Eubanks, D. (2021) Assessing for student success. *Intersection: A Journal at the Intersection of Assessment And Learning, 2(2).*

Assessing for Student Success

By David Eubanks

Intersection: A Journal at the Intersection of Assessment and Learning
Vol 2, Issue 2, Winter 2021

Abstract: Limitations of assessment reporting are analyzed in order to reorient practice toward improving the equity of student success. Reporting practice is contextualized by scholarship within assessment and accreditation organizations, and the implicit assumptions driving current practice are examined. Two novel findings are presented: an assessment of the number of discrete learning goals that comprise an academic program and a demonstration of how course grades can provide insights into curricular barriers in way that standard reporting does not. New assumptions are proposed that are intended to contribute to the evolution of practice so that equity goals may be achieved.

Introduction

Outcome disparities in higher education are well-documented (Chetty et al., 2013). For example, students from higher-income families attend more selective colleges and earn more after graduation than their lower-income peers, challenging the idea that education can be an economic equalizer (New York Times, 2017). Similar disparities can exist within institutions, so that thinking of an "average student experience" elides the experiences of different types of students. Unfortunately, typical assessment reporting is not equipped to meet this challenge. This essay seeks to understand that problem and suggest changes.

The need for change in assessment to rectify learning outcome disparities was addressed in Stitt-Bergh et al. (2019). The authors, representing the national assessment organizations AALHE, NILOA, and AAC&U, charge that "The assessment of student learning in higher education has been headed down an unproductive path for too long" (p.43), and describe the problems stemming from a report-writing culture:

> Too many campuses maintain a view of learning assessment that limits its uses to gatekeeping and providing evidence to external entities such as regional accreditors. An expanded view would position assessment as a tool for equity, program understanding, and improvement of the learning system, all in service to the broader public good.

Similar sentiments can be heard from accrediting agencies (CHEA, 2017, p. 17).

> [I]nstitutional and programmatic accreditors were nearly unanimous in their concern about the trend toward standardized "cookie cutter" definitions and approaches as well as the use of "blunt" measures without regard for the rich heterogeneity of the missions, cultures and student populations of institutions and programs.

The debilitating effects of standardized reporting on data quality and analysis are addressed previously (Eubanks, 2017). The need for more attention to student success across the demographic spectrum has been taken up by the Lumina Foundation and the Higher Learning Commission (2018), and the Council for Higher Education Accreditation (2019). The U.S. Department of Education (2020) recently advocated such changes by hosting a joint presentation by institutions showing significant improvements in student success metrics. These case studies illustrate that concerted action based on good data can result in significant improvements to student success, including for traditionally vulnerable groups.

This essay takes up this challenge by inviting you to take a step back from the report-writing rules in order to reimagine a practice focused on student success. The analysis below shows how the assumptions of report writing stand in the way of equity goals and proposes new assumptions that can better serve the needs of students.

Assessment terminology can sometimes be ambiguous. In this essay, a *learning goal* is what we would like students to learn, a *learning outcome* is what we assessed them to have learned, and *student achievement* refers to course completion, retention, graduation, licensing exam pass rates, employment, loan repayment, and so on. I will use “student success” as a general term to include both learning outcomes and achievement.

There are many levels and types of academic programs that report assessment results. For simplicity, I will focus on four-year undergraduate degrees. These typically comprise 120 credit hours of course work split into general education, elective, and major courses – about 40 three-credit classes (or equivalent) in all. A nominal course meets three times a week over 15 weeks for 50 minutes each time.

Reporting Assumptions

Assessment report-writing, which drives much of the activity in many assessment offices, implies assumptions about learning goals and measurement that inform practice. Here is my attempt to describe the key assumptions.

Assumption 1. Student learning and achievement are sufficiently distinct that projects to improve these are separate and distinct efforts. The major institutional accreditors have separate standards for achievement and learning (e.g., SACSCOC has 8.1 and 8.2, respectively, and HLC has 4B and 4C). A consequence of this assumption is that course completion, for example, cannot be used for most assessment reports as a measure of learning.

Assumption 2. Evidence of learning must be based on discrete student work samples. Assessment data comes from isolated work products like written papers or exams or performances that are scored or graded. Other types of data, like course grades or student surveys, are disparaged as being “indirect measures” (Suskie, 2018, p. 49).

