobs

Does assessment have a vision? At best, one might clumsily say the vision is to improve student learning; when pressed on what the vision looks like, one might falter. Two trailblazers among us have defined tangible visions – Dr. Keston Fulcher and Dr. David Eubanks. At first, I thought the two perspectives were wildly different. But on reflection, I see great commonalities. Here, I will note the similarities and comment on each perspective.

The primary similarity between Eubanks and Fulcher is a feeling of disillusionment with our current assessment practices, both mentioning forms of bureaucracy. If the vision is to ‘improve’ student learning and achievement, our current system essentially isn’t working. We need space for more innovative work that may truly impact students. To that end, both authors share a sense of optimism for the future. Finally, each visionary incorporates educational development into their dreams.

The primary similarity between Eubanks and Fulcher is a feeling of disillusionment with our current assessment practices

Both authors are creating new systems. However, their systems look quite different. In Fulcher’s vision, assessment practice seems to remain largely the same (presumably, there are still student learning outcomes, curriculum mapping, and aligned measurement components). Assessment practice, however, becomes subordinate to educational development in the new learning system. The vision behind Fulcher’s learning system is clear - achieve evidence of improved student learning. I love that educational development is front and center. Indeed, it is clear that faculty and administrators need assistance in improvement efforts. However, I wonder – would Fulcher’s future still mandate the “dreaded” assessment reporting requirements? And I wonder what evidence of learning improvement would achieve? It sounds nice! But senior administrators are not bothered with this metric now, why would they be in the future? The public is certainly not demanding it (though the COVID-19 pandemic may change this).

Dr. Eubanks’s future creates an entirely new system. It seems assessment is replaced with a more robust analytic system using as much data as possible (notably including grades) to understand student learning in real-time. The analytic system would yield findings for faculty, administrators, and educational developers to improve. Students could be helped when they need it. In such a system, it seems the workload could shift from the faculty to the

analysts, creating space that might be used for improvement conversations. And, although I was trained that grades are not assessment measures, I recognize that using them as Eubanks described could be powerful and create new faculty partners. Finally, I appreciate that Eubanks has acknowledged educational development’s role in the new system (though perhaps to a lesser degree than Fulcher). I have one primary question/concern for this future – how do we know that the new robust system doesn’t yield reports that sit on a shelf like current assessment reports? In the new system, why would faculty and administrators be any more motivated to apply these new data to drive change?

Perhaps there could be a hybrid vision that combines Eubanks’s and Fulcher’s perspectives. I imagine Eubanks’s analytic system could replace assessment and could exist within Fulcher’s learning system. In this way, the data would be different, but the primary focus would still be on improvement with educational developers central to the system’s success. Regardless of the future, these visionary pieces have left me excited for the possibilities to change higher education for the better.

Regardless of the future, these visionary pieces have left me excited for the possibilities to change higher education for the better.

Assessment practice has certainly been a beneficial practice in higher education over the last thirty years. Much good has come from it, notably faculty working together within a program as opposed to merely independent contractors. Now, it’s time to level-up and build on the progress made. I leave you with three hopes for the future. First, I hope accrediting agencies will recognize the need to create space to engage in more interesting work. It’s hard to imagine innovation while turning the crank to ensure assessment reports are neatly filed ahead of an accreditation visit. Second, I hope assessment practitioners continue to create partnerships with educational developers. Regardless of vision, the next assessment movement must rely more heavily on the expertise of educational developers. Last, I hope we can create a strategy to inspire higher education to truly value learning and improvement. Learning is often taken for granted (of course it’s happening!) and senior administrators are generally not obsessed with ensuring we’re getting this right and making it better. The pandemic has opened this conversation – many parents and students have started asking tough questions about learning as modalities shifted. Can we harness this attention to support our next vision? We can and we must.

References

- Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice*, 34, 31–40. <https://doi.org/10.1111/emip.12096>
- 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). <http://www.bu.edu/provost/files/2015/09/From-Gathering-to-Using-Assessment-Results-Lessons-from-the-Wabash-Study-C.-Blaich-K.-Wise1.pdf>
- Council for Higher Education Accreditation. (2019). *Accreditation and student success outcomes: Perspectives from accrediting organizations*. CHEA/CIQG Publication Series. <https://www.chea.org/sites/default/files/pdf/Accreditation%20and%20Student%20Learning%20Outcomes%20-%20Final.pdf>
- Eubanks, D. (2017). A guide for the perplexed. *Intersection*, 4-14. https://www.aalhe.org/assets/docs/AAHLE_Fall_2017_Intersection.pdf
- Eubanks, D. A., Vanovac, S. (2020) Divergent Development of Writers in College, *Journal of Writing Assessment*, 4.
- Lederman, D. (2019, April 17). Harsh take on assessment ... from assessment pros. *Inside Higher Ed*. <https://www.insidehighered.com/news/2019/04/17/advocates-student-learning-assessment-say-its-time-different-approach>
- Millet, I. (2010) Improving grading consistency through grade lift reporting. *Practical Assessment Research and Evaluation*, 15(4), 1–8. <https://doi.org/10.7275/zp5s-xe72>
- Millet, I. (2018) The relationship between grading leniency and grading reliability. *Studies in Higher Education*, 43(9), 1524-1535. <https://doi.org/10.1080/03075079.2016.1265497>
- United States Department of Education. (2020). *Predictive analytics to improve student outcomes* [Webinar]. <https://mediasite.ed.gov/webcast/Play/44bab8fe1b54425792e62f452837a5641d>



Podcast: Leading Improvements in Higher Education

Stay connected to hear the latest research-based topics
featured on IUPUI Assessment Institute!

<https://assessmentinstitute.iupui.edu/overview/podcast-episodes.html>.