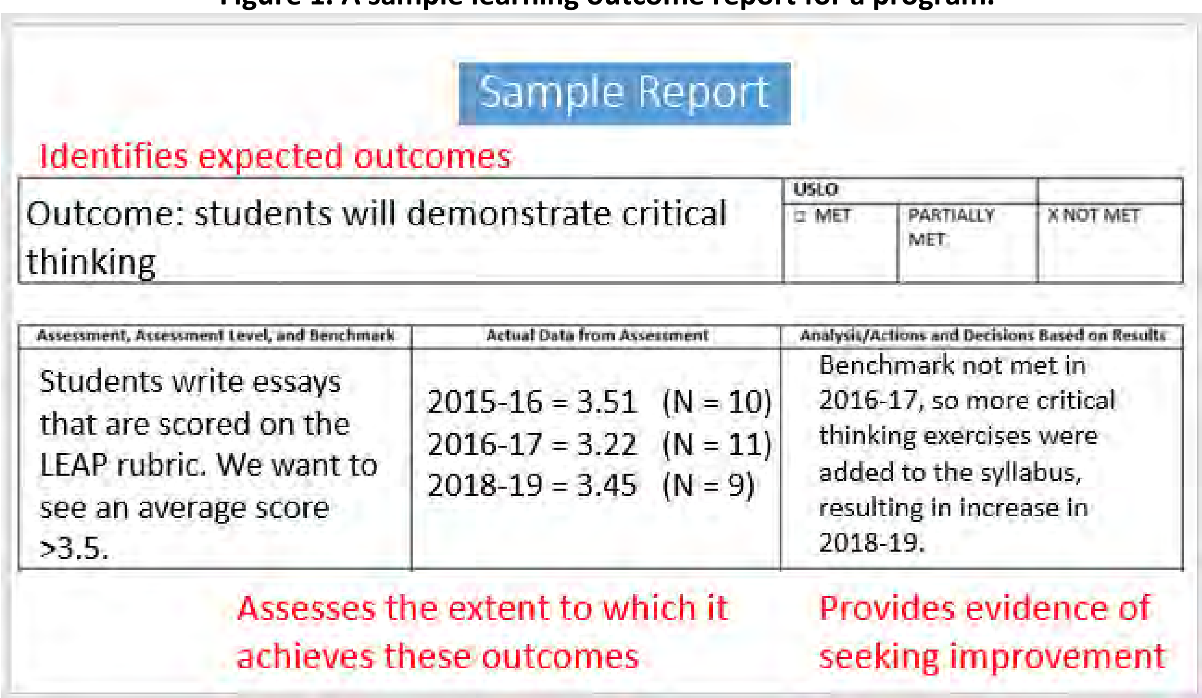
Assumption 3. Reliability and validity of assessment data are good enough to use without checking. As a practical matter, it would be impossible to validate the many sources of data used by an assessment office, so this assumption is necessary whether or not it is true. Sometimes assessment is called “action research” or “design-based research” to excuse the lack of rigor (Suskie, 2018, p. 34; Swarat & Chase, 2020).

Assumption 4. <u>Program quality is evident from the end products.</u> The idea is that the effect of an academic program is not complete until the student is near graduation, and the assessment is of the nearly "final product" with respect to program learning.

Assumption 5. <u>A few higher-order learning goals are sufficient to assess a program.</u> Bloom's Taxonomy or AAC&U LEAP outcomes are often used to define aggregate learning goals that develop over the course of a program. These overarching learning goals are usually the focus of program assessment. Sometimes programs are required to fit learning goals into a hierarchy of course, program, and institutional ones, as with the course, program, and institutional learning goals in WSCUC accreditation.

A typical assessment report is the reification of these assumptions. For example, The SACSCOC *Principles of Accreditation* (2018) Standard 8.2a reads: "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 [for] Student learning outcomes for each of its educational programs" (p. 20). This formal process will be limited to a few outcomes per program (typically five or fewer). Each report includes definitions of the chosen learning goals, summarizes data from a capstone course or other summative assessment where ratings are produced from test scores or rubric ratings of student work samples. The statistical quality of the data is usually not assessed. Not all institutions or programs will follow this outline, but the description aligns with recommendations from Suskie (2018), Middle States (2007), Higher Learning Commission (2018), and SACSCOC (2020).

Figure 1. A sample learning outcome report for a program.


Sample Report

Identifies expected outcomes

Outcome: students will demonstrate critical thinking	USLO		
	☐ MET	PARTIALLY MET	X NOT MET

Assessment, Assessment Level, and Benchmark	Actual Data from Assessment	Analysis/Actions and Decisions Based on Results
Students write essays that are scored on the LEAP rubric. We want to see an average score >3.5.	2015-16 = 3.51 (N = 10) 2016-17 = 3.22 (N = 11) 2018-19 = 3.45 (N = 9)	Benchmark not met in 2016-17, so more critical thinking exercises were added to the syllabus, resulting in increase in 2018-19.

Assesses the extent to which it achieves these outcomes

Provides evidence of seeking improvement

The report template in Figure 1 checks all the boxes needed to be approved by a SACSCOC reviewer. Though the sample sizes are too small to justify the conclusions here, this would not cause a compliance problem (Assumption 3).

Each of the assumptions listed above creates problems for meaningful assessment, and the combination of them adds up to the counterproductive report writing that has been criticized by assessment leaders and accreditors. The focus of the next section is on the last assumption, about consolidation of learning goals. Because standard assessment reports can only accommodate a few learning goals, the number of goals that comprise a program of study is crucial.

Counting Learning Goals

A reasonable lower limit on the number of learning goals in a typical undergraduate degree program can be obtained by assuming that each week of a class, a student learns a new topic. Forty (three-credit) classes in a bachelor's degree and 15 weeks per class sums to 600 learning goals. However, only one outcome per week seems like an underestimate. If it were one outcome per class meeting, that triples the count to 1,800 learning goals. Thinking that the answer is probably somewhere between those bounds, I consulted the table of contents of a Calculus textbook I have taught from.

Figure 2. A chapter outline from a calculus textbook

Limits and Derivatives

The selection in Figure 2 is from the table of contents in Stewart (2015), an introductory calculus text, showing the chapter on limits and derivatives. The topics are hierarchical, so that the sections of this chapter all concern limits and derivatives (a special kind of limit), and the section topics are also partly sequential, so that the section on continuity requires knowledge of limits found in the prior section. This organization has evolved over time to somewhat standardize what comprises an introductory calculus class.

Each section listed in the chapter has a learning goal implied by the title, with this section organizing material on mathematical limits and derivatives. The learning goals within each subsection are still complex topics. Here is the precise definition in Section 2.4 (Stewartm 2015, p. 109):

Figure 3. Textbook definition of a limit

3 Definition of Left-Hand Limit

$$\lim_{x \to a^-} f(x) = L$$

if for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that

$$\text{if} \quad a - \delta < x < a \quad \text{then} \quad |f(x) - L| < \varepsilon$$

Concepts addressed in the definition in Figure 3 include: (1) using absolute values with inequalities, (2) recognizing and writing Greek letters, and (3) the understanding that epsilon is envisioned as a very small number. This last point relates to a paradox that bedeviled mathematicians for hundreds of years: *how can a number be infinitely small but greater than zero?* In a general education context, one could even mention the western Church's reaction to this paradox (Alexander, 2014). Different applications of the definition add more complexity to the learning goal. These cases can look very different to students depending on the algebra involved.

Practice problems at the end of the chapter are designed as assessments for various ways that the definition in Figure 3 can be applied. There are standard analytical tricks to use on these problems, like adding and subtracting the same quantity to turn one complicated expression into two simpler ones. For mathematics majors, these techniques develop into the topic of real analysis that extends through the undergraduate and graduate curricula.

My conclusion from the textbook review is that the material in the chapter subsections each easily contain enough content to be considered at least one learning goal, each with identifiable assessments. These learning goals are highly structured, both in sequence and in hierarchy. A typical first course in calculus would cover about seven chapters in a class meeting four periods per week for a semester, for around 50-chapter sections total. Prorating to a three-credit class gives an estimate of about 37 learning goals for a three-credit class (about one per class meeting, with time out for exams and activities). This estimate is somewhere between one new learning objective per class day and one per week, which seems reasonable. Reviews of an organic chemistry text and a psychology text give similar numbers to my calculation for introductory calculus.

For a 40-course bachelor's degree, 37 learning goals per course totals to about 1,500. If a degree is nominally one-third each for general education, electives, and major courses, then a major program might be seen to include only 500 learning objectives. However, since not every student will take every course offering in most programs, the curriculum is likely to be larger, and the count higher. The mathematics curriculum at my university has 30 standard courses, while history has about twice as many. These are four-credit classes, so a reasonable range of learning goals at 50 per course is 1,500-3,000 per program. Because of the diversity of course offerings, a count of general education program learning goals could be even larger.

In summary, I estimate that a typical undergraduate will be exposed to around 1,500 learning goals over four years, and that a typical academic major program or general education curriculum will have at least that many each.

Learning Hierarchies

The number of learning goals in an academic program is evidently large. Practicality demands that only a few of these can receive formal treatment, with curated definitions, data-gathering, and reports. Suskie (2018) recommends to “Focus assessment efforts on a limited number of learning goals that you and your colleagues feel are most important – perhaps no more than three to six […] (p. 153)” and cautions that (p. 50):

> I have seen associate degree programs with 45 learning goals and general education curricula with over 100. There is no way that students can learn so many things well in the time we have with them or that faculty can assess them meaningfully. Consider focusing on just three to six key learning goals for each program and for each general education requirement.

No evidence is given to support the claim that students cannot achieve 45-100 learning goals while earning a degree. As seen in the previous section, there are more than that in a single year of college. But the point about the number of projects that the faculty in an academic program can formally manage is valid, leaving the question open about how the large remainder is to be reported on, if at all.

One method of handing the large number of learning goals is to only assess “higher order” outcomes that are thought to subsume many “lower order” ones. This is described the Middle States guide (2010, p. 13) (emphasis added).

> An SLO refers to an <u>overarching</u> outcome for a course, program, degree or certificate, or student services area (such as the library). SLOs describe a student’s ability to <u>synthesize many discreet [*sic*] skills</u> using higher level thinking skills and to produce something that asks them to apply what they’ve learned. SLOs usually encompass a <u>gathering together of smaller discrete objectives</u> […] through analysis, evaluation and synthesis into more sophisticated skills and abilities.

A glossary referenced in the Middle States guide (2010, p. 10) refers to the limited course-based goals as “learning objectives,” which are:

> small steps that lead toward a goal, for instance the discrete course content that faculty cover within a discipline. Objectives are usually more numerous and create a framework for the overarching student learning goals which address synthesizing, evaluating and analyzing many of the objectives.

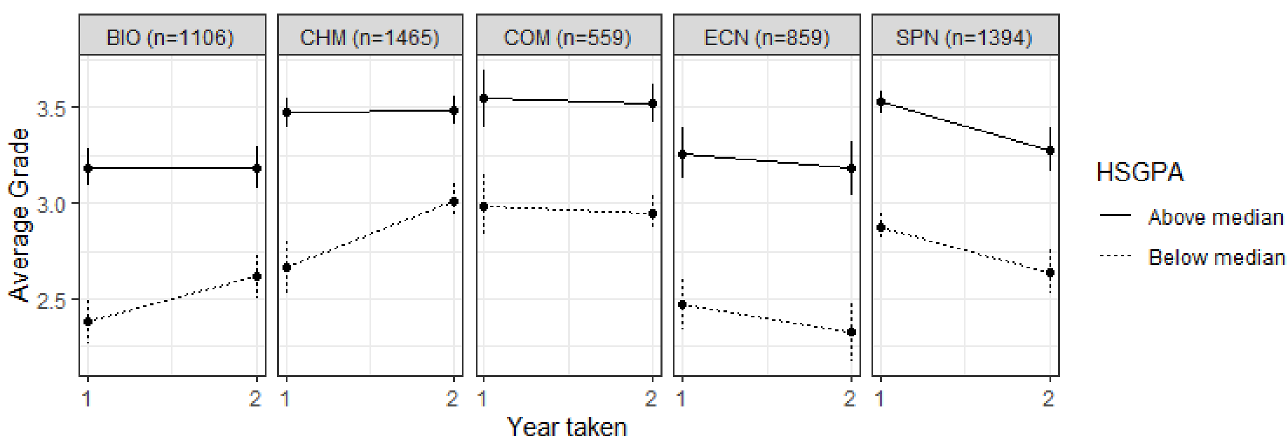
These statements describe the basis for Assumption 5, that assessing a few “higher order” learning goals is sufficient to understand program quality. The choice of verbs in the quote above is from Bloom’s Taxonomy (Bloom, 1956), which is suggested as one way of organizing the objectives hierarchically into broad learning goals that in principle subsume the proliferation of more focused ones.

However valuable assessing cumulative skills may be, assessing program quality surely still needs to include the myriad learning goals within courses. It is unreasonable to believe that an assessment of thinking or communication skills (for example) can simultaneously give us information on the hundreds of discrete learning goals a student was expected to achieve. As a result of the focus on cumulative goals, the content taught in courses is largely invisible to assessment reporting and barriers to student success will probably go unnoticed in official reports. The importance of such detailed understanding of the curriculum is illustrated with an example in the next section.

Example: First Year Class Schedules
Should first-year college students continue study in class subjects they recently had in high school, like foreign language or biology, or should they wait a year if the course seems difficult, to give them time to adjust? On one hand, students who wait will begin to forget what they learned and be less prepared when they do enroll in the course. On the other hand, we do not want to overburden new students with a schedule that is too difficult during the critical first year. How does the answer this question depend on student characteristics?

Figure 4 shows grade averages for students in the first course they take at my university in Biology (BIO), Chemistry (CHM), Communication Studies (COM), Economics (ECN), or Spanish (SPN). Grade averages are disaggregated by high school grades (above or below median GPA) and whether they take the first course in the first year of college or the second.

Figure 4. Grade averages for introductory courses in various disciplines, disaggregated by median high school grade averages and the year of college that the course was taken.



Notes: Sample sizes appear in the titles. Error bars are two standard errors of the mean.

Two trends appear in Figure 4. Students with lower grade averages in high school also have lower grade averages in the language courses. For most subjects, the higher-HSGPA group did not see a change in average grades, but the lower-HSGPA group shows more sensitivity to a delay. For Biology and Chemistry, it appears that waiting may be beneficial for that group, whereas Spanish shows a decline for both groups.

The results suggest that for higher-HSGPA students, it does not matter when they first take these subjects in college, except for foreign language, which they should take right away. For lower-HSGPA students, it is better to wait for science courses but not for foreign language ones. Because of selection effects and other student characteristics, these recommendations are subject to review by faculty who teach these courses, especially to identify confounding variables.

This example illustrates that we should not focus solely on senior-level achievement and higher-order learning goals, and that student success rates vary by student type: an equity concern. It also illustrates the usefulness of grade data. The groups with lower grade averages in Figure 4 are likely subject to a disadvantage that impedes learning, like having forgotten language vocabulary and grammar or having a difficult schedule. Early impediments can have consequences throughout a student's career. But for such an analysis, we need data from all courses and all students. These concerns are invisible to report-writing work that focuses on high-level outcomes in senior classes. An obvious solution is to routinely analyze grade data.

Course Grades

The inability of program assessment reports to assess most learning within a program is a critical flaw. The focus on a few data points at the end of a program also makes it difficult to assess differences between student types, cannot discover subtle issues like the waiting effect illustrated above, and the results come too late to help the students being assessed. There is a need for data that is timely, easily obtained, and that speaks to course content. Course grades are ideally suited for the purpose, particularly since they are attached to student identifiers that can be used to easily match to demographic or other categories. And course grades can create "gatekeeping" that limits student opportunities. For example, students must maintain a high enough grade average to avoid suspension and remain in a major program.

Given the obvious utility of course grades as assessment data, it is unfortunate that they are not widely used in accreditation reports for assessment of learning. This proscription seems to have been motivated for a desire for more objective measures of learning (Ewell, 2009). The consequence is that various sources simultaneously acknowledge that grades can have value as indicators of learning while discouraging the use of grades for assessment (Middle States, 2007; Suskie 2018). For example, on the topic of grades topic, Suskie (2018, p. 10) notes that:

> We can conclude from a grade of B in an organic chemistry course, for example, that the student has probably learned a good deal about organic chemistry. But that grade alone cannot tell us exactly what aspects of organic chemistry she has and has not mastered.

But, she later describes grades as "indirect" measures of learning, unsuitable as a primary source. Messick (1994) finds no use for the direct/indirect distinction; it circumvents validity-checking by arbitrarily favoring some sources over others.

It is curious that most accreditors seem unconcerned by their implicit assumptions that grades are low-quality indicators of student learning. This collective shrug is difficult to square this view with the attention otherwise paid to the integrity of transcripts. We care that course instructors are qualified, that transfer credit is validated, and that accurate records are kept. But are we not to care that grades actually indicate that learning occurred? In a study of grade inflation, Kamber (2008) surveyed accreditors and higher education associations on the matter, concluding that

"The implicit message here is 'don't ask, don't tell'" (p. 175).

New Assumptions

The need for institutions to understand and address disparities in student outcomes is becoming dire. Stitt-Bergh et al. (2019) urge us to look beyond accreditation reports, and to do so we need a new set of assumptions. Kuh et al. (2015, p. 50) took up the question of evidence needed for improvement, concluding that the focus should be on the evidence.

> Moving from a compliance model of assessment to one focused on improving students' educational experiences means putting a premium on evidence. It also means, we have suggested in this chapter, being smart about what constitutes evidence and how to use it effectively. What matters for assessment for the improvement of student learning is not the amount of information gathered, but its usefulness to the community—and usefulness, or actionability, is determined by a number of factors: the quality of the evidence; its technical properties; and its match with the interests, questions, and dispositions of those who will be using it. Moving from a compliance model of assessment to one focused on improving students' educational experiences means putting a premium on evidence.

The need for such rich evidence was noted as far back as the 1992 "Principles of Good Practice for Assessing Student Learning" (Hutchings et al., 2012) in its second principle: "Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time."

I propose the following assumptions, which are designed to bridge the evidence gap between report writing and the aspirations of more useful data and student equity.

Assumption 1. Student learning and achievement are closely related and should be studied together. Separate accrediting standards on learning should be integrated into the student success standards, so that assessment offices "get credit" for increasing pass rates, retention, graduation, placement, and so on, in addition to custom measures of learning.

Assumption 2. Evidence of achievement and learning should be defined by the institution with respect to its mission. Existing reporting assumptions severely limit data sources and methods that can be used in reports, which leads to small sample sizes and (often) inappropriate emphasis on quantitative assessments. This includes the ability to customize all aspects of the assessment program: definitions, organization, and assessment data and methods. If learning goal hierarchies are not useful to an institution, they should not be required.

Assumption 3. The statistical qualities of assessment data should be tested if they are to be published or used statistically. We should study the reliability and validity of grades and other achievement data so that transcripts and published reports on student achievement can be trusted as meaningful. Data that are not going to be used statistically do not need this level of scrutiny, and we should distinguish between the two.

Assumption 4. Student achievement should be assessed throughout a program, not just at the end. To understand barriers to student success, we must analyze the whole curriculum to understand the biases and barriers represented by courses and to understand which student populations are most vulnerable. Course grades are a good (free) resources for this, and nicely complement custom assessments of learning (see Eubank, 2020).

These assumptions are mirror images of the ones driving current report writing practice and describe a quantitative assessment program that complements two other activities: program review and faculty professional development through assessment offices and teaching/learning centers. Together these efforts form a triad that: (1) works toward equity in outcomes and success in all groups, (2) assures meaningfulness of transcripts, (3) ensures that programs have sufficient resources and human capital to be effective, (4) ensures that curricular are reviewed and remain current, and (5) provides faculty with resources and motivation to continually improve teaching.

The last of these is the most important for continual improvement of student learning. Current report writing is described as empirical research that improves teaching and learning. However, even when it succeeds, the focus on one learning goal at a time is very inefficient. Given the large number of learning goals in a program, improving teaching generally through faculty development has a much larger impact. For example, a course redesign will affect 30-40 learning goals. High quality research (local or in publications) on teaching can inform practices, and assessment's role can center on: (1) informal (non-statistical) classroom data that the faculty find useful, (2) course grades, and (3) a small number of large-scale, high-quality research efforts.

Conclusions

The rigid requirements of current reporting practice have the laudable aim of objectively understanding student learning. Unfortunately, the bureaucratization of this goal has resulted in "cookie-cutter" reporting rules that are so inflexible that the work is often at odds with ensuring the equity of student outcomes. The emphasis on report-writing must change to value student success as defined by the institution, within the scope of its mission, using data and methods that it finds credible. As a replacement to report-writing, I have suggested a flexible triad of student success research, program review, and faculty development. There are signs that this will work, from my own research (Eubanks, 2020), from innovative assessment programs (Metzler & Kurtz, 2018), and from impressive results in student success (U.S. Department of Education, 2020).

Stitt-Bergh et al.'s (2019) call for reform will go nowhere without reflection and action within the assessment community. Ultimately this must result in practitioners, rather than accreditors and consultants, defining standards of best practice, which may vary by institution and program. Achieving student success cannot be reduced to a list of checkboxes, no matter how well-intentioned.

References

Alexander, A. (2014). The secret spiritual history of calculus. *Scientific American.* https://www.scientificamerican.com/article/the-secret-spiritual-history-of-calculus/.

Bloom, B. (1956). *A taxonomy of cognitive objectives.* McKay.

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2013). The equality of opportunity project. Brookings Institution.

Council for Higher Education Accreditation. (2019). Accreditation and student success outcomes: Perspectives from accrediting organizations. *CHEA/CIQG Publication Series.*

Eubanks, D.A., (2017) A guide for the perplexed. *Intersection: A Journal at the Intersection of Assessment and Learning,* Fall 2017, 14-17.

Eubanks, D.A., Vanovac, S. (2020). Divergent development of writers in collect. *Journal of Writing Assessment,* 4, 15-54. https://doi.org/10.37514/JWA-J.2020.4.1.03

Ewell, P. T. (2009). Assessment, accountability, and improvement. *NILOA Occasional Papers*, *1.* University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Higher Learning Commission. (2018). Defining student success data: Recommendations for changing the conversation.

Hutchings, P., Ewell, P., & Banta, T. (2012, May). AAHE principles of good practice: Aging nicely. University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Ewell, P. T., Hutchings, P., & Kinzie, J. (2015). *Using evidence of student learning to improve higher education*. John Wiley & Sons.

Kamber, R. (2008). Combating grade inflation: Obstacles and opportunities. In Hunt, L. H. (Ed.), *Grade inflation: Academic standards in higher education* (pp. 171-189). SUNY Press.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(3), 13-23.

Metzler, E. T., & Kurz, L. (2018). Assessment 2.0: An organic supplement to standard assessment procedure. *NILOA Occasional Papers*, 36*.* University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Middle States Commission on Higher Education. (2007). Student learning assessment: Options and resources.

New York Times. (2017). Economic diversity and student outcomes at America's colleges and universities. https://www.nytimes.com/interactive/projects/college-mobility/

Southern Association of Colleges and Schools, Commission on Colleges. (2018). Principles of accreditation: Foundation for quality enhancement.

Suskie, L. (2018). *Assessing student learning: A common sense guide.* John Wiley & Sons.

Stitt-Bergh, M., Wehlburg, C. M., Rhodes, T., & Jankowski, N. (2019). Assessment for student learning and the public good. Change: The Magazine of Higher Learning, 51(2), 43-46.

Swarat, S., Chase, D. (2020) Right to the point: Student learning and advanced topics in assessment. WASC Senior College and University Commission. https://www.wscuc.org/events/right-point-student-learning-and-advanced-topics-assessment

U.S. Department of Education (2020) Predictive analytics to improve student outcomes, https://mediasite.ed.gov/webcast/Play/44bab8fe1b54425792e62f452837a5641d

Young, D. (2020, December 2-9) Student outcomes 8.2[Conference session], Annual meeting of the Southern Association of Colleges and Schools, Commission on Colleges, Virtual.

About the Author

Dr. David Eubanks is the Assistant Vice President of Assessment and Institutional Effectiveness at Furman University in Greenville, South Carolina. He can be reached at: david.eubanks@furman.edu